



# **XVII Congreso de Metodología de las Ciencias Sociales y de la Salud**

**Universidad de Zaragoza**

**18-22 de julio de 2022**

**LIBRO DE RESÚMENES**

**(Versión 14/07/2022)**

# Tabla de Contenidos

Meta-Análisis: Avances Estadísticos .....	1
Aplicaciones de la Metodología Observacional en Ámbitos Profesionales.....	3
Creación y Análisis de Pruebas Psicométricas .....	5
Validación del Constructo de Cognición Social.....	6
Investigando Metodología en el Multiverso de la Academia: Comunidad, Recursos y Estrategias por y para Jóvenes .....	8
Diseños de Medidas Repetidas: Del Modelo Lineal General al Modelo Lineal Mixto.....	10
Meta-Análisis de Generalización de la Fiabilidad .....	12
Innovaciones Metodológicas en Evaluación de Programas: Calidad Metodológica y Tamaño de Efecto de Programas de Intervención .....	14
Medición de la Recuperación en Personas Diagnosticadas con Trastorno Mental Severo .....	17
Mejorando los Test y su Uso: Hecho y por Hacer.....	19
Meta-Análisis: Avances Metodológicos.....	21
Metodología Cualitativa para el Avance en Psicología.....	23
Retos y Propuestas para Conciliar la Teoría de la Validez y las Prácticas de Validación .....	25
Mesa 1 .....	27
Mesa 2 .....	29
Mesa 3 .....	31
Mesa 4 .....	33
Mesa 5 .....	34
Mesa 6 .....	36
Mesa 7 .....	38
Mesa 8 .....	39
Mesa 9 .....	41
Mesa 10 .....	43
Mesa 11 .....	44
Mesa 12 .....	46
Sesión de Pósteres 1.....	48
Sesión de Pósteres 2.....	59
Sesión Premio AEMCCO Jóvenes Investigadores 2022 .....	68

## SIMPOSIO

### Meta-Análisis: Avances Estadísticos

Coordina: José Antonio López López

**Resumen.** A lo largo de las últimas décadas, el meta-análisis se ha consolidado como una metodología imprescindible para una acumulación eficiente del conocimiento científico. Un meta-análisis consiste en la integración cuantitativa de los resultados reportados por un conjunto de estudios primarios que examinaron un mismo fenómeno que permite, en primer lugar, la obtención de un valor medio de dichos resultados. Además, es común encontrar un cierto grado de heterogeneidad entre los hallazgos de los estudios individuales, y por ello un campo de investigación muy activo es el de los métodos estadísticos para identificar variables que ayuden a entender esta heterogeneidad. Este campo es conocido como análisis de moderadores (o modificadores del efecto), y en él se centran las cinco propuestas de este simposio. La primera comunicación, presentada por Juan Botella Ausina, pone el foco en el problema de analizar variables moderadoras a nivel del individuo como si se tratara de variables a nivel del estudio. La segunda comunicación, presentada por Belén Fernández Castilla, analiza las características metodológicas de los meta-análisis en red llevados a cabo en psicología y en ciencias de la educación. La tercera comunicación corre a cargo de Belén Fernández Castilla, y en ella se presenta una aplicación de meta-análisis clásico y meta-análisis en red dentro del ámbito educativo. La cuarta comunicación, presentada por José Antonio López López, ofrece una nueva perspectiva para el análisis de moderadores gracias al uso de modelos de localización y escala aplicados al meta-análisis. Finalmente, Desirée M<sup>a</sup> Blázquez Rincón presenta la quinta comunicación, en la que ofrece los resultados de un estudio de simulación dirigido a examinar el funcionamiento de los modelos de localización y escala en meta-análisis.

Este simposio ha sido fundado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104033GA-I00 /MCIN/AEI/10.13039/ 50110 00110 33

#### **Problemas para evaluar moderadoras dicotómicas de individuo como moderadoras de estudio en meta-análisis.**

Juan Botella<sup>1</sup>, Mario Calabria<sup>1</sup> y Manuel Suero<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

En meta-análisis es frecuente que interese evaluar el papel moderador de variables como el sexo en el tamaño del efecto. El sexo es una variable dicotómica a nivel

individual, pero la información del nivel de los individuos no está disponible en un meta-análisis. Más bien se suele disponer de la frecuencia relativa de hombres y mujeres (operacionalizada, por ejemplo, como la proporción de mujeres) a nivel de estudio. El potencial papel moderador de esta variable se evalúa ajustando un modelo de meta-regresión con dicha frecuencia relativa como predictora. En esta comunicación se exponen algunos problemas que surgen en la evaluación de este tipo de variables. En concreto, las dificultades que se derivan de las restricciones de rango que suelen acompañar a su muestreo y el papel que desempeñan las estimaciones con muestras puras. También se expondrán formas alternativas de evaluar su papel.

#### **Meta-análisis de red en psicología y ciencias de la educación: Una revisión sistemática de sus**

**características** Belén Fernández-Castilla<sup>1</sup> y Wim Van den Noortgate<sup>2</sup>

<sup>1</sup>Universidad Nacional de Educación a Distancia, <sup>2</sup>Universidad de Leuven

El meta-análisis en red (MR) es una de las mejores técnicas para combinar los resultados de estudios que exploran la eficacia de varias intervenciones. El MR se ha aplicado principalmente en el campo de la medicina, mientras que en el ámbito de la psicología y las ciencias de la educación su uso es menos frecuente. En consecuencia, las revisiones sistemáticas que describen las características de las MR se limitan al ámbito médico, y no se sabe nada sobre las características de las MR publicadas en el ámbito de la psicología y las ciencias de la educación. Sin embargo, esta información sigue siendo relevante para el diseño de futuros estudios de simulación, y para detectar buenas y malas prácticas de investigación en nuestro ámbito. Así, este estudio describe las características de los MR publicados en psicología y ciencias de la educación, así como sus características metodológicas, y las compara con las observadas en el ámbito médico. Los resultados muestran que, en psicología, el número de estudios incluidos en las síntesis es mayor, el tamaño del efecto más utilizado es la diferencia de medias tipificada (a diferencia de la razón de ventajas en medicina), el tamaño de la muestra es menor, se incluyen más grupos de intervención y se observan efectos inconsistentes con mayor frecuencia. Estos resultados pueden utilizarse en futuros estudios de simulación para generar conjuntos de datos realistas. Por último, se advierte de la escasa calidad del informe de algunos aspectos técnicos del MR, como el modelo estadístico utilizado.

## **El meta-análisis de red aplicado a investigación en ciencias de la educación: Potenciales y limitaciones.**

Belén Fernández-Castilla<sup>1</sup>, Zi Yan<sup>2</sup>, Hongling Lao<sup>2</sup>, Ernesto Panadero<sup>3</sup>

<sup>1</sup>Universidad Nacional de Educación a Distancia, <sup>2</sup>The Education University of Hong Kong, <sup>3</sup>Universidad de Deusto

El meta-análisis de red es una técnica que permite combinar cuantitativamente los resultados de estudios que comparan la eficacia de múltiples intervenciones. Esta técnica permite obtener efectos indirectos entre intervenciones a través de un comparador común, que suele ser el grupo control. Por lo tanto, una de sus ventajas es que el meta-análisis de red cuenta con más tamaños del efecto (efectos directos + indirectos) que un meta-análisis tradicional, y por lo tanto la estimación del tamaño de efecto combinado es más precisa. Además, gracias a la obtención de efectos indirectos, es posible comparar intervenciones que nunca han sido comparadas en investigación primaria. Sin embargo, para poder aplicar esta técnica hay que cumplir con el supuesto de transitividad: las características de los estudios tienen que ser homogéneas entre sí. En este trabajo se discute si es posible cumplir con este supuesto en el ámbito de las ciencias de la educación, donde las características de los estudios primarios suelen ser tan heterogéneas. Específicamente, expondremos las dificultades encontradas por este equipo para llevar a cabo un meta-análisis de red sobre la eficacia de diferentes estrategias de evaluación (i.e., autoevaluación, evaluación por pares, evaluación del profesorado) en el rendimiento académico. La imposibilidad de reunir un conjunto homogéneo de estudios y cumplir así con el supuesto de transitividad hizo inviable la ejecución de un meta-análisis de red, siendo finalmente más efectivo aplicar un meta-análisis tradicional por pares. Finalmente, se discutirán posibles estrategias para futuros meta-análisis de red en este campo.

## **Análisis de moderadores de la cantidad de heterogeneidad en meta-análisis.** José Antonio López-López<sup>1</sup>, Wolfgang Viechtbauer<sup>2</sup>

<sup>1</sup>Universidad de Murcia, <sup>2</sup>Maastricht University

En meta-análisis, es habitual encontrar heterogeneidad entre los tamaños del efecto extraídos o calculados a partir de los datos de los estudios primarios. Por ello, una etapa adicional del análisis consiste en la búsqueda de moderadores asociados con dicha heterogeneidad. En este sentido, los modelos de meta-regresión constituyen una opción flexible que permite incorporar variables moderadoras tanto categóricas como cuantitativas. Hasta la fecha, los modelos de meta-regresión disponibles solo

permitían examinar la asociación entre los moderadores y la magnitud de los tamaños del efecto, asumiéndose un valor constante para el parámetro de heterogeneidad del modelo, la varianza inter-estudios. Sin embargo, este supuesto podría no cumplirse en determinadas situaciones, y esto sugiere la conveniencia de disponer de modelos más flexibles que permitan más de un valor para la varianza inter-estudios y, en consecuencia, la incorporación de moderadores de la cantidad de heterogeneidad encontrada en el meta-análisis. En esta comunicación se presentará una extensión de los modelos de meta-regresión tradicionales que relaja el supuesto de un único valor de la varianza inter-estudios para el meta-análisis. Se utilizará la etiqueta “location-scale models” (modelos de localización y escala) para referirse a estos modelos, con los componentes “location” y “scale” referidos a la magnitud de los tamaños del efecto y a su varianza, respectivamente. La estimación de parámetros de estos modelos se lleva a cabo mediante técnicas de máxima verosimilitud (completa o restringida) y es posible incorporar moderadores diferentes para cada una de las dos partes del modelo. Los nuevos modelos están disponibles para su uso a través del paquete metafor de R. En esta comunicación se ilustrará el ajuste de “location-scale models” en metafor mediante un ejemplo ilustrativo.

## **Rendimiento de los modelos de localización y escala en meta-análisis: Un estudio de simulación.** Desirée Blázquez-Rincón<sup>1</sup>, Wolfgang Viechtbauer<sup>2</sup>, y José Antonio López-López<sup>1</sup>

<sup>1</sup>Universidad de Murcia, <sup>2</sup>Maastricht University

Los modelos de localización y escala suponen una mejora en el campo del meta-análisis, ya que permiten estudiar al mismo tiempo el influjo de variables moderadoras sobre la media (localización) y la varianza (escala) de la distribución de efectos. Aunque recientemente se ha añadido al paquete metafor una implementación de estos modelos, todavía no se ha examinado su rendimiento. Se examinaron las propiedades estadísticas de los procedimientos de estimación e inferencia con respecto a la parte de escala del modelo, comparando diferentes métodos de estimación (máxima o máxima verosimilitud restringida), pruebas de significación (de tipo Wald o de razón de verosimilitudes) y métodos para construir intervalos de confianza (de tipo Wald y de perfil de verosimilitud). Para ello, realizamos un estudio de simulación de Montecarlo en el que las condiciones variaban según el tipo de moderador del parámetro de heterogeneidad (cualitativo o continuo) y la cantidad de heterogeneidad fijada para cada nivel del moderador. En general, la prueba de razón de verosimilitudes mostró

tasas de error tipo I más cercanas al nivel nominal y una mayor potencia estadística que la prueba de tipo Wald. En la mayoría de los escenarios, el intervalo de perfil de verosimilitud mostró probabilidades de cobertura más bajas que el intervalo de tipo Wald, pero más cercanas al nivel nominal del 95%. En cuanto al método de estimación, se obtuvieron tasas de rechazo más deseables e intervalos más estrechos cuando se utilizó la estimación REML en lugar de la ML. Finalmente, se obtuvieron tasas de rechazo y probabilidades de cobertura ligeramente superiores cuando se examinó un moderador cualitativo en lugar de uno cuantitativo. Contando con ciertas restricciones estadísticas y con la falta de convergencia de algunos procedimientos, se comprobó que los modelos de localización y escala son una herramienta útil para modelar el parámetro de heterogeneidad.

## SIMPOSIO

### Aplicaciones de la Metodología Observacional en Ámbitos Profesionales

**Coordinan:** Triana Aguirre Delgado y Elena Rodríguez Naveiras

**Resumen.** En Ciencias Sociales, de la Salud y del Comportamiento es común el uso de la metodología observacional, un procedimiento de registro sistemático objetivo, que posibilita el estudio del comportamiento humano en situaciones naturales y cuasi-naturales. Es por ello que resulta relevante contar con instrumentos de probada bondad psicométrica que den información de lo que sucede en una situación natural en diferentes ámbitos profesionales. Si bien se trata de una metodología muy utilizada tanto en el ámbito del deporte, como en estudios observacionales desarrollados en el contexto educativo, este procedimiento adquiere especial relevancia desde un punto de vista formativo, más concretamente en la adquisición de competencias profesionales. En este simposio se presentan diversos trabajos en los que se hace uso de instrumentos de observación. Por un lado, algunos de ellos se enfocan en la adaptación del protocolo de observación para evaluar la adquisición de competencias empáticas en el contexto clínico, así como el posterior estudio de este proceso y entrenamiento de estas competencias. Y, por otro lado, se presenta el entrenamiento del equipo de observación que realizó la depuración de Protocolo de Observación de la Mediación, Técnicas y Estrategias (PROME-TE, v.2) realizado completamente online, dada la situación de crisis sanitaria por la COVID-19 y, posteriormente, se expone la fiabilidad interobservadores de dicho instrumento.

### Adaptación de un protocolo de observación para evaluar la adquisición de competencias empáticas en el contexto clínico en estudiantes de psicología. Emilio Verche<sup>1,3</sup>, Dylan Gutiérrez-Bachiri<sup>1</sup>, Paula Lijó<sup>2</sup>, Ricardo de Pascual-Verdú<sup>1</sup> y Sònia Gregorio<sup>1</sup>

<sup>1</sup>Universidad Europea de Madrid, <sup>2</sup>Universidad Autónoma de Madrid, <sup>3</sup>Centro de Estudios Superiores Cardenal Cisneros

El empleo de la empatía en entornos clínicos conlleva múltiples beneficios a pacientes y profesionales. Sin embargo, no existen protocolos de observación capaces de medir su adquisición en Psicología. El objetivo de este estudio es el desarrollo de un protocolo específico para la evaluación de competencias empáticas. En este estudio se adaptó el sistema de observación por categorías ACOVEO de clasificación de verbalizaciones en categorías funcionales (de Pascual-Verdú et al., 2019) para la evaluación de la adquisición de competencias empáticas. Se añadieron categorías morfológicas que se correspondían con contenidos relacionados con la empatía. Este sistema se utilizó con un total de 37 estudiantes del grado en psicología a los que se les realizó una evaluación de competencias objetiva y estructurada de simulación clínica con un actor/actriz que hacía de pacientes. Al estudiante se le pidió que realizara el rol de terapeuta. La sesión tenía una duración de 10 minutos. Dos investigadores entrenados ciegos al objetivo de la investigación y la condición de los participantes realizaron la observación. Tras confeccionar la versión inicial de la adaptación del sistema de registro, los observadores registraron un 10% de la muestra (16 situaciones) y compararon sus registros. En los casos en los que surgían desacuerdos, se analizaba el motivo y, de ser necesario, se mejoraba la definición de la categoría o se alcanzaban acuerdos relativos a cómo se codificarían verbalizaciones determinadas. A continuación, se procedía de nuevo al registro de las situaciones, repitiéndose el proceso hasta que el acuerdo interjueces alcanzó un valor de kappa de Cohen aceptable. El sistema final de categorías adaptado se compone de 3 categorías para verbalizaciones empáticas: específicas personalizadas, inespecíficas personalizadas e inespecíficas generales; 4 categorías de verbalizaciones emotivas: específica general, de cadena apetitiva, de cadena aversiva y coloquial; 2 categorías para instrucciones discriminativas, y otras 2 para instrucciones de refuerzo y castigo. En conclusión, el sistema de categorías adaptado para la evaluación de las competencias empáticas muestra una fiabilidad y validez óptima. Este sistema permitiría desarrollar metodologías de enseñanza que capaciten para el desempeño clínico empático.

**Una aproximación multimétodo al estudio de la adquisición de competencias empáticas.** Paula Lijó<sup>1</sup>,

Dylan Gutiérrez-Bachiri<sup>2</sup>, Margarita G. Márquez<sup>2</sup>, Sara Vilas<sup>2</sup> y Emilio Verche<sup>3</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad Europea de Madrid, <sup>3</sup>Universidad Complutense de Madrid

La investigación de procesos tiene como objetivo averiguar cómo se producen los cambios terapéuticos y cuál es el proceso que subyace dicho cambio. En este sentido, la empatía se erige como una habilidad importante en ese proceso de cambio. Sin embargo, se trata de un problema de investigación complejo para el que los enfoques multimétodo conforman una aproximación comprensiva. El objetivo de este trabajo es estudiar el proceso de adquisición y entrenamiento de competencias empáticas en estudiantes universitarios mediante el registro de verbalizaciones empáticas. Se realizó una aproximación multimétodo empleando métodos experimentales y observacionales para el análisis de verbalizaciones empáticas. Se diseñaron cuatro condiciones experimentales: simulación, role-playing, clase magistral y control, para poder observar las diferencias pre-post de 37 estudiantes del grado de psicología en una evaluación de competencias objetiva y estructurada. Para la evaluación observacional se adaptó el sistema ACOVEO (de Pascual-Verdú et al., 2019) para la clasificación de verbalizaciones empáticas de los participantes. Los resultados se analizaron con un ANOVA mixto 4x2 comparando entre grupos pre-post antes y después de pasar por las condiciones experimentales en las distintas categorías de las verbalizaciones empáticas. No se encontraron diferencias significativas en las condiciones pre-post, ni entre las condiciones experimentales ni la interacción, para la adquisición de competencias empáticas. La aproximación multimétodo permite un acercamiento cuantitativo a través de diferentes metodologías al estudio de la adquisición de competencias empáticas de una forma más ecológica, recogiendo la exhaustividad de los diseños experimentales y la naturalidad de la observación sistemática.

**Entrenamiento de observadores en tiempos de pandemia. Un reto metodológico.** Triana Aguirre<sup>1</sup>, África Borges<sup>1</sup> y Aránzazu Calzadilla<sup>1</sup>

<sup>1</sup>Universidad de La Laguna

La metodología observacional es un procedimiento científico que pone de manifiesto la aparición de conductas directamente observables para proceder a su registro mediante un instrumento elaborado ad hoc. Se caracteriza por la perceptividad del comportamiento siendo adaptable y flexible a los comportamientos y a los contextos, lo que supone seguir un proceso disciplinado y riguroso. La fiabilidad de un instrumento de observación depende, en buena parte, de la competencia del observador. Asimismo, un

adiestramiento riguroso de las personas observadoras previene posibles fuentes de error, garantizando de esta forma la calidad del dato. El entrenamiento de observadores es entendido como un proceso de formación en el que adquieren el conocimiento conceptual, empírico y técnico necesario para desempeñar adecuadamente esta tarea. En el proceso de elaboración del Protocolo de Observación de la Mediación-Técnicas y Estrategias (PROME-TE) y, dada la situación de crisis sanitaria por la COVID-19, el entrenamiento del equipo de observación se llevó a cabo de forma completamente online, desarrollándose desde marzo a mayo del 2020. En esta comunicación se presenta la estructura del entrenamiento de observadores en un formato virtual lo que supuso un esfuerzo añadido a la ardua tarea de la observación, siendo clave el aspecto motivacional de las personas observadoras.

**Fiabilidad interobservadores del protocolo de observación de la mediación. Técnicas y estrategias (PROME-TE V.2).** Triana Aguirre<sup>1</sup>, África Borges<sup>1</sup> y Elena Rodríguez-Naveiras<sup>1</sup>

<sup>1</sup>Universidad de La Laguna

La mediación como método para resolver conflictos es una actividad profesional reciente que no tiene perfectamente sistematizadas las bases de la actuación profesional, más allá de los principios marcados por la literatura existente. Determinados aspectos, como la formación práctica en mediación, se encuentran aún en desarrollo y para conocer en profundidad el desempeño de esta actividad es necesario contemplar las habilidades y técnicas necesarias para ello. La persona que desempeña la labor mediadora puede contar con infinidad de habilidades innatas para ello, pero existe especial interés en aquellas que deben ser aprendidas y entrenadas. Ante la gran variabilidad de características, funciones, técnicas y estrategias que forman parte del ejercicio de la mediación y, por tanto, de la persona mediadora, es necesario entender que la calidad del proceso pasa ineludiblemente por la debida formación y cualificación de las personas mediadoras que han de llevarla a cabo. A pesar de la importancia de la formación práctica en mediación, en la actualidad no existen instrumentos que evalúen el comportamiento de la persona mediadora durante el proceso de mediación y, por ende, a las personas que se están formando en la materia. Por ello, el objetivo de este trabajo es comprobar la bondad psicométrica del Protocolo de Observación de la Mediación. Técnicas y estrategias en su versión definitiva. En cuanto a los resultados, tanto los índices de fiabilidad de Kappa de Cohen como los coeficientes de generalizabilidad relativo y absoluto superan cumplen con los criterios establecidos

previamente en el estudio. Se concluye que El PROME-TE v.2 cuenta con la adecuada bondad psicométrica en términos de fiabilidad interobservadores.

## SIMPOSIO

### Creación y Análisis de Pruebas Psicométricas

**Coordina:** Eduardo García-Cueto

**Resumen.** La medición adecuada de los constructos psicológicos es la base de gran parte de la investigación de la conducta humana. Es por esto que conocer cómo realizar una adecuada evaluación es de máxima importancia. En este sentido, este simposio va a analizar la evaluación a través de cuestionarios desde aspectos básicos de la creación de ítems hasta el uso práctico de los mismos. Por un lado, se va a exponer una investigación sobre el tipo de ítems más usados en psicología: el tipo Likert. Desde hace un tiempo existe un gran número de estudios que hablan de la posibilidad de que sea más preciso el uso de escalas continuas que las escalas graduadas tipo Likert. Con el primer trabajo se va a discutir las posibilidades que ofrecen estos dos tipos de respuesta y sus limitaciones. Seguidamente, se va a exponer los pasos dados para la creación de un instrumento para la medida de los usos del humor. Es reseñable la novedad de la investigación dada la práctica inexistencia de instrumentos en este campo en español, así como el nuevo marco teórico en el que el trabajo se inscribe. Otra de las investigaciones versa sobre el estudio de la dimensionalidad del cuestionario Dirty Dozen y la invarianza de medida en 8 regiones del mundo. Finalmente, se va a tratar de conocer el efecto de mediación del uso comparativo de las redes sociales entre el tiempo de uso y la ansiedad. Se llevó a cabo un modelo de mediación simple con sólo un mediador entre el antecedente y el consecuente utilizando un modelo de regresión lineal y se realizó el mismo modelo mediante ecuaciones estructurales.

**Ítems Tipo Likert versus Ítems Continuos.** Jaime García-Fernández<sup>1</sup>, Alvaro Postigo<sup>1</sup>, Covadonga García-Nuevo<sup>1</sup>, Marcelino Cuesta<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

Los ítems tipo Likert son comúnmente utilizados en investigaciones del ámbito de las ciencias sociales y de la salud. El presente estudio trata de comprobar si se pudiera, de alguna forma, mejorar las propiedades psicométricas de los autoinformes. Por tanto, no se trata de ver la utilidad (o no) del autoinforme, si no de ver si cambiando el formato de respuesta (de graduada a continua) hay una variación en las propiedades

psicométricas de las pruebas. Para ello, 537 personas (74% mujeres) con una edad media de 33.07 años (DT=13.56) respondieron a dos cuestionarios online en dos ocasiones, una con formato Likert (0-4) y otra con formato continuo (0-100). La discriminación, fiabilidad y pesos factoriales del formato Likert han sido mejores que las del formato continuo. Los coeficientes de congruencia indican que el formato de respuesta no afectó a la estructura factorial. Las escalas con ítems continuos dan puntuaciones significativamente más altas. Por tanto, estos resultados parecerían indicar que los ítems tipo Likert muestran un mejor funcionamiento que los continuos. Sin embargo, estos resultados podrían estar fatalmente ocasionados por el formato de presentación de la escala continua. El hecho de haber configurado el test en un formato electrónico forzó a que la longitud de la línea de respuesta fuera pequeña para así adaptarse a todo tipo de dispositivos. Por tanto, es posible pensar que influyó de forma decisiva sobre los resultados obtenidos.

**Escala CUHO: Usos del humor en población española.** Álvaro Menéndez-Aller<sup>1</sup>, Covadonga Gonzalez-Nuevo<sup>1</sup>, Jaime García-Fernández<sup>1</sup>, Eduardo García-Cueto<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

El estudio del humor se ha centrado en sus usos específicos. El instrumento más utilizado en investigación y/o evaluación de estos usos es el Humor Styles Questionnaire (HSQ), que propone cuatro usos del humor, dos positivos y dos negativos, en función de su impacto en el bienestar psicológico. El cuestionario ha manifestado ciertas incongruencias cuando se utiliza en población española. Esto es lo que ha propiciado que desde el grupo de investigación de Psicometría de la Universidad de Oviedo se haya iniciado la construcción de un nuevo instrumento más acorde con la cultura y la idiosincrasia de la población española. Se trata del Cuestionario de Usos del Humor de Oviedo (CUHO) basado en un nuevo modelo teórico que contempla siete usos diferentes del humor. Método. Con la finalidad de someter a prueba el citado modelo se creó un banco de ítems que fue examinado por un grupo de 12 expertos en evaluación psicológica para evaluar tanto el ajuste de los ítems a la definición de su uso como su representatividad. Seguidamente, la prueba se aplicó a una muestra piloto de 241 participantes (Medad = 39,89; DTedad = 14,60), de los cuales el 65,6% eran mujeres, para iniciar el estudio de las propiedades psicométricas de los ítems. Finalmente, se aplicó la prueba a una muestra de 527 participantes (Medad = 32,21; DTedad = 15,85), de los cuales 57,3% eran mujeres. Resultados. Se eliminaron 28 ítems por presentar inadecuadas propiedades en algunos de los parámetros psicométricos estimados. También se estudian las relaciones iniciales entre los Usos del Humor

y otras variables psicológicas. Conclusiones. El CUHO es una prueba que parece presentar adecuadas propiedades psicométricas para la evaluación de los usos del humor.

**Estudio de la Estructura Factorial del Dirty Dozen en Ocho Regiones del Mundo.** Álvaro Postigo<sup>1</sup>, Álvaro Menéndez-Aller<sup>1</sup>, Jaime García-Fernández<sup>1</sup>, Eduardo García-Cueto<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

Los rasgos de la triada oscura (narcisismo, psicopatía y maquiavelismo) recogen las diferencias individuales en la personalidad aversiva. La triada oscura ha mostrado relaciones importantes con conductas que afectan a la vida de las personas. Uno de los instrumentos más conocidos para evaluar la triada oscura es el Dirty Dozen. Sin embargo, sigue habiendo controversia acerca del uso de una puntuación general de triada oscura o, en cambio, de tres puntuaciones diferentes. El objetivo del presente trabajo fue el estudio de la estructura factorial del Dirty Dozen a través de ocho regiones del mundo. Se contó con 11.477 participantes pertenecientes de 49 países y agrupados en ocho regiones. Se estudiaron diferentes estructuras factoriales a través de Análisis Factoriales Confirmatorios. Tanto los modelos tridimensionales como los modelos bifactor (simétricos y no simétricos) mostraron un ajuste adecuado a los datos. El modelo bifactor no simétrico (contemplando la psicopatía como factor de referencia) mostró una mayor coherencia metodológica, apoyado por los diferentes índices bifactor. Referido al modelo bifactor no simétrico, se cumplió la invarianza configural y métrica, además de la invarianza escalar parcial. Los resultados indican que la psicopatía puede tener una posición superior y los elementos malignos del narcisismo y el maquiavelismo pueden ser esencialmente características subordinadas de la psicopatía. Se defiende una puntuación general de psicopatía y puntuaciones específicas de maquiavelismo y narcisismo, mostrando ser invariante entre las ocho regiones estudiadas.

**El uso comparativo de Redes Sociales como mediador entre el tiempo de uso y la ansiedad.** Covadonga González-Nuevo<sup>1</sup>, Álvaro Postigo<sup>1</sup>, Álvaro Menéndez-Aller<sup>1</sup>, Marcelino Cuesta<sup>1</sup>

<sup>1</sup>Universidad de Oviedo

Dentro del campo de la medición del uso problemático de Redes Sociales (RRSS) es de suma importancia conocer qué usos se relacionan más con la ansiedad para una mejor prevención. El objetivo de este trabajo es conocer si el uso comparativo de RRSS es un mediador de la relación del tiempo de uso de dichas redes con la ansiedad. Método: La muestra estuvo formada por 1003

participantes mayores de edad (M = 42.33; SD = 14.32). Se evaluó el tiempo de uso de RRSS, el uso comparativo de RRSS y el nivel de ansiedad. Se llevó a cabo un modelo de mediación simple con sólo un mediador entre el antecedente y el consecuente utilizando un modelo de regresión lineal y se realizó el mismo modelo mediante ecuaciones estructurales. Resultados: Los resultados obtenidos mostraron que el tiempo de uso de las redes sociales y el nivel de ansiedad se relacionan de manera directa e indirecta, a través del uso comparativo de RRSS. Se obtuvieron resultados similares mediante ambos métodos. Conclusiones: El uso comparativo de RRSS media la relación entre el tiempo de uso y la ansiedad.

## SIMPOSIO

### Validación del Constructo de Cognición Social

Coordina: África Borges del Rosal

**Resumen.** La cognición social es un constructo muy complejo y amplio, que ha sido definido como un conjunto de procesos que subyacen en las interacciones sociales, haciendo referencia al procesamiento mental de la información acerca del mundo social, procesos que actúan a varios niveles (neurobiológico, psicológico y social) y permiten percibir, reconocer y valorar estímulos sociales, con los que se construyen representaciones de uno mismo y de los demás, guiando el comportamiento social de los individuos. Dada la generalidad del término, se ha tratado de dividir y fragmentar la Cognición Social en procesos, habilidades o dominios, para hacer más fácil el estudio y la comprensión del constructo: a) las habilidades para comprenderse a sí mismo, englobando constructos como el autoconcepto y habilidades como la autoreflexión, que permiten tener un mejor conocimiento de la persona analizando las situaciones ya vividas y aprendiendo de ellas para próximos encuentros sociales; b) los procesos de autoregulación, como pueden ser al control de los impulsos o la expresión verbal de las emociones, que fomentan las interacciones sociales positivas ya que la persona ajusta su respuesta de acuerdo al tipo de contexto; c) habilidades sociales, entendidas como todas aquellas que permiten conectar emocionalmente con otros y negociar en las situaciones percibidas como injustas; y d) habilidades para entender a los demás, centradas en la comprensión, por un lado emocional y por otro cognitiva de las otras personas. Aunque con cierta permeabilidad en la distinción, las primeras engloban constructos como la empatía y las segundas otros, como la Teoría de la Mente (ToM por sus siglas en inglés). Dado que el constructo de cognición social está aún definiéndose, la validación de constructo es de gran relevancia, enfocándolo desde diversas investigaciones para establecer las relaciones de las



variables que se suponen sustentan el constructo. En el presente simposio se presentan cuatro trabajos: la relación entre empatía y neuronas espejo, así como la ToM medida con dos instrumentos, y su relación con el género, la empatía y la función ejecutiva.

#### **Diferencias de género en teoría de la mente en adultos.**

Andrea García Casanova<sup>1</sup>, Lorena Fumero Negrín<sup>1</sup>, Marina Gómez Madrid<sup>1</sup> y África Borges<sup>1</sup>

<sup>1</sup>Universidad de La Laguna

La Cognición Social es un constructo de estudio relativamente reciente en Psicología y neurociencias, cuyo interés reside permite percibir, reconocer y evaluar elementos sociales con los cuales construir representaciones, tanto de uno mismo como de los demás y que serán usadas posteriormente para guiar el comportamiento social de la persona. Dentro de los aspectos interpersonales, siendo una de las variables más relevantes la Teoría de la Mente (ToM). El objetivo de la presente investigación es analizar las diferencias que existen entre el género en ToM. Se ha contado con una muestra de 179 personas, seleccionadas mediante un muestreo de conveniencia. El instrumento que se ha utilizado ha sido el Test de la Miradas y la recogida de datos se ha realizado a través de una plataforma online creada ad hoc para ellos. La comparación entre géneros se ha hecho mediante el estadístico t de Student, observándose ausencia de diferencias. Se concluye la relevancia de estas conclusiones en el entorno de la igualdad de género.

#### **Cognición social: Relación entre funciones ejecutivas y teoría de la mente.** Raúl González García<sup>1</sup> y África Borges<sup>1</sup>

<sup>1</sup>Universidad de La Laguna

La cognición social es un conjunto de procesos complejos que modulan nuestra forma de pensar sobre los demás, además de reconocer y evaluar elementos sociales, siendo una herramienta clave para comprender las relaciones interpersonales. Entre los numerosos subdominios que componen este constructo destacan las funciones ejecutivas y la teoría de la mente, ambos procesos cognitivos que se complementan y construyen un espacio para comprender el pensamiento ajeno y actuar en consecuencia de forma adecuada. El objetivo de esta investigación es precisamente analizar la relación que subyace a ambos, empleando para ello una muestra de 145 participantes seleccionados mediante muestreo de conveniencia con edades comprendidas entre los 18 y los 65 años. Para evaluar ambos subdominios, se pasaron las pruebas DEX y Test de las Miradas para medir las

funciones ejecutivas y la teoría de la mente respectivamente. La metodología empleada fue metodología de encuestas, diseño transversal. Los cuestionarios se implementaron en formato electrónico a través de una plataforma creada ad hoc para la presentación y recogida de datos. Se realizó una regresión múltiple, siendo la variable dependiente la puntuación en el test de las miradas y como variables independientes las escalas del DEX. Se muestran las relaciones existentes entre estas variables, concluyendo la relevancia de ampliar el conocimiento sobre las demás partes que componen la base de la cognición social.

#### **Las neuronas espejo y la empatía.** Nicole Privato<sup>1</sup>, Julio Manuel Plata Bello<sup>1</sup> y África Borges<sup>1</sup>

<sup>1</sup>Universidad de La Laguna

A partir de su descubrimiento las neuronas espejo han proporcionado una base neuronal para diferentes funciones sociales como la capacidad de entender las acciones de los demás, la imitación, la habilidad de prever las intenciones del otro y la empatía. El objetivo de esta investigación es estudiar la relación entre los niveles de empatía de los participantes, relevados por el cuestionario IRI, y la activación de las zonas cerebrales donde están las neuronas espejo mientras se observan imágenes de personas con y sin dolor por medio de resonancia magnética funcional (RMNf). Las imágenes utilizadas fueron un total de doce, seis representando personas con expresiones de dolor y seis retrayendo las mismas personas con expresión neutra. Cada imagen aparece durante 5 segundos en tres bloques de tres imágenes siguiendo el siguiente esquema: imagen con dolor-resting-imagen sin dolor. Los resultados obtenidos indican que la mayor activación obtenida fue en el lóbulo parietal inferior izquierdo ( $FWE < 0.05$ ); en el lóbulo occipital, en específico en el giro occipital medio izquierdo y en el giro lingual izquierdo; finalmente, en el lóbulo temporal izquierdo se ha encontrado activación significativa en el giro temporal superior. En el lóbulo derecho se ha encontrado una activación superior solo en el giro temporal medio derecho. El estudio de regresión múltiple entre los resultados obtenidos en el IRI y las imágenes obtenidas por la resonancia magnética indican una relación positiva entre las puntuaciones en la subescala "preocupación empática" y el lóbulo superior frontal medio y las regiones parietotemporales, ambas zonas cerebrales ricas en neuronas espejo. Así, a mayor puntuación obtenida en la escala "preocupación empática" mayor activación cerebral en las regiones espejo al observar imágenes de personas con dolor.

**Diferencias de género en teoría de la mente en adolescentes.** Ricardo Quintero Rodríguez<sup>1</sup>, Triana Aguirre<sup>1</sup> y África Borges<sup>1</sup>

<sup>1</sup>Universidad de La Laguna

El reconocimiento de sentidos no literales se puede considerar como uno de los niveles más complejos de la Teoría de la Mente. Se conceptualiza como la capacidad para atribuir intencionalidad a partir de comunicaciones metafóricas. Comúnmente, se ha estudiado el reconocimiento de sentidos no literales en niños y adolescentes diagnosticados con diferentes patologías. Sin embargo, las investigaciones en población típica han sido poco tratadas en la literatura clásica y la investigación. La literatura pone en manifiesto que al género como una variable relevante a la hora de analizar las diferencias interindividuales en la Teoría de la Mente. El presente estudio busca identificar la existencia de diferencias de género en el reconocimiento de sentidos no literales en adolescentes. Participaron 509 adolescentes, 259 mujeres y 250 hombres, entre 12 y 18 años, pertenecientes a diferentes instituciones educativas de las Islas Canarias. Se administró la prueba de Historias Extrañas de Happé (1994, versión española). Los resultados permitirán profundizar en el estudio de los procesos cognitivos subyacentes al funcionamiento social. Se debate la relevancia de estos aspectos interindividuales entre el alumnado.

## SIMPOSIO

**Investigando Metodología en el Multiverso de la Academia: Comunidad, Recursos y Estrategias por y para Jóvenes**

**Coordinan:** Ricardo Rey-Sáez y Pablo Nájera

**Resumen.** Los primeros pasos dentro del mundo académico fuerzan a aquellos estudiantes que quieren dedicar su vida a la ciencia a convivir con situaciones para las que no hay una formación específica. Emociones como la incertidumbre, la soledad o la frustración pueden acompañarnos mientras adquirimos una serie de competencias profesionales concretas: búsqueda y resumen de información, uso de software y programación, establecer relaciones entre teoría y modelado estadístico, y la síntesis, redacción y transmisión de la información. Estas habilidades, estrictamente vinculadas a nuestra labor científica, se tendrán que ver forzosamente complementadas con la adquisición de competencias burocráticas: contratos predoctorales, estancias en el extranjero y ayudas económicas que dependen de un sinfín de documentos

que deben ser presentados en un estricto plazo de tiempo. El compromiso con la ciencia acarrea un número muy elevado de responsabilidades y decisiones; pero, si bien el camino no es fácil, no tenemos por qué recorrerlo solos. En base a esta idea, el Grupo de Jóvenes Investigadores en Metodología (GJIM) pretende compartir en este simposio una serie de recursos y estrategias elaboradas por y para jóvenes, con el objetivo de allanar los primeros pasos del camino. Generar espacios de comunicación fluida y formar parte de una red de personas con objetivos e intereses comunes puede facilitar que nuestros barcos (y los de otras personas) lleguen a buen puerto.

**Investigación, paso 1: ¿Cómo y dónde busco información? Nuevas herramientas para la detección eficiente de bibliografía de interés.** Ricardo Rey-Sáez<sup>1</sup>

<sup>1</sup>Universidad Complutense de Madrid

Todo trabajo académico parte del mismo punto: una revisión de la literatura previa. Aunque este paso es la piedra angular sobre la que se cimienta el marco teórico de nuestros trabajos, las primeras aproximaciones a nuestro objeto de estudio suelen ser poco formales (pudiendo escucharlo en la universidad, en un curso, en una ponencia en un congreso, en una charla distendida con otros investigadores, con amigos...). Es tras esta primera toma de contacto cuando iniciamos la búsqueda de artículos para documentarnos y aprender, al mismo tiempo que surgen nuestras primeras hipótesis concretas. No obstante, el cómo y el dónde buscar información puede volverse un reto mayúsculo, sobre todo cuando queremos invertir el menor número posible de recursos. Algunas de las estrategias habituales pasan por aplicar búsquedas por booleanos en numerosas bases de datos con varios filtros, consultar determinadas revistas o introducir el título de una publicación (habitualmente, citadas en trabajos relacionados que ya hemos leído) en distintos buscadores. Si bien este "modus operandi" es conocido por todos, existen alternativas recientes que simplifican en gran medida este proceso, pudiendo obtener en pocos segundos un screening global de muchos artículos de interés (y de la relación entre ellos). El objetivo de esta comunicación oral es presentar varios de estos recursos y mostrar algunas de sus características: su funcionamiento intuitivo, su forma de resumir la información (a distintos niveles) y su facilidad para complementar los métodos tradicionales, representando una nueva estrategia que, en el futuro, permitirá búsquedas aún más exhaustivas y eficientes para estudiantes, docentes e investigadores.

**Trucos y consejos para simular en R.** Pablo Nájera<sup>1</sup>, Rodrigo Schames Kreitchmann<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Desde su creación en 1993, el software estadístico R se ha consolidado como un referente para la investigación en ciencias sociales gracias a la inmensa variedad de procedimientos estadísticos que contiene y la flexibilidad que otorga a la hora de diseñar métodos nuevos. Entre las múltiples funcionalidades que ofrece, R permite realizar estudios de simulación. Los estudios de simulación constituyen una herramienta muy atractiva para la investigación en metodología, puesto que permiten evaluar de forma sistemática el rendimiento de distintos procedimientos en función de las condiciones especificadas, permitiendo así comparar y estudiar la pertinencia de los métodos para su posterior aplicación en situaciones reales. Aunque las simulaciones son especialmente interesantes para la investigación teórica (p.ej., evaluación del funcionamiento de estadísticos o técnicas psicométricas), también pueden ser útiles para predecir el rendimiento de procedimientos estadísticos en estudios aplicados mediante diseños post hoc (i.e., simulaciones en las que se emplean las condiciones específicas de la aplicación real). Esta presentación tiene el fin de introducir y facilitar la realización de estudios de simulación con R a aquellas personas que no estén muy familiarizadas con esta metodología y/o con el software estadístico. Así, se empleará una perspectiva aplicada en la que se mostrará el esqueleto básico que vertebrará un script de simulación en R y se compartirán distintos trucos y consejos que pueden facilitar la organización, eficiencia y velocidad de nuestras simulaciones. Las contribuciones por parte de las personas asistentes y el intercambio de experiencias y recursos en relación con los estudios de simulación en metodología serán totalmente bienvenidas.

**Investigación psicológica mediante Machine Learning: recursos para que los primeros pasos se den en la dirección adecuada.** Diego Iglesias<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Observado un fenómeno de interés psicológico, el modelado estadístico permite estimar ecuaciones matemáticas (i.e. modelos) que tratan de, o bien predecir, o bien comprender dicho fenómeno a partir de un conjunto de posibles variables explicativas. Dado el amplio bagaje de algoritmos que pueden utilizarse en el proceso de estimación, todo modelo estadístico puede entenderse dentro del continuo de la interpretabilidad-complejidad. Generalmente, cuanto mayor es la complejidad del algoritmo subyacente, menor es la interpretabilidad del modelo al que da lugar. Si bien la

investigación en Psicología ha centrado su interés en modelos explicativos e interpretables (i.e. lineales), la incorporación de modelos de mayor complejidad desarrollados en las ciencias de la computación mediante machine learning (p.ej. redes neuronales, árboles de decisión, etc.) supone una contribución de gran relevancia para la Psicología, mejorando su capacidad predictiva y el nivel de formalización de los constructos teóricos que propone. Adentrarse en la investigación psicológica mediante machine learning requiere conocimientos matemáticos y de programación específicos. No saber dónde ni cómo adquirir estos conocimientos es una situación a la que se enfrenta cualquier investigador/a novel. El objetivo de esta ponencia es presentar una serie de recursos (p.ej. manuales y cursos especializados accesibles) y una metodología de trabajo con el propósito de hacer lo más llevadero y provechoso posible el proceso de inmersión en el mundo del machine learning.

**Ya hice la ciencia, ¿ahora cómo la cuento?: desde los congresos científicos hasta las redes sociales.** Alicia Franco-Martínez<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Cuando elaboramos nuestras investigaciones, sobre todo en el campo de la metodología, tendemos a adentrarnos en mundos de modelos y propuestas muy específicas, con las que pasamos meses y años trabajando y para los que acumulamos cantidades ingentes de información. Plasmar todo ese conocimiento en un manuscrito es una tarea muy exigente, centrada principalmente en convencer a editores y revisores de que nuestro trabajo merece la pena. Pero, ¿qué pasa más allá de las JCR? Los congresos, como puntos de encuentro científicos, con el tiempo han ido dando forma al formato tradicional de ponencias que hoy conocemos, definida por unas características muy particulares (diapositivas de PowerPoint, inicios bajo la fórmula “buenos días, mi nombre es [inserte nombre] y mi presentación versa sobre [inserte título completo]”, estructura similar a la de los papers...). En este breve taller reflexionaremos sobre si esta modalidad de presentación está cambiando en los últimos años, sobre si debería hacerlo y en qué sentido. Además, compartiremos consejos y recursos que pueden servir de apoyo creativo para elaborar futuras ponencias. Pero, ¿qué pasa más allá de los congresos? Negar la potencia de difusión científica que suponen los espacios virtuales, principalmente las redes sociales, sería negar una realidad cada día más evidente. En estos entornos, las citas se transforman en likes y retuits, mientras se abren foros para discutir el contenido científico, preguntar

dudas y compartir inquietudes. En este simposio hablaremos de cómo adaptar el trabajo académico a, por ejemplo, un hilo de Twitter; de qué tonos funcionan mejor en estos contextos, y de la extensa variedad de recursos audiovisuales que ofrecen.

**“La ciencia no se hace sola”: Primeros pasos en la investigación y la relevancia de los grupos de apoyo.**  
Cristina Rodríguez-Prada<sup>1,2</sup>

<sup>1</sup>Universitat Pompeu Fabra, <sup>2</sup>Universidad Autónoma de Madrid

Las palabras de un sabio tuitero decían así: “la ciencia no se hace sola, ahí que acerla” (sic). Tenía razón porque, primero, la ciencia es una actividad humana; segundo, porque es un trabajo fundamentalmente colaborativo (“a hombros de gigantes”). Las primeras etapas en el mundo de la investigación suelen ir acompañadas, junto con aspectos muy positivos, de circunstancias no pocas veces hostiles. Por un lado, nos introducimos en un contexto altamente competitivo, con pocos recursos dedicados desde las instituciones que nos inicia en un proceso de enculturación sobre el funcionamiento del engranaje y la violencia burocrática. La preparación de documentos probatorios, la presentación a un sinfín de convocatorias con escasos números de plazas con el objetivo de obtener financiación para el trabajo que realizamos y las malas praxis que pueden venir asociadas pueden ser una fuente importante de malestar, junto con las condiciones precarias en términos laborales y de expectativas de futuro. Por otro lado, la urgencia de producción científica durante los primeros años, la sobrecarga laboral, la confusión de roles y las fechas límites; todo ello sumado a la soledad que suele caracterizar el desarrollo de las investigaciones, especialmente en un contexto de teletrabajo. ¿Cómo podemos hacer que estas aguas sean más navegables? En este espacio nos centraremos en la necesidad de organización colectiva como herramienta para (1) la mejora de las condiciones laborales en los primeros pasos en el mundo académico; (2) la mejora del bienestar de las personas que nos iniciamos en investigación a través de compartir nuestras experiencias y recursos ante las problemáticas que, aunque inicialmente puedan percibirse como individuales, no lo son. De esta manera, enfatizamos la necesidad de establecer redes sociales de apoyo para una mejor estancia en el contexto académico.

## SIMPOSIO

### Diseños de Medidas Repetidas: Del Modelo Lineal General al Modelo Lineal Mixto

Coordinan: Paula Fernández y María J. Blanca

**Resumen.** El diseño de medidas repetidas es uno de los diseños más utilizados en ciencias sociales y de la salud. Entre las alternativas de análisis de datos, cuando la variable dependiente es cuantitativa, se encuentran el Modelo Lineal General (MLG) y el Modelo Lineal Mixto (MLM). Dentro del primer enfoque, el análisis de varianza (ANOVA) de medidas repetidas sigue siendo el más prevalente en la literatura científica aplicada. El ANOVA asume independencia, normalidad, y esfericidad. Conocer las consecuencias de la violación de estos supuestos es primordial para establecer el grado en que las inferencias derivadas del análisis son correctas. Existe un notable número de estudios de simulación Monte Carlo para analizar la robustez del ANOVA. Sin embargo, los resultados no son consistentes y no arrojan estrategias claras para guiar a los investigadores aplicados en el uso adecuado de estos modelos. Por ello, el primer y segundo trabajo se centra en analizar las consecuencias de la violación de la normalidad y esfericidad, respectivamente. Por otro lado, el MLM permite modelizar la estructura de la matriz de covarianza por lo que puede resultar más adecuado cuando la esfericidad no se satisface. Sin embargo, la principal dificultad de este procedimiento radica precisamente en la selección correcta de esta estructura. El tercer trabajo se centra en analizar la eficacia de los Criterios de Información para seleccionar la matriz de covarianza que se adecúa a los datos, y la robustez del MLM. Ambos se analizan en condiciones adversas de heterogeneidad de matrices de covarianza y de pérdida de datos MNAR. Finalmente, el objetivo del cuarto trabajo es examinar el comportamiento del MLM junto a otras alternativas univariadas cuando a los datos subyace una matriz autorregresiva estacionaria, condición muy habitual en investigaciones aplicadas. Se ponen de manifiesto algunas fortalezas y algunas debilidades de los procedimientos estadísticos analizados.

**Violación de la normalidad en el ANOVA de medidas repetidas.** María J. Blanca<sup>1</sup>, Jaume Arnau<sup>2</sup>, F. Javier García-Castro<sup>1</sup>, Rafael Alarcón<sup>1</sup>, Roser Bono<sup>2</sup>

<sup>1</sup>Universidad de Málaga, <sup>2</sup>Universidad de Barcelona

El diseño de medidas repetidas es uno de los diseños más utilizados en ciencias sociales y de la salud. Aunque hay otras alternativas de análisis más avanzado, la realidad es

que el estadístico F del ANOVA de medidas repetidas sigue siendo el procedimiento estadístico más empleado para analizar las diferencias de medias. Este análisis requiere normalidad, esfericidad, e independencia entre los sujetos. El impacto de la violación de la normalidad ha recibido más atención en el ANOVA intersujeto en comparación con el de medidas repetidas. En este último, la mayoría de los estudios se centran en analizar la influencia de la violación de la esfericidad en el error de Tipo I o en la potencia estadística, por lo que son escasos los estudios que analicen de forma extensiva la robustez del estadístico F bajo violación de la normalidad. El objetivo de este trabajo es realizar un estudio de simulación Monte Carlo para analizar si se altera el error de Tipo I y la potencia estadística del estadístico F en un diseño unifactorial de medidas repetidas. Se manipulan el tamaño muestral, incluyendo tamaños pequeños, medianos y grandes, y distintas formas de distribución de la variable dependiente. Estas distribuciones representan desviación leve, moderada y grave con respecto a la distribución normal. Los resultados muestran que tanto el error Tipo I como la potencia del estadístico F no se alteran con la violación de la normalidad. Futuros estudios deberán ampliar las variables manipuladas para determinar la generalización de estos resultados a otras condiciones. Agradecimientos: este estudio es parte del proyecto de I+D+i PID2020-113191GB-I00, financiado por MCIN/AEI/10.13039/501100011033/.

**Violación de la esfericidad en el ANOVA de medidas repetidas.** María J. Blanca<sup>1</sup>, Jaume Arnau<sup>2</sup>, F. Javier García-Castro<sup>1</sup>, Rafael Alarcón<sup>1</sup>, Roser Bono<sup>2</sup>

<sup>1</sup>Universidad de Malaga, <sup>2</sup>Universidad de Barcelona

El ANOVA unifactorial de medidas repetidas asume independencia, normalidad, y esfericidad. Este último se refiere a la igualdad de las varianzas de las diferencias en la variable dependiente entre los niveles del factor intrasujeto. La evidencia previa muestra que la violación de este supuesto tiene un gran impacto en el error Tipo I. Aunque existen otros procedimientos alternativos más avanzados, una opción recomendada en la mayoría de los textos clásicos es el uso de las pruebas F ajustadas cuando la esfericidad se viola. La realidad es que estas pruebas son frecuentemente empleadas porque son intuitivas, fáciles de aplicar y por su disponibilidad en la mayoría de los programas estadísticos de análisis de datos. Las pruebas F ajustadas más utilizadas son las conocidas como Greenhouse-Geisser (F-GG) y Huynh-Feldt (F-HF), que varían dependiendo del procedimiento de estimación del factor correctivo  $\epsilon$ . Existe un notable número de estudios de simulación Monte Carlo para analizar la robustez del estadístico F y de las pruebas F ajustadas. Sin

embargo, los resultados no son consistentes y de ellos no se pueden derivar estrategias claras para seguir en el análisis de datos. Esta inconsistencia en los hallazgos encontrados muestra la necesidad de realizar nuevos estudios que puedan proporcionar directrices claras a los investigadores. El objetivo es analizar el impacto de la violación de la esfericidad en el error de Tipo I asociado a los estadísticos F, F-GG y F-HF en un diseño unifactorial de medidas repetidas, con distintos tamaños muestrales, y con violaciones leves, moderadas y graves del supuesto. Los resultados muestran que el estadístico F es liberal ante violaciones moderadas y severas, independientemente del tamaño muestral. Asimismo, se exponen estrategias analíticas a seguir ante la violación de la esfericidad. Agradecimientos: este estudio es parte del proyecto de I+D+i PID2020-113191GB-I00, financiado por MCIN/AEI/10.13039/501100011033/.

**La influencia de la topografía de la heterogeneidad de las matrices de covarianza en el análisis de los diseños de medidas repetidas mediante el MLM cuando se ha producido una pérdida de datos MNAR.** M. Paula Fernández<sup>1</sup>, Guillermo Vallejo<sup>1</sup>, Pablo Livácic-Rojas<sup>2</sup>, Feliciano Ordóñez<sup>3,4</sup> y Ellián Tuero-Herrero<sup>1</sup>

<sup>1</sup>Universidad de Oviedo, <sup>2</sup>Universidad de Santiago de Chile

<sup>3</sup>Facultad Padre Ossó de Oviedo, <sup>4</sup>Universidad Internacional de la Rioja

Cuando la variable dependiente es métrica, el MLM permite, además, tener visibilidad del resultado del análisis de los diseños de medidas repetidas desde la óptica de todos los enfoques contemplados por el MLG. El investigador podrá tomar, de este modo, una decisión sobre qué resultado es el más ajustado a la teoría, y también a sus datos, y podrá excusar las amenazas a la validez de la conclusión estadística, de haberlas. En la investigación aplicada es habitual que el tamaño de los grupos de sujetos que se comparan en su comportamiento a través del tiempo esté desproporcionado y, por esa razón, y por otras también, es habitual que las matrices de covarianza sean heterogéneas. Es habitual también tener pérdida de datos. En el presente trabajo se muestran los resultados de dos estudios en los que se ha examinado el efecto que tiene la heterogeneidad de las matrices de covarianza (en razón positiva y negativa, y ambas cuando el grado de variación del tamaño de los grupos es media y grave) sobre la eficacia de los Criterios de Información para seleccionar la matriz de covarianza que se adecúa a los datos, y, con independencia de ello, sobre la robustez que tiene el MLM para identificar correctamente las fuentes de variación cuando se produce pérdida de datos MNAR. Agradecimientos: esta investigación ha sido financiada por el Proyecto PGC2018-101574-B-I00 (AEI/FEDER, UE),

concedido por el Ministerio Español de Ciencia, Innovación y Universidades.

**El análisis de datos de un diseño de medidas repetidas para el investigador aplicado y para el investigador metodológico.** M. Paula Fernández<sup>1</sup>, Guillermo Vallejo<sup>1</sup>, Feliciano Ordóñez<sup>2,3</sup>, Pablo Livácic-Rojas<sup>4</sup>, Ellián Tuero-Herrero<sup>1</sup>

<sup>1</sup>Universidad de Oviedo, <sup>2</sup>Facultad Padre Ossó de Oviedo,

<sup>3</sup>Universidad Internacional de la Rioja, <sup>4</sup>Universidad de Santiago de Chile

La investigación mediante simulación Monte Carlo es el modo de conocer el comportamiento de las técnicas estadísticas, sean cuales sean las condiciones contextuales de distribución de los datos, de características de diseño, y del proceso o estado observado de una o varias variables dependientes. Poniendo el foco en los diseños de medidas repetidas y asumiendo que la variable dependiente es métrica, el MLG y el MLM son las opciones disponibles. Se realiza una investigación de simulación Monte Carlo para examinar la robustez del MLM junto a otras alternativas univariadas cuando a los datos de un diseño de medidas parcialmente repetidas subyace una matriz autorregresiva estacionaria con distintos de grado de autocorrelación positiva y negativa. A continuación, se muestra el resultado de investigaciones cuasireales generados con las mismas características que las simuladas, y serán analizadas desde la óptica de un investigador aplicado y de un investigador metodológico. Ambos, el investigador aplicado y el investigador metodológico se van a encontrar en alguna disyuntiva, y ambos tendrán que tomar una decisión razonada sobre el resultado que finalmente deciden publicar. Agradecimientos: esta investigación ha sido financiada por el Proyecto PGC2018-101574-B-I00 (AEI/FEDER, UE), concedido por el Ministerio Español de Ciencia, Innovación y Universidades.

## SIMPOSIO

### Meta-Análisis de Generalización de la Fiabilidad

**Coordina:** Julio Sánchez Meca

**Resumen.** Desde que en 1998 Tami Vacha-Haase acuñara el término 'reliability generalization' para referirse a los meta-análisis de generalización de la fiabilidad (GF), este tipo de meta-análisis psicométrico se ha desarrollado de forma prolífica en el ámbito de las Ciencias del Comportamiento y de la Salud. Un meta-análisis de GF tiene por objeto investigar cómo varían los coeficientes de fiabilidad que se obtienen al aplicar un determinado

test en diferentes muestras de participantes y diferentes contextos de aplicación. Se presentan en este simposio los resultados de cinco investigaciones realizadas por el equipo de la Unidad de Meta-análisis de la Universidad de Murcia, en colaboración con investigadores de otras universidades españolas y extranjeras, en las que se ofrecen resultados y técnicas de análisis novedosos, como son los modelos multinivel y la aplicación de los modelos de ecuaciones estructurales en el contexto meta-analítico (MASEM). Las dos primeras comunicaciones, que corren a cargo de Alejandro Veas Iniesta (Universidad de Murcia) y Raimundo Aguayo (UNED), presentan dos aplicaciones del enfoque MASEM a sendos tests psicológicos y se comparan los resultados con los de otros enfoques estándar. La tercera comunicación, presentada por Carmen López Ibáñez (Universidad de Murcia), presenta un estudio comparativo del uso de los modelos multinivel y el enfoque estándar en los meta-análisis de GF. La cuarta comunicación, presentada por Alejandro Sandoval Lentisco (Universidad de Murcia), presenta los resultados de un meta-análisis de GF con análisis estadísticos novedosos. Y la última comunicación, presentada por Julio Sánchez Meca (Universidad de Murcia), presenta los resultados de un estudio sobre las tasas de inducción de la fiabilidad en el ámbito de las investigaciones psicológicas.

**Meta-análisis de Generalización de Fiabilidad de la Escala Fear of COVID-19: Una comparación entre la aproximación univariada y MASEM.** Desirée M<sup>a</sup> Blázquez-Rincón<sup>1</sup>, Raimundo Aguayo-Estremera<sup>2</sup>, Zainab Alimoradi<sup>3</sup>, Elahe Jafari<sup>3</sup> y Amir H. Pakpour<sup>4</sup>

<sup>1</sup>Universidad de Murcia, <sup>2</sup>Universidad Complutense de Madrid, <sup>3</sup>Qazvin University of Medical Sciences, <sup>4</sup>Jönköping University, Jönköping

**Objetivo:** El meta-análisis de generalización de la fiabilidad (GF) se ha realizado tradicionalmente sintetizando coeficientes de fiabilidad. Este enfoque univariado tiene la desventaja de que los coeficientes reportados (comúnmente el alfa de Cronbach) pueden no ser el índice de fiabilidad que mejor refleja la estructura interna de la escala. Para superar esta dificultad, recientemente se ha propuesto utilizar el enfoque meta-analytic structural equation modeling (MASEM). El objetivo del presente trabajo es usar dos técnicas MASEM para estudiar la fiabilidad de la Escala Fear of COVID-19 (FCV-19S), y comparar los resultados con los del enfoque univariado. **Método/Diseño:** Se aplicaron dos técnicas diferentes (One-Stage and Two-Stage MASEM) para estudiar la estructura interna de la FCV-19S. A partir del modelo que mejor reflejaba la estructura interna de la escala (un factor congénico, un factor tau-equivalente

o dos factores), se eligió un índice de fiabilidad y se calculó su promedio junto con el intervalo de confianza. Por último, se llevaron a cabo análisis moderadores para las correlaciones promediadas. Resultados: Se obtuvo una matriz de correlaciones de Pearson para los 7 ítems de la FCV-19S en 43 estudios independientes. Se computó una matriz combinada siguiendo un modelo de efectos aleatorios, dado que había evidencia de heterogeneidad entre estudios para las correlaciones sintetizadas. De los tres modelos, el modelo de dos factores tuvo un ajuste estadísticamente mejor, aunque los tres alcanzaron índices de ajuste razonablemente buenos. Los resultados fueron equivalentes con las dos técnicas MASEM utilizadas. Conclusiones: A pesar de las ventajas de estas técnicas MASEM, un inconveniente consiste en que el análisis de moderadores se realiza sobre las correlaciones (Two-Stage MASEM) o las cargas factoriales (One-Stage MASEM), a diferencia del enfoque univariado en el que se analizan directamente los índices de fiabilidad. En nuestra opinión, estas dos técnicas MASEM y el GF univariado podrían utilizarse como enfoques complementarios.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**Meta-análisis de ecuaciones estructurales (MASEM): Principios y aplicaciones para la mejora de los procesos de generalización de resultados en test y escalas de medida.** Alejandro Veas<sup>1</sup>, José Antonio López-López<sup>1</sup>, Rubén López-Nicolás<sup>1</sup>, Julio Sánchez-Meca<sup>1</sup>

<sup>1</sup>Universidad de Murcia

En los últimos años se ha mostrado evidencia de ciertas limitaciones de los métodos tradicionales aplicados a los meta-análisis de generalización de la fiabilidad, tales como la ausencia de equivalencia entre la fiabilidad total y las subescalas o el incumplimiento del principio de independencia de errores. En este sentido, se han desarrollado métodos estadísticos multivariados que permiten una estimación más eficaz de los instrumentos de medida, tales como el meta-análisis de ecuaciones estructurales (MASEM), cuyas ventajas residen en la posibilidad de combinar matrices de correlaciones de distintos estudios, estimar y comprobar el ajuste de modelos factoriales en la matriz media de correlaciones, así como la estimación de modelos aleatorios. La presente comunicación muestra la utilidad de esta técnica a partir de su aplicación a una escala validada recientemente en España, la School Attitude Assessment Survey-Revised. Para ello, se describen los pasos empleadas para la estimación de parámetros y la comprobación de ajuste del modelo en la matriz de correlaciones sintetizada,

empleando el software estadístico R. Por último, se discuten las implicaciones de los resultados obtenidos.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**El metaanálisis de generalización de la fiabilidad a través de la perspectiva multinivel.** Carmen López-Ibáñez<sup>1</sup>, Wim Van Den Noortgate<sup>2</sup>, Julio Sánchez-Meca<sup>1</sup>

<sup>1</sup>Universidad de Murcia, <sup>2</sup>Katholieke Universiteit Leuven, Kortrijk

El metaanálisis de generalización de la fiabilidad (GF) se ha revelado como una buena herramienta para la síntesis de evidencia sobre las propiedades psicométricas de los instrumentos de medida en Psicología. Cada vez es más común que los estudios primarios reporten los resultados psicométricos de sus propios datos en lugar de inducirlos de estudios previos o de los estudios originales de dichos instrumentos, lo que permite que el metaanálisis estime estas propiedades de forma más precisa. Sin embargo, una de las limitaciones que sigue teniendo esta técnica es la forma de lidiar con la dependencia dentro del metaanálisis. Tradicionalmente, cuando una escala presenta una estructura multidimensional con varias subescalas, el análisis de la fiabilidad se ha realizado por cada subescala, haciendo metaanálisis independientes para cada una de ellas. No obstante, esas subescalas forman parte de un mismo cuestionario y, en definitiva, de un mismo constructo, por lo que no es de extrañar que surjan relaciones de dependencia entre ellas. Separar cada subescala en un metaanálisis independiente elimina esa posible dependencia, con la consiguiente pérdida de información. Es por ello por lo que parece necesaria la implantación de alguna técnica que ayude en el manejo de dicha dependencia. El enfoque multinivel podría ser una buena alternativa. Este estudio tiene como objetivo aplicar el enfoque multinivel y la estrategia tradicional para el cálculo de metaanálisis de generalización de la fiabilidad en 4 escalas que ya han sido objeto de metaanálisis GF, comprobando así el comportamiento de los resultados en términos de coeficiente promedio y su intervalo de confianza. Dichas diferencias se observaron a través de un índice de discrepancia que informaba sobre el porcentaje de cambio que se producía entre las comparaciones.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**Meta-análisis de Generalización de la Fiabilidad de las escalas FOCI y C-FOCI.** Alejandro Sandoval-Lentisco<sup>1</sup>, Rubén López-Nicolás<sup>1</sup>, José A. López-López<sup>1</sup>, Julio Sánchez-Meca<sup>1</sup>

<sup>1</sup>Universidad de Murcia

Entre la multitud de escalas que existen para evaluar la sintomatología obsesivo-compulsiva el Florida Obsessive Compulsive Inventory (FOCI) y su versión para niños, el Child-Florida Obsessive Compulsive Inventory (C-FOCI) destacan por ser las únicas escalas breves y auto-administradas que permiten evaluar de forma simultánea síntomas y severidad. En este estudio llevamos a cabo un meta-análisis de generalización de la fiabilidad para estimar una fiabilidad media e identificar las características de los estudios que explicarán la heterogeneidad. Para las dos subescalas de la FOCI y C-FOCI, encontramos fiabilidades que iban desde aceptables a buenas (FOCI Symptom Checklist KR-20  $\alpha+$  = .826 y FOCI Symptom Severity  $\alpha+$  = .882; C-FOCI Symptom Checklist KR-20  $\alpha+$  = .740 y FOCI Symptom Severity  $\alpha+$  = .794). Para el análisis de moderadores, primero aplicamos modelos mixtos de meta-regresión y análisis de subgrupos para las variables continuas y categóricas, respectivamente. Con los moderadores que fueron estadísticamente significativos o que explicaron más de un 10% de la varianza, realizamos modelos de moderadores incluyendo todas las posibles combinaciones de estos. Para la comparación de modelos, hicimos uso del Akaike Information Criteria (AIC). Notablemente, encontramos que, para la subescala Symptom Severity de la FOCI, los modelos mejor valorados fueron los que incluían la variable “Fuente de la que se obtuvo el coeficiente de fiabilidad”, es decir, si este fue reportado por los autores, reportado tras contactar con los autores o estimado por nosotros. Para la C-FOCI Symptom Checklist y Symptom Severity, los modelos mejor posicionados fueron los que incluían únicamente las variables “Tamaño muestral” y “Género”, respectivamente.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**El fenómeno de la Inducción de la Fiabilidad en Psicología: Una meta-revisión.** Julio Sánchez-Meca<sup>1</sup>, Manuel Jesús Albaladejo Sánchez<sup>1</sup>, Carmen López-Ibáñez<sup>1</sup> y José Antonio López-López<sup>1</sup>

<sup>1</sup>Universidad de Murcia

La fiabilidad no es una propiedad inherente al test, sino que depende de las puntuaciones obtenidas en una aplicación del mismo. Sin embargo, es común

encontrarse con frases del tipo “la fiabilidad del test es 0.80”, así como con estudios que no la calculan con los datos de su muestra o incluso la omiten por completo. Si bien la práctica correcta es reportar la fiabilidad obtenida con los datos calculados en la aplicación del estudio, la más común es la denominada inducción de la fiabilidad, que se da cuando se asume como propia la obtenida en una aplicación anterior. La inducción de la fiabilidad se puede dividir en inducción por omisión (no se menciona en absoluto la fiabilidad), por reporte vago (no se da una estimación concreta de esta) y por reporte preciso (se da una estimación concreta calculada en una aplicación anterior). El objetivo de esta comunicación es estimar las tasas de inducción de fiabilidad de los tests psicológicos y comprobar si las tasas de reporte correcto han mejorado a lo largo de los años. Para ello se realizó una meta-revisión de los estudios de Generalización de la Fiabilidad (GF) de test psicológicos de entre 1998 y 2021, consultando en PsycInfo, Pubmed, Scopus y Google académico con las palabras clave “reliability generalization”, “reliability induction” o “reporting reliability” en título o abstract. Se discutirán los resultados, así como sus implicaciones para el reporte correcto de los resultados de la investigación psicológica.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

## SIMPOSIO

### Innovaciones Metodológicas en Evaluación de Programas: Calidad Metodológica y Tamaño de Efecto de Programas de Intervención

**Coordinan:** Susana Sanduvete-Chaves y Salvador Chacón-Moscoso

**Resumen.** En la práctica habitual en evaluación de programas, no se explicitan todos los aspectos fundamentales de cómo se llevó a cabo la intervención ni se justifica si se solventó el problema de la manera más eficaz posible. Esto ocurre en los diferentes ámbitos de intervención, tales como el social, educativo, de la salud, deportivo o clínico, entre otros. Se entiende por tanto que existe cierta debilidad metodológica en la evaluación de programas de intervención que, si se solventara, facilitaría la replicabilidad de las intervenciones exitosas y la acumulación integrada del conocimiento y, por tanto, el crecimiento y avance de la ciencia. En este simposio, se expondrán los avances del grupo de investigación ‘Innovaciones metodológicas en evaluación de programas’ (HUM-649, Junta de



Andalucía) en esta línea de trabajo, junto con otros grupos de investigación colaboradores. Se presentarán cinco comunicaciones: la primera muestra una escala de diez ítems para medir la calidad metodológica en estudios experimentales y cuasi-experimentales con niveles adecuados de fiabilidad y evidencias de validez en base al contenido y la estructura interna del instrumento; la segunda comunicación explica el proceso de evaluación del clima laboral en urgencias hospitalarias desde una perspectiva de método mixto a través de un estudio de caso; la tercera presentación se trata de un estudio de invarianza factorial de una escala que mide clima laboral en urgencias; la cuarta comunicación muestra un estudio de factores predictores de la salud mental de mujeres víctimas de malos tratos por parte de sus parejas, con el fin de tener información para dirigir las intervenciones en este ámbito hacia acciones eficaces; finalmente, la quinta presentación utiliza el análisis de redes para la detección de síntomas nucleares y puentes entre la depresión y la ansiedad.

Este simposio ha sido financiado por el Fondo Nacional de Desarrollo Científico y Tecnológico FONDECYT Regular, ANID, Gobierno de Chile [1190945]; el proyecto FEDER 2014-2020, Junta de Andalucía, España [US-1263096]; el VI Plan de Investigación y Transferencia-VI-PPITUS, Universidad de Sevilla, España [VIPP PRECOMPETI 2020/1333]; el proyecto PID2020-115486GB-I00 financiado por MCIN/AEI/10.13039/501100011033; y Xarxa R+D+I en Tecnologies de la Salut (XarTEC SALUT) [Red R+D+I en Tecnologies de la Salut] (2020-2022), financiado por AGAUR (Agència de Gestió d'Ajuts Universitaris de Recerca) y Unión Europea (Fons Europeu de Desenvolupament Regional) [2018 XARDI 00016].

**Facetas sobre la calidad metodológica de los estudios primarios: una validación empírica.** Salvador Chacón-Moscoso<sup>1,2</sup>, Susana Sanduvete-Chaves<sup>1</sup>, José Antonio Lozano-Lozano<sup>2</sup>, Francisco Pablo Holgado-Tello<sup>3</sup>

<sup>1</sup>Universidad de Sevilla, <sup>2</sup>Universidad Autónoma de Chile,

<sup>3</sup>Universidad Nacional de Educación a Distancia

Existe una amplia variedad de enfoques para medir la calidad metodológica en estudios primarios sobre intervenciones. Los profesionales, generalmente no expertos en metodología, no tienen criterios claros para evaluar los programas de intervención o estudios primarios. Además, los instrumentos de medida disponibles carecen de evidencia de sus propiedades métricas y no explican las razones por las que se seleccionan unos indicadores y no otros. Este trabajo propone la Escala de Calidad Metodológica (Methodological Quality Scale -MQS-) para estudios de

intervención experimentales y cuasi-experimentales como una herramienta simple y útil con evidencia de fiabilidad y validez en base al contenido y la estructura interna del instrumento. Dos codificadores aplicaron la MQS de forma independiente a un conjunto de estudios primarios sobre programas de formación en organizaciones obtenidos de una revisión sistemática. Se realizaron análisis descriptivos de las puntuaciones de los ítems. Para obtener las facetas de validez en la escala, se realizó un análisis paralelo utilizando la implementación óptima para determinar el número de dimensiones antes de realizar diferentes análisis factoriales exploratorios para identificar las dimensiones principales. Para cada faceta de validez obtenida, presentamos estadísticas descriptivas básicas (media, desviación estándar, fiabilidad y discriminación promedio). Además, las puntuaciones de las facetas de validez se sometieron a una interpretación teórica basada en el modelo de validez de Shadish, Cook y Campbell. Los resultados incluyeron (a) una validación empírica de posibles facetas de validez sobre la calidad metodológica desde un marco teórico; y (b) una interpretación de las puntuaciones en base a las facetas de validez obtenidas. La MQS, de 10 ítems, tiene en cuenta la validez interna, de constructo y externa en estudios con intervención. Además, se especifican los criterios de inclusión de los ítems, es de fácil aplicación y permite obtener perfiles de calidad metodológica en diferentes áreas de intervención.

**Proceso de evaluación del clima laboral en urgencias hospitalarias desde una perspectiva de método mixto: un estudio de caso.** M. Teresa Anguera<sup>1</sup>, José Antonio Lozano-Lozano<sup>2</sup>, Susana Sanduvete-Chaves<sup>3</sup>, Salvador Chacón-Moscoso<sup>3,2</sup>

<sup>1</sup>Universidad de Barcelona, <sup>2</sup>Universidad Autónoma de Chile,

<sup>3</sup>Universidad de Sevilla

A partir de estudios previos sobre el constructo clima laboral en servicios de urgencias hospitalarias, se delimitó un modelo teórico con cuatro dimensiones: satisfacción, productividad, relaciones interpersonales y ejecución en el trabajo. En este estudio, la evaluación del clima laboral en el servicio de urgencias de un hospital se realiza desde el marco de los mixed methods. En el estudio de caso que presentamos, se ha llevado a cabo una entrevista en profundidad a un médico de urgencias y se encuadra, por tanto, en la observación indirecta. El planteamiento de los mixed methods se concreta, en metodología observacional, en la sucesión de las tres macro etapas CUAL-CUAN-CUAL. La recogida de datos (CUAL), que se concreta en las respuestas a la entrevista en profundidad requiere, a partir del marco teórico, la construcción de un instrumento de observación, formado por dimensiones y

sistemas de categorías (con sus respectivos códigos) exhaustivas y mutuamente excluyentes elaborado a partir de cada una de las dimensiones. Asimismo, mediante el criterio sintáctico, se segmenta el texto en unidades textuales, marcando el inicio y final de cada una. Es posible ya proceder a la codificación de las respuestas a la entrevista, obteniéndose una matriz de códigos cuyas columnas corresponden a las respectivas dimensiones (que son 4), y cuyas filas son las respectivas unidades temáticas. La información que ahora se dispone es cualitativa y muy sistematizada. A partir de ella, se lleva a cabo la etapa CUAN, realizándose un análisis de coordenadas polares, que permite obtener un mapa de interrelaciones entre cada conducta focal (una categoría en cada análisis) y las conductas condicionadas (el conjunto de todas las categorías). Además, se pueden graficar estas interrelaciones mediante vectores, con su longitud y ángulo. A partir de los resultados, tiene lugar la segunda etapa CUAL, interpretándose los resultados.

**Factores predictores de salud mental en mujeres víctimas de violencia de pareja en Chile.** Paloma Labra-Valerdi<sup>1</sup>, Salvador Chacón-Moscoso<sup>2,1</sup>, Susana Sanduvete-Chaves<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Chile, <sup>2</sup>Universidad de Sevilla

Los estudios sobre violencia contra las mujeres específicos en población chilena son escasos. Como resultado, los programas gubernamentales de tratamiento carecen de una perspectiva local. Se analizaron variables predictoras de la salud mental de mujeres chilenas que habían sufrido maltrato por parte de sus parejas. Participaron en el estudio 202 mujeres que acudían periódicamente a Centros de la Mujer públicos; en promedio, habían sufrido 11 años de maltratos. Se realizaron análisis de regresión logística para determinar qué variable/s de sus historias de violencia predecía/n mejor sus estados de salud mental. Entre otros factores protectores, el aumento tanto de la resiliencia como del tiempo transcurrido desde el último episodio violento se tradujo en una reducción de los niveles de malestar psicológico general ( $B = -1,836$ ,  $p < ,001$  y  $B = 1,117$ ,  $p < ,001$  respectivamente), trastorno de estrés postraumático ( $B = -1.243$ ,  $p = .002$  y  $B = 1.221$ ,  $p < .001$  respectivamente) y depresión ( $B = -1.822$ ,  $p < .001$  y  $B = 1.433$ ,  $p < .001$  respectivamente). Como factores de riesgo, se encontró tener un alto nivel de factores estresantes adicionales, asociado a mayores niveles de angustia psicológica general ( $B = 1.007$ ,  $p = .005$ ), trastorno de estrés postraumático ( $B = 0.928$ ,  $p = .013$ ) y depresión ( $B = 1.061$ ,  $p = .016$ ). Como conclusión, cabe destacar que el Centro de la Mujer es el lugar donde las mujeres se sienten más apoyadas. Para mejorar la

eficacia de los tratamientos en estos centros y ayudar a la recuperación de las mujeres que han sufrido violencia de pareja, se deben tener en cuenta los factores predictivos significativamente asociados al estado de salud mental. Esto implicaría priorizar los casos en los que el último episodio de violencia fue más reciente, abordar factores estresantes adicionales y promover la resiliencia.

**Estudio de invarianza de medida de la versión reducida de la escala de clima laboral en servicios de emergencias hospitalarias.** José Antonio Lozano-Lozano<sup>1</sup>, Susana Sanduvete-Chaves<sup>2</sup>, Salvador Chacón-Moscoso<sup>1,2</sup>, Francisco Pablo Holgado-Tello<sup>3</sup>

<sup>1</sup>Universidad Autónoma de Chile, <sup>2</sup>Universidad de Sevilla,

<sup>3</sup>Universidad Nacional de Educación a Distancia

El presente estudio se basa en una escala de 24 ítems tipo Likert que mide clima laboral en servicios de emergencias hospitalarias. Esta escala presenta índices de ajuste globales apropiados (RMSEA = 0,06 con un intervalo del 90% de 0,05 a 0,07, SRMR = 0,08, GFI = 0,9, AGFI = 0,96, CFI = 0,98, NFI = 0,95 y NNFI = 0,98), evidencias de validez de criterio ( $\rho_{XY} = 0,68$ ,  $p < 0,001$ ) y una excelente fiabilidad ( $\alpha = 0,94$  y  $\omega = 0,94$ ), con muestra chilena. Este trabajo presenta un estudio empírico de la invarianza de medición en el área sustantiva del clima laboral en servicios de emergencia en función del país de origen (Chile y España) y del género (hombres y mujeres). Participaron 242 trabajadores de Chile ( $M = 32,71$ ;  $DT = 9,25$ ) y 244 de España de 20 a 58 años ( $M = 41,60$ ;  $DT = 10,23$ ), perteneciente al servicio de urgencias de una organización sanitaria de Santiago (Chile). El análisis factorial confirmatorio proporcionó índices globales adecuados para apoyar la existencia de una estructura común, un patrón invariante de cargas factoriales e invariancia del modelo estructural. La versión reducida de la escala mostró una invarianza de medición adecuada para las variables analizadas, confirmando su usabilidad en España y Chile y con personas de diferente género.

**Análisis de redes en el estudio de la comorbilidad en los problemas emocionales infanto-juveniles.** Francisco Pablo Holgado-Tello<sup>1</sup>, Milagros Ocalin Sánchez-Hernández<sup>2</sup>, Miguel Ángel Carrasco<sup>1</sup>

<sup>1</sup>Universidad Nacional de Educación a Distancia, <sup>2</sup>Universidad Nacional Autónoma de Nicaragua

El Análisis de Redes ha abierto nuevas puertas y posibilidades en la evaluación y medida de constructos psicológicos. Se ha mostrado de gran utilidad en el

campo de la psicopatología dado que permite encontrar patrones de covariación entre los síntomas, permitiendo identificar grupos de síntomas y síntomas claves que actúan como nodos que activan otros síntomas. La identificación de dichos nodos facilita la intervención, ya que permite focalizarla sobre los aspectos nucleares. En el presente trabajo se trata de poner de relieve las principales aportaciones del Análisis de Redes en el estudio de la comorbilidad. Desde un enfoque práctico e instrumental se abordarán los principales indicadores proporcionados por esta técnica de análisis para el estudio de la morbilidad conjunta. En concreto, una vez expuestos los principales índices para poder interpretar la comorbilidad en el ámbito de la psicopatología, expondremos un ejemplo aplicado sobre los problemas emocionales infanto-juveniles centrados sobre la depresión y la ansiedad. Los instrumentos utilizados han sido el CES-DC (Sánchez-Hernández, Delgado, Carrasco & Holgado-Tello (2018) y el YSR (Achenbach, 1991). La muestra estuvo formada por 986 niños y adolescentes españoles procedentes de la población general, 540 niñas (55%) y 446 niños (45%), con edades que oscilan entre los 9 y los 18 años (M= 13,09; DT= 2,01). Los principales síntomas que actúan de mediadores o puentes entre la depresión y la ansiedad serán puestos de manifiesto.

## SIMPOSIO

### Medición de la Recuperación en Personas Diagnosticadas con Trastorno Mental Severo

**Coordinan:** Juana Gómez-Benito y Nekane Balluerka

**Resumen.** Las políticas públicas de atención sanitaria de los trastornos más graves se iniciaron con modelos basados en la guardia y custodia, transitaron a modelos centrados en el cuidado y la remisión y, más recientemente, han apostado por un modelo basado en la recuperación. Esta nueva perspectiva, que modifica y cambia el papel de los diferentes protagonistas, debe modificar sus métodos de evaluación e intervención, ahora orientadas a la recuperación subjetiva y el empoderamiento personal, un nuevo modelo que prioriza la autodeterminación y el respeto por las preferencias y valores de las personas. La incorporación de este modelo requiere informar, capacitar y medir sus resultados. Una de las revisiones sistemáticas más consultadas sobre instrumentos de medición de la recuperación (Sklar, Groessl, O'Connell, Davidson y Aarons, 2013) evidencia que se han desarrollado numerosos instrumentos para evaluar la recuperación en salud mental, que utilizan diferentes

conceptualizaciones de la recuperación y cuyas propiedades psicométricas no siempre son óptimas. Además, gran parte de ellos no están adaptados al contexto español. Teniendo en cuenta que las personas en recuperación, los profesionales y los responsables de políticas sanitarias abogan por una atención de salud mental orientada a la recuperación, la promoción de la recuperación constituye una característica destacada de la política de salud mental a nivel internacional. Partiendo de dicha premisa y tratando de cubrir las lagunas arriba señaladas, las contribuciones de este simposio pretenden avanzar en esta línea, intentando clarificar la definición de recuperación, aportar instrumentos de medida con excelentes propiedades métricas para evaluar la recuperación de los trastornos mentales graves en la población española, analizando además la calidad de los instrumentos ya disponibles, y proponer y evaluar la eficacia de una intervención basada en el modelo de recuperación. En conjunto, todo ello favorecerá el cumplimiento de las actuales directrices de la OMS y la ONU.

**Definición, evaluación e intervención en el proceso de recuperación en salud mental.** M<sup>a</sup> Dolores Hidalgo<sup>1</sup>, Nekane Balluerka<sup>2</sup>, Georgina Guilera<sup>3</sup>

<sup>1</sup>Universidad de Murcia, <sup>2</sup>Universidad del País Vasco UPV/EHU, <sup>3</sup>Universitat de Barcelona

Las políticas de salud mental internacionales (OMS, 2013; OMS-Europa, 2013) están cada vez más comprometidas con un cambio de paradigma hacia un modelo de recuperación en el que la salud mental no es equivalente a la ausencia de síntomas o a la adaptación funcional a la sociedad, sino a la posibilidad de desarrollar una vida satisfactoria que responda a las preferencias y a los valores de la persona. El presente proyecto trata de impulsar el avance del paradigma de la recuperación en las políticas de salud mental en España incidiendo en la conceptualización, la evaluación y la intervención vinculadas al proceso de recuperación. La creación de un sistema de salud orientado a la recuperación requiere, en primer lugar, definir el concepto de recuperación. Este concepto se utiliza con diferentes significados en el paradigma de recuperación en salud mental, lo que genera confusión respecto a los objetivos y la orientación de los servicios que tratan de implicarse en este paradigma. Además, esta polisemia conceptual genera desacuerdos en la forma en la que debe evaluarse la recuperación y en el enfoque de las intervenciones que tratan de fomentarla. Por ello, además de una adecuada conceptualización, la implantación de un sistema sanitario basado en la recuperación requiere adaptar y validar instrumentos de medida que evalúen con rigor

psicométrico la recuperación de los trastornos mentales graves en la población española. Por último, el avance hacia este nuevo sistema de salud en España también requiere proponer programas de intervención eficaces que promuevan el bienestar y el desarrollo de un proyecto de vida que respete las preferencias de las personas que padecen problemas de salud mental. En esta comunicación se presentan los elementos clave que configuran la propuesta de nuestro equipo para promover el avance del paradigma de la recuperación en España.

### **Elementos clave de la recuperación en salud mental a través del método Delphi.** Estefanía Guerrero<sup>1</sup>, Alba Aza<sup>1</sup>, Maite Barrios<sup>1</sup>

<sup>1</sup>Universitat de Barcelona

Pese a la implementación cada vez más extendida del paradigma de recuperación en salud mental que toma como centro a los usuarios de los servicios de salud, no existe todavía un consenso internacional sobre sus elementos clave. En este contexto, la opinión de personas expertas resulta esencial para eliminar ambigüedades y encaminar esfuerzos de políticas públicas y servicios de salud para mejorar la calidad de vida de los usuarios de los servicios de salud mental. El objetivo de este estudio es establecer un consenso internacional sobre los elementos que la recuperación en salud mental tiene para cada uno de los colectivos implicados: usuarios de los servicios de salud, profesionales de la salud, y familiares y cuidadores. Con esta finalidad, se diseña un estudio Delphi de tres rondas para cada colectivo utilizando el programa Qualtrics. La metodología Delphi es sumamente eficaz y ampliamente utilizada en investigación para obtener opinión consensuada y representativa de un grupo de expertos. La primera ronda plantea diferentes preguntas abiertas relacionadas con la definición, los elementos clave y los factores facilitadores e inhibidores de la recuperación. En la segunda se valora la relevancia de los aspectos identificados a partir del análisis cualitativo de la ronda anterior. Finalmente, la tercera ronda consiste en reevaluar aquellos elementos que obtuvieron un mayor grado de acuerdo en la segunda ronda. Los participantes se seleccionan mediante el contacto con asociaciones de usuarios y familiares y la búsqueda sistemática en bases de datos y planes estratégicos de salud mental, asegurándose que todos cumplan los criterios de inclusión y exclusión. Se espera que los resultados contribuyan a la definición de la recuperación en salud mental para cada uno de los colectivos estudiados y permita el diseño de instrumentos específicos para medir y monitorizar adecuadamente el proceso, así como mejorar las intervenciones actualmente ya desarrolladas.

### **Adaptación de instrumentos para medir el proceso de recuperación.** Maite Barrios<sup>1</sup>, Nuria Martín-Ordiales<sup>2</sup>, Hernán María Sampietro<sup>1</sup>

<sup>1</sup>Universitat de Barcelona, <sup>2</sup>Universidad de Murcia

El modelo de recuperación en salud mental surge en la década de 1990 como respuesta a las críticas a la psiquiatría por parte de diferentes movimientos de derechos civiles, los cuales desembocarían en una amplia aceptación de la desinstitucionalización. Desde entonces, el concepto de recuperación ha cobrado gran relevancia. Dada la importancia de la temática es esencial disponer de instrumentos con evidencias de validez adecuadas y demostrar que las puntuaciones que de ellos se obtienen son fiables. No obstante, mientras que en lengua inglesa se dispone de un amplio abanico de instrumentos, en español sólo se cuenta con alguna adaptación. Disponer de instrumentos adaptados a nuestro contexto es esencial para alcanzar a una mayor comprensión del constructo, permitir la comparación entre estudios de otras regiones y países y llegar a una síntesis en la construcción del conocimiento. Nuestro equipo tiene como objetivo adaptar y validar diferentes instrumentos que valoran el proceso de recuperación en salud mental utilizando para ello una amplia muestra de personas diagnosticadas según la Clasificación Internacional de Enfermedades (CIE-11) de un trastorno mental grave. Para la adaptación se ha seguido el proceso de traducción y retrotraducción recomendado por la International Test Commission (2017). El significado psicológico de los enunciados de los ítems de las versiones en inglés (original y retrotraducida) se ha comparado incorporando a los autores de la versión original en el proceso de adaptación. Para el proceso de validación, se han llevado a cabo los análisis psicométricos propios de un estudio de este tipo. Nuestra investigación supone un estudio pionero en España, dado que la selección de herramientas que proponemos adaptar y validar tiene en cuenta la concepción más actual y multidimensional del proceso de recuperación en salud mental, incorporando los instrumentos de más reciente creación, los cuales presentan visiones más actualizadas del constructo.

### **Análisis de la calidad de los instrumentos de recuperación.** Georgina Guilera<sup>1</sup>, Jorge Torres-Marín<sup>1</sup>, Arantxa Gorostiaga<sup>2</sup>

<sup>1</sup>Universitat de Barcelona, <sup>2</sup>Universidad del País Vasco UPV/EHU

El impulso del modelo de recuperación en salud mental de las últimas décadas ha provocado un incremento en la disponibilidad de instrumentos de evaluación de la recuperación. Los profesionales que trabajan en este

campo deben conocer cuál es la calidad de los instrumentos existentes para poder tomar decisiones que se adecuen a su contexto de evaluación. Con este estudio pretendemos avanzar en el análisis de la calidad psicométrica de los instrumentos existentes para la medida de la recuperación, con la intención de facilitar la toma de decisiones de los profesionales que potencialmente utilizarán dichos instrumentos en su práctica clínica o investigadora. Se presentan los resultados de la valoración de la calidad de la Recovery Assessment Scale (RAS; Corrigan et al, 2004; Giffort et al., 1995), una escala breve que evalúa el proceso de recuperación en salud mental, adoptando dos estrategias metodológicas distintas. Por un lado, se utiliza la 'COSMIN Risk of Bias checklist' (Mokkink et al, 2018), una herramienta que posibilita valorar de forma sistemática las propiedades psicométricas de los instrumentos de autoreporte del estado de salud. Y, por otro lado, se plantea un estudio de generalización de la fiabilidad de dicha escala. En la búsqueda electrónica realizada en enero de 2022 con el nombre de la escala como palabra clave, se obtuvieron 361 documentos, cuyos títulos y resúmenes fueron revisados para seleccionar aquellos que analizaron alguna de las propiedades psicométricas incluidas en COSMIN. Los estudios que además reportaron datos sobre la fiabilidad de la RAS pasaron a formar parte del estudio de generalización de la fiabilidad. La síntesis de las propiedades psicométricas de la RAS presentada en este estudio permitirá ofrecer recomendaciones sobre su uso para la medición de la recuperación.

**Eficacia de los programas de intervención.** Hernán María Sampietro<sup>1</sup>, Viviana Carmona<sup>2</sup>, Juana Gómez-Benito<sup>1</sup>

<sup>1</sup>Universitat de Barcelona, <sup>2</sup>Actívament Catalunya Associació

Existen en el mundo anglosajón diversas herramientas creadas por personas usuarias de servicios de salud mental, basadas en su propia experiencia, orientadas a promover la autodeterminación en los procesos de recuperación, empoderamiento y bienestar. Hasta la actualidad, ninguna de estas herramientas ha sido implementada y evaluada en nuestro contexto cultural. Para realizar una evaluación de la eficacia de estos programas de recuperación se ha seguido un doble objetivo: (a) Identificar las características distintivas de estas herramientas y conocer cómo fueron implementadas y evaluadas; (b) Implementar y evaluar la efectividad de un taller basado en el modelo de recuperación usando una de estas herramientas. Para el primer objetivo, se realizó una revisión de alcance, siguiendo el enfoque de Arksey y O'Malley y la extensión PRISMA para revisiones de alcance (PRISMA-ScR). Se

utilizaron cinco estrategias de búsqueda: contacto con redes internacionales de usuarios y supervivientes de la psiquiatría, búsquedas en bases de datos académicas, en la literatura gris, en Google Scholar y revisión de referencias. Se encontraron 6 herramientas diferentes. Sólo una de ellas, el WRAP, ha sido ampliamente evaluada. Pese a que no es una intervención clínica, la variable más evaluada es la mejora de síntomas. Para el segundo objetivo, se realizó un diseño cuasiexperimental, pre-post test con grupo control. Se ha reclutado una muestra aproximada de 200 participantes de 13 Servicios de Rehabilitación Comunitaria colaboradores que implementan el Taller de Recuperación, incluyendo 100 personas en el grupo experimental (participantes del taller) y 100 personas en el grupo control (usuarios del servicio, que no participan del taller). Se realizarán análisis multivariados de los datos para diseños de medidas repetidas para las variables: recuperación, esperanza, apoyo social y empoderamiento. Se realizarán comparaciones entre e intragrupos. Si se demuestra la eficacia del programa, podría pasar a formar parte de la cartera de servicios.

## SIMPOSIO

### Mejorando los Test y su Uso: Hecho y por Hacer

Coordina: Ana Hernández

**Resumen.** Los test son una herramienta fundamental para los psicólogos. Por ello es esencial velar por su calidad psicométrica y su adecuado uso, especialmente en el contexto actual, donde la tecnología está cambiando todo lo relacionado con la evaluación mediante test.

Para garantizar la calidad de los test, las casas editoriales de test juegan un papel primordial. Pero las revisiones externas de los test por parte de expertos también contribuyen a garantizar dicha calidad, introduciendo recomendaciones para mejorar las pruebas y sus manuales. En cuanto al buen uso de los test, éste requiere que los usuarios tengan la formación y competencias evaluativas y psicométricas necesarias, pero también que sigan toda una serie de códigos y estándares profesionales. Este simposio presenta algunas de las acciones llevadas a cabo por la Comisión de Test del Consejo General de la Psicología de España con el fin de mejorar los test psicométricos y su uso. El primer trabajo se centrará en los resultados de las dos últimas rondas de revisiones de test realizadas y las sugerencias de mejora en el proceso, poniendo en perspectiva 10 años de revisiones. El segundo se ocupa del análisis de distintas sentencias judiciales

relacionadas con el uso de los test en los procesos selectivos de las administraciones públicas, dando una serie de recomendaciones sobre el uso de los test en dichos procesos. El tercero aporta el punto de vista una editorial de test sobre cuáles son los principales avances que se han producido y los desafíos existentes. El cuarto trabajo analiza el impacto que la tecnología está teniendo tanto en las pruebas como en su uso. Finalmente, el simposio se cierra con el análisis global por parte del discussant de la sesión de los logros y los retos a los que nos enfrentamos, con el fin de enriquecer el debate.

### **Evaluación de test editados en España: Una década revisando test.** Francisco J. Abad<sup>1</sup>, Luis M. Lozano<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad de Granada

Desde 2010, la Comisión de Test del Consejo General de la Psicología del Colegio Oficial de Psicólogos lleva a cabo una revisión de la calidad de test publicados en España. En el presente trabajo se resumen los resultados de las dos últimas ediciones de este proceso, la novena y la décima. Participaron un total de 26 expertos, que evaluaron un total de 13 test, a través del Cuestionario de Evaluación de Test-Revisado, CET-R. En la ponencia se describen los puntos más novedosos de cada convocatoria y se comentan las valoraciones de las pruebas, en términos absolutos y en términos relativos, en comparación con las obtenidas en los procesos previos. Para ambas ediciones, se recogen los puntos fuertes y los posibles aspectos de mejora, en cada uno de los apartados (Materiales y documentación, Fundamentación teórica, Análisis de ítems, Adaptación, Validez, Fiabilidad, Baremos), aunque la conclusión general es que los test en España gozan de buena salud. En segundo lugar, se recogen algunas sugerencias de mejora en el proceso de evaluación y sobre el cuestionario CET-R, valorando la fiabilidad interjueces de los expertos, así como las posibilidades de automatizar algunos puntos a través del uso de una plataforma informatizada. Por último, se valora la repercusión que tiene la existencia de este proceso de revisión en la calidad de los test desarrollados y se reflexiona sobre las posibles medidas para aumentar el conocimiento de estos procesos de revisión por parte de los psicólogos españoles.

### **Los test y los procesos de selección en las administraciones públicas.** Ana Hernández<sup>1</sup>, Paula Elosua<sup>2</sup>, Francisco J. Abad<sup>3</sup>, José R. Fernández-Hermida<sup>4</sup>

<sup>1</sup>Universidad de Valencia, <sup>2</sup>Universidad del País Vasco,

<sup>3</sup>Universidad Autónoma de Madrid, <sup>4</sup>Universidad de Oviedo

Durante los últimos años ha aumentado sustancialmente el número de recursos y sentencias judiciales relacionadas con el uso de test psicométricos en los procesos selectivos de las administraciones públicas. Entre las sentencias publicadas cobra especial relevancia la Sentencia de la Sala Tercera del Tribunal Supremo de 27 de enero de 2022 (rec. 8179/2019) que establece jurisprudencia sobre el uso de estas pruebas. La revisión de esta y otras sentencias por parte de la Comisión de Test del Consejo General de la Psicología de España, ha dado lugar a un informe que recoge algunos de los puntos que pueden resultar más controvertidos relacionados con el uso de test en procesos selectivos de las administraciones públicas. Asimismo, a partir de las normas y códigos de buenas prácticas profesionales en los procesos de evaluación de personas, en general, y en contextos organizacionales, en particular, en el informe se ofrece una serie de recomendaciones dirigidas a las administraciones públicas convocantes y a los tribunales y servicios de evaluación implicados, así como una serie de aclaraciones a los organismos judiciales sobre la idiosincrasia de la evaluación psicológica, no equiparable a pruebas de rendimiento. El objetivo de este trabajo es compartir el análisis de los principales problemas detectados y de las principales medidas y recomendaciones que se proponen para mejorar la transparencia y verificabilidad de los procesos selectivos de las administraciones públicas. Todo ello sin olvidar la necesidad de salvaguardar las pruebas y sus materiales y mejorar la confianza en la profesión de la Psicología.

### **Avances y desafíos en el desarrollo de tests psicológicos: Visión desde una editorial.** Pablo Santamaría<sup>1</sup>

<sup>1</sup>Hogrefe TEA ediciones

El objetivo de esta comunicación será presentar una panorámica amplia, desde el punto de vista de una editorial, de qué avances se han realizado en el mundo de los test psicológicos en los últimos años y, por otra parte, cuáles son los principales desafíos que se presentan actualmente y los retos que suponen. Con relación a los avances, se describirán diversos aspectos tanto puramente técnicos como de uso profesional que se han consolidado en la práctica habitual, tales como el continuous norming (sistema de baremos continuos para grupos más pequeños de edad que permiten una corrección más precisa), sistemas alternativos de corrección que ponderan por edad y nivel educativo para obtener baremos ajustados (especialmente en el campo de neuropsicología donde el nivel educativo es un factor importante para el análisis del deterioro cognitivo), sistemas de aplicación y corrección on-line, informes automáticos interpretativos... En cuanto a los desafíos y

retos existentes, se abordará la necesidad de modelos cada vez más sólidos de las distintas áreas de psicología (por ejemplo, modelo HiTOP en psicopatología), la brecha existente entre la psicometría más puramente académica y la psicometría aplicada o la problemática de la proliferación de pseudoherramientas basadas en nuevas tecnologías, entre otras cuestiones. Finalmente, se tratará de presentar un cierto balance de estos avances y retos y se plantearán algunas implicaciones que tiene esta situación tanto a nivel profesional como institucional.

### **Evaluación apoyada en tecnologías digitales: Desafíos y directrices.** Paula Elosua<sup>1</sup>

<sup>1</sup>Universidad del País Vasco

El impacto de la tecnología en el diseño y construcción de ítems y test, la posibilidad de acceso a datos personales obtenidos en ambientes no controlados, o el desarrollo de métodos computacionales para su análisis están abriendo nuevas líneas de trabajo en el campo de la medición y evaluación en ciencias sociales y de la salud. Las prácticas relacionadas con el uso de test como instrumentos de evaluación están sufriendo una profunda transformación. Hoy conviven los test tradicionales de lápiz y papel, con aplicaciones construidas sobre la omnipresencia de lo móviles, o formas de evaluar apoyadas en datos que no han sido generados con objetivos de evaluación específicos. En esta presentación intentamos mostrar algunos de los hitos relacionados con la evaluación y medición apoyada en tecnologías digitales como base para la reflexión sobre la situación actual y desarrollo futuros; desarrollo que no pueden olvidar los principios éticos y formales construidos a lo largo de la corta o larga historia relacionada con la construcción y uso de test.

## **SIMPOSIO**

### **Meta-Análisis: Avances Metodológicos**

**Coordina:** Juan Botella

**Resumen.** La metodología meta-analítica sigue experimentando continuos avances que permitan optimizar este método de investigación para obtener el máximo provecho de la síntesis de las investigaciones y para adaptarla a las características de los estudios y a las preguntas a las que esta metodología puede dar respuesta. Se presentan en este simposio cinco trabajos desarrollados por investigadores de diferentes universidades españolas, algunos de ellos en colaboración con investigadores de universidades

extranjeras, que suponen un avance en la metodología del meta-análisis. El primer trabajo, presentado por Fulgencio Marín Martínez (Universidad de Murcia), aborda de forma empírica la frecuencia con la que se producen diferencias importantes en la variabilidad de las puntuaciones cuando se comprueba la eficacia de una intervención mediante la aplicación de un diseño pretest-posttest, un fenómeno que tiene importantes consecuencias en la realización de meta-análisis. El segundo trabajo es presentado por Rubén López Nicolás (Universidad de Murcia) y aborda el estudio de la replicabilidad de los meta-análisis, en concreto, en términos de los análisis estadísticos llevados a cabo en la síntesis de un conjunto de estudios. El tercer trabajo, presentado por Alejandro Sandoval Lentisco (Universidad de Murcia) también se enmarca en el estudio de la replicabilidad de los meta-análisis, en concreto, en el grado de precisión con la que se elaboran los preregistros de los meta-análisis. El cuarto trabajo, presentado por Laura Badenes Ribera (Universidad de Valencia), presenta los resultados de un estudio del progreso sobre la reproducibilidad de los meta-análisis sobre la eficacia de intervenciones. Finalmente, el quinto trabajo es presentado por Susana Sanduvete-Chaves (Universidad de Sevilla) y consiste en los resultados de un meta-análisis sobre la eficacia de las intervenciones psicológicas para disminuir la fusión cognitiva en pacientes con dolor crónico, haciendo especial hincapié en la relación entre calidad metodológica de los estudios y el tamaño del efecto.

**El supuesto de homogeneidad de varianzas en diseños pretest-posttest: evidencia empírica.** Fulgencio Marín Martínez<sup>1</sup>, Julio Sánchez Meca<sup>1</sup>, José Antonio López López<sup>1</sup>, María Rubio Aparicio<sup>2</sup>, Manuel Jesús Albaladejo Sánchez<sup>1</sup>, Mar Torregrosa Almagro<sup>1</sup>, David Pina Castillo<sup>1</sup> y Raquel Martínez Olivares<sup>1</sup>

<sup>1</sup>Universidad de Murcia, <sup>1</sup>Universidad de Alicante

Muchas técnicas e índices estadísticos asumen la homogeneidad de varianzas en diseños pretest-posttest. Sin embargo, desde un punto de vista teórico, se estima que la aplicación de una intervención o tratamiento entre el pretest y el posttest acompañada de una interacción sujeto-tratamiento, puede incrementar la variabilidad del posttest. Con el propósito de evaluar empíricamente la frecuencia con la que se produce este efecto en la población, así como la magnitud y el sentido de la razón entre varianzas, se seleccionaron 5 meta-análisis sobre la eficacia de intervenciones psicológicas, y se anotaron los valores de las medias y desviaciones típicas en los registros pretest-posttest de las tablas de resultados de los estudios de tales meta-análisis. En lo que respecta a los 555 registros pretest-posttest con intervenciones psicológicas clínicas entre el pretest y el posttest, en el

72,7% de las comparaciones pretest-postest no se encontraron diferencias significativas entre las varianzas del pretest y el postest, en el 10,2% se detectó una variabilidad significativamente superior en el postest y en el 17,1% restante una variabilidad significativamente superior en el pretest. Se discute el interés de estos resultados en la estimación de determinados tamaños del efecto y en la planificación de estudios de simulación Monte Carlo.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**Reproducibilidad analítica de 217 meta-análisis sobre intervenciones clínicas en psicología.** Rubén López-Nicolás<sup>1</sup>, Daniël Lakens<sup>2</sup>, Jose A. López-López<sup>1</sup>, Julio Sánchez-Meca<sup>1</sup>, María Rubio-Aparicio<sup>3</sup>, Alejandro Sandoval-Lentisco<sup>1</sup>, Carmen López-Ibáñez<sup>1</sup>, Desirée Blazquez-Rincón<sup>1</sup>

<sup>1</sup>Universidad de Murcia, <sup>2</sup>Eindhoven University of Technology,

<sup>3</sup>Universidad de Alicante

Actualmente, los estudios meta-analíticos son considerados como una de las herramientas de investigación más útiles. En un contexto en el que el número de publicaciones científicas es cada vez mayor, enfoques como este, centrados en el desarrollo de herramientas para la síntesis de la evidencia disponible, pasan a un primer plano. En los últimos años, diferentes aspectos de la credibilidad de los resultados publicados se han visto comprometidos. La evaluación empírica de los resultados publicados ha emergido como un área útil y necesaria, ante los resultados de diversos proyectos centrados en la evaluación de la replicabilidad y la reproducibilidad de estudios publicados. En este trabajo se trataron de reproducir analíticamente los resultados de 217 meta-análisis publicados en 100 artículos diferentes seleccionados aleatoriamente para un trabajo previo. En primer lugar, se encontraron problemas para reproducir estos meta-análisis en etapas tempranas. Para el 33% de los meta-análisis incluidos en este estudio, no se pudieron recuperar los datos originales necesarios para llevar a cabo los intentos de reproducción del resultado. Por otro lado, en cuanto a los meta-análisis en los que se pudo recuperar los datos originales, un 19% mostraron algún tipo de discrepancia menor. Estos resultados se discuten e integran con investigación previa en reproducibilidad de meta-análisis. Finalmente, se enumera una serie de recomendaciones a la luz de los resultados de este y otros trabajos relacionados.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**Prácticas de prerregistro en meta-análisis.** Alejandro Sandoval-Lentisco<sup>1</sup>, Tom Hardwicke<sup>2</sup>, José A. López-López<sup>1</sup>, Julio Sánchez-Meca<sup>1</sup>

<sup>1</sup>Universidad de Valencia, <sup>2</sup>Universidad de Amsterdam (Países Bajos)

Hacer un meta-análisis, al igual que ocurre con los estudios primarios, requiere que el investigador que lo lleva a cabo tome multitud de decisiones a la hora de diseñarlo, analizarlo y reportar los resultados. Estos “grados de libertad” presentan riesgos de sesgos, puesto que el investigador puede tomar decisiones que hagan que los resultados varíen a su conveniencia. Una de las formas más eficaces de minimizar estos posibles sesgos es prerregistrar un protocolo antes de llevar a cabo el meta-análisis. No obstante, el grado en el que un prerregistro es beneficioso depende de lo exhaustivo que sea el prerregistro y de que las decisiones sean, o bien, finalmente llevadas a cabo, o bien, justificadas en el caso de que se desvíen de lo prerregistrado. En este estudio, se examinarán las prácticas de prerregistro de una muestra de meta-análisis de Psicología. Primero, se examinará la frecuencia con la que los meta-análisis son prerregistrados. En segundo lugar, se examinará cómo de exhaustivos fueron los prerregistros de los meta-análisis. Por último, se compararán las decisiones establecidas en el prerregistro con las que finalmente aparecen en el artículo y si estas desviaciones fueron mencionadas y justificadas. Las implicaciones de estos resultados serán discutidas.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**Reproducibilidad de los meta-análisis sobre eficacia de intervenciones: Estudio de progreso.** Laura Badenes Ribera<sup>1</sup> y Julio Sánchez Meca<sup>2</sup>

<sup>1</sup>Universidad de Valencia, <sup>2</sup>Universidad de Murcia

En los últimos 10 años, la reproducibilidad y replicabilidad de los meta-análisis se ha convertido en una línea de investigación muy relevante, dado el importante papel que ejerce este tipo de investigaciones en el enfoque de la práctica basada en la evidencia. La gran cantidad de decisiones que el meta-analista tiene que tomar durante el proceso de extracción de los datos para la obtención o el cálculo de los tamaños del efecto, hace que esta etapa sea muy sensible a la comisión de errores. Estos errores son todavía más frecuentes en los meta-análisis sobre la eficacia de intervenciones, ya que se requiere extraer datos estadísticos de los estudios primarios para calcular los tamaños del efecto. Se presenta en este trabajo un



estudio de progreso de la reproducibilidad de este tipo de meta-análisis. Para ello, se seleccionó aleatoriamente un meta-análisis de entre 100 meta-análisis previamente identificados sobre la eficacia de intervenciones psicológicas, y se extrajeron los datos estadísticos de los estudios individuales de ese meta-análisis, se calcularon los tamaños del efecto y se compararon con los reportados en el meta-análisis original. Se discuten las implicaciones de los posibles errores e inconsistencias que se pueden producir en este tipo de meta-análisis.

Estudio financiado por la Agencia Estatal de Investigación (Gobierno de España). PID2019-104080GB-I00/MCIN/AEI/10.13039/50110 00110 33.

**Eficacia de las intervenciones psicológicas para disminuir la fusión cognitiva en pacientes con dolor crónico: una revisión sistemática y meta-análisis.** Alejandro Sandoval-Susana Sanduvete-Chaves<sup>1</sup>, Salvador Chacón-MoscOSO<sup>1,2</sup>, Francisco J. Cano-García<sup>1</sup>

<sup>1</sup>Universidad de Sevilla, <sup>2</sup>Universidad Autónoma de Chile

El papel de la fusión cognitiva en pacientes con dolor crónico, uno de los seis componentes clave del modelo de flexibilidad psicológica, aún no ha sido bien establecido. El objetivo de este estudio fue medir la eficacia de las intervenciones psicológicas para reducir la fusión cognitiva (PROSPERO-CRD42021255028). Se realizaron búsquedas de los estudios primarios relevantes en las bases de datos WOS, SCOPUS, Medline y PsycInfo realizados hasta enero de 2022. Los criterios de inclusión para los estudios fueron que se tratara de pacientes diagnosticados de dolor crónico, que hubiera intervenciones psicológicas para los pacientes, y que se dieran mediciones de fusión cognitiva. Se midió la calidad metodológica de los estudios primarios incluidos aplicando una escala previamente elaborada, y se calcularon los tamaños de efecto promedio (g de Hedges). Esta revisión sistemática y meta-análisis incluyó 12 artículos con 16 informes (13 estudios de pre-post/seguimiento y tres ensayos controlados aleatorios -ECA-). Según los hallazgos, la fusión cognitiva disminuyó significativamente después de la intervención. Los tamaños del efecto fueron pequeños/medianos en el post-test,  $g = -0.333$ ,  $p < .001$ , IC 95% [-0.478, -0.187]; y medio en el seguimiento a largo plazo,  $g = -0.534$ ,  $p < .001$ , IC 95% [-0.759, -0.309]. Similar tendencia se encontró en los ECA en el post-test,  $g = -0.622$ ,  $p = .042$ , IC 95% [-1.221, -0.023] y en el seguimiento a corto plazo,  $g = -0.793$ ,  $p < .001$ , IC 95% [-1.176, -0.401]. Se identificaron variables moderadoras como la situación laboral, el género, la intensidad del dolor, el nivel de depresión antes de la intervención, la duración y la modalidad de la intervención. Los resultados muestran la eficacia de las

intervenciones psicológicas para reducir la fusión cognitiva en pacientes con dolor crónico, y la influencia de variables moderadoras relevantes.

Trabajo financiado por el Fondo Nacional de Desarrollo Científico y Tecnológico FONDECYT Regular, ANID, Gobierno de Chile [1190945]; el proyecto FEDER 2014-2020, Junta de Andalucía, España [US-1263096]; el VI Plan de Investigación y Transferencia-VI-PPITUS, Universidad de Sevilla, España [VIPP PRECOMPETI 2020/1333]; el Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI 2020), Junta de Andalucía, España [ref. PY20-01122]; y el proyecto PID2020-115486GB-I00 financiado por MCIN/AEI/ 10.13039/501100011033.

## SIMPOSIO

### Metodología Cualitativa para el Avance en Psicología

**Coordinan:** Leire Aperribai y África Borges

**Resumen.** Las distintas disciplinas de la Psicología tratan de dar respuesta a fenómenos de diversa índole a través de las metodologías empíricas. Es larga la trayectoria de los métodos en psicología, y permiten explicar fenómenos humanos importantes. La metodología cualitativa es conocida por poder ofrecer datos que serían difíciles de obtener en situaciones experimentales o cuasiexperimentales. Sin embargo, los análisis de datos cualitativos han estado sujetos a una labor muy tediosa y a la subjetividad de las y los investigadores, lo que ha supuesto un empleo limitado de esta metodología. Sin embargo, las tecnologías han permitido elaborar programas que han facilitado superar las limitaciones mencionadas y promover avances en el conocimiento de distintos ámbitos de la Psicología. Así, se han podido abordar temas complejos por medio de esta metodología. Concretamente, en este simposio se analizarán temas como la atención y evaluación por parte de las y los profesionales de la salud de los/las niños/as y jóvenes con altas capacidades intelectuales, las habilidades sociales y la empatía como temas primordiales de la cognición social, las estrategias de afrontamiento del alumnado de secundaria, y las experiencias vividas por las mujeres que han sufrido la mutilación genital. Se puede constatar que temas tan diversos y complejos pueden abordarse desde la metodología cualitativa.

**El estudio de la atención y evaluación de la salud de los/as niños/as y adolescentes con altas capacidades intelectuales por medio del método mixto.** Leire Aperribai Unamuno<sup>1</sup>, Lorea Cortabarría Cortazar<sup>1</sup>

<sup>1</sup>Universidad del País Vasco

En la diversa literatura sobre los/las niños/as con alta capacidad intelectual se ha recogido poco sobre la atención que reciben en los servicios de salud. Sin embargo, es importante que las y los profesionales de la salud sepan de sus características y de cómo abordar las distintas alteraciones que puedan aparecer en este colectivo. Para poder abordar este aspecto se carece de instrumentos estandarizados que aclaren con qué formación y experiencia cuentan éstos. Así, el empleo de los diseños de métodos mixtos en este ámbito puede ser una opción importante. En este estudio, se pretenden estudiar estos aspectos desde el método mixto. Para ello, se ha enviado un cuestionario online a profesionales de la salud de la Comunidad Autónoma de Euskadi y se han recogido respuestas de 25 participantes. Los resultados muestran que las y los profesionales de la salud mencionan alteraciones relacionadas con el TEA o el TDAH y a psicopatologías tales como la ansiedad o la depresión. Se concluye que el empleo de este método puede contribuir a un mejor conocimiento del abordaje que realizan estos profesionales.

#### **El estudio de la cognición social por medio del método mixto.** Adalberto González<sup>1</sup>, Leire Aperribai<sup>2</sup>

<sup>1</sup>Universidad de La Laguna, <sup>2</sup>Universidad del País Vasco UPV/EHU

El estudio de la cognición social recobra importancia a medida que las sociedades van desarrollando una mayor dependencia en las relaciones sociales. Así, la cognición social se estudia por medio de distintas disciplinas y métodos cada vez más avanzados. También supone un avance el empleo de los diseños de métodos mixtos en el ámbito de la psicología. En este estudio, se pretenden estudiar dos variables propias de la cognición social, la empatía y las habilidades sociales, desde el método mixto. Para ello, se ha enviado un cuestionario online a personas de entre 18 y 65 años de características sociodemográficas distintas y se han recogido respuestas de 143 participantes a las preguntas “En la interacción social, ¿tú consideras que es necesario “ponerse en el lugar del otro”? ¿Por qué?” y “¿Te consideras una persona socialmente habilidosa? ¿Por qué?”. Los resultados muestran que en el constructo de la empatía destacan los conceptos entender, situación de uno mismo, la necesidad de ponerse en el lugar del otro y la comunicación son fundamentales en este constructo. En cuanto al constructo de habilidades sociales, se muestran conceptos tales como saber establecer relaciones en las situaciones sociales, la facilidad o capacidad de tratar a la persona en un entorno, conocer a la gente y sus problemas, y entablar una conversación y las dificultades inherentes a ésta. Se concluye que el empleo de este

método puede contribuir a la comprensión de conceptos complejos como éstos que se emplean en el ámbito de conocimiento de la cognición social.

#### **Estrategias de Afrontamiento en Alumnado de Secundaria.** Marina Gómez-Madrid<sup>1</sup>, Triana Aguirre Delgado<sup>1</sup>, Tania Martín Sabina<sup>1</sup>, África Borges del Rosal<sup>1</sup>

<sup>1</sup>Universidad de La Laguna

El afrontamiento es un proceso contextual referido a lo que la persona piensa y hace ante situaciones de estrés, por lo que resulta de gran importancia conocer las estrategias de afrontamiento que utilizan los jóvenes. El objetivo del estudio ha sido conocer las estrategias de afrontamiento que usa el alumnado de secundaria ante un problema. La muestra del estudio han sido 1.500 alumnos y alumnas de primero a tercero de ESO en centros educativos de la Comunidad Autónoma de Canarias, que respondían a la siguiente pregunta abierta: “Anota a continuación cualquier otra cosa que sueles hacer para afrontar tus problemas”. La recogida de la muestra se hizo de manera presencial en los centros y en formato electrónico mediante formulario de Google. Se realizó el análisis cualitativo mediante el programa de análisis de textos ALCESTE. Los resultados categorizan las respuestas emitidas en varias clases, que informan sobre cómo el alumnado afronta los problemas. Concluye la relevancia de las estrategias usadas.

#### **Una aproximación cualitativa a las consecuencias psicológicas de la mutilación genital femenina.** Maida A. S. Hussein<sup>1</sup>, Rubén García-Sánchez<sup>2</sup>, Emilio Verche<sup>3</sup>

<sup>1</sup>Universidad Europea de Madrid, <sup>2</sup>Universidad Autónoma de Madrid, <sup>3</sup>Universidad Complutense de Madrid

La mutilación genital femenina (MGF) es una práctica nociva que ha demostrado inducir complicaciones físicas aversivas. A pesar de no generar ningún beneficio, millones de niñas son sometidas a la MGF incluso más allá de sus países de origen. Los resultados físicos de la práctica están escasamente investigados y menos aún los psicológicos, al ser muestras de difícil acceso. Por esta razón, este estudio tiene como objetivo investigar las consecuencias psicológicas de la práctica utilizando un enfoque cualitativo. Participaron en el estudio seis mujeres que sufrieron mutilación genital femenina. Se realizó una entrevista biográfica semiestructurada que contenía temas que examinaban la historia de las participantes sobre la MGF, los sistemas de salud, las relaciones y la auto-imagen, las percepciones sobre la MGF y las perspectivas sobre el futuro. Se utilizó un análisis temático para examinar y determinar la experiencia de las participantes con la MGF. El estudio identificó cuatro temas principales de interés: la aflicción

psicológica, que engloba todos los síntomas traumáticos que experimentaron las participantes; los sistemas de atención médica, que analiza la experiencia de las mujeres con los sistemas de atención médica, incluidos ginecólogos, matronas y otro personal médico; amor propio y sexualidad, que explora la vida sexual de las mujeres, sus matrimonios y relaciones, así como también cómo experimentaron la menstruación después de ser mutiladas; por último, MGF: presente y futuro, que examina la percepción de las participantes sobre la práctica actual y cómo la ven en el futuro. Los hallazgos indican que las mujeres que sufrieron mutilaciones se vieron afectadas psicológicamente. Se observaron consecuencias tanto físicas como psicológicas aversivas en las participantes. El análisis temático proporciona una investigación precisa y completa de la experiencia de estas mujeres y se conforma como una metodología apropiada para el estudio de la MGF.

## SIMPOSIO

### Retos y Propuestas para Conciliar la Teoría de la Validez y las Prácticas de Validación

**Coordina:** José Luis Padilla y Stephen Sireci

**Resumen.** Desde la última edición de los Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014), hay un amplio consenso sobre la definición de validez y los contenidos actuales de la teoría de la validez. Al mismo tiempo, hay aún un desfase entre los contenidos de la teoría de la validez y las prácticas y evaluaciones de validez que se llevan a cabo con fines de investigación o aplicados en diferentes ámbitos psicológicos y educativos. A este desfase, hay que sumar los retos que plantean a la teoría y prácticas de validación los cambios tecnológicos en los modos de administración, desarrollos psicométricos como los tests adaptativos computerizados, la evaluación psicológica en el contexto organizacional, los juicios de valor asociados a la identidad de género, las consecuencias del uso de los tests para los diferentes agentes implicados, etc. Por otro lado, abordar un proyecto de validación desde el marco teórico de "validación por argumentos" puede resultar complejo. El objetivo general del simposio es abordar y plantear propuestas para conciliar teoría de validez y prácticas de validación a partir de la experiencia de las y los participantes en diferentes contextos aplicados y de investigación. Cada aportación intentará responder a dos cuestiones generales: a) ¿Qué contenido de la teoría actual de la validez resulta más útil para mejorar la práctica de validación?; y b) ¿cómo se podría superar el desfase entre la teoría y la práctica de validación? Las contribuciones al simposio intentan

responder a las dos preguntas a partir de trabajos de validación en sus diferentes contextos, o proponiendo un enfoque más general para abordarlas.

### ¿Alguna vez ha diseñado e implementado un proceso completo de validación psicométrica? Albert Sesé<sup>1</sup>

<sup>1</sup>Universitat de les Illes Balears

La medida de lo psicológico, latente, inobservable... requiere de validación rigurosa, si por validar entendemos, por ejemplo, lo que la RAE expresa genéricamente como "dar fuerza o firmeza a algo". Hemos de dar firmeza a las medidas de las variables psicológicas porque de ellas depende un adecuado conocimiento de los procesos en que operan, y en consecuencia, la calidad de las posteriores intervenciones psicológicas. Después de que Joseph Lee Cronbach proclamara hace ya casi 70 años la validez como un concepto unitario, no sin oposición ni caza de brujas, y pusiera las bases de lo que conocemos como validación psicométrica, la validez sigue siendo la clave de bóveda de la Psicometría. Aunque existe un cierto consenso sobre la definición teórica contemporánea de la validez, la pléyade de procedimientos disponibles para obtener evidencias de validez de las puntuaciones de un test hace que un proceso de validación psicométrica se convierta en un complejo diseño de investigación específico con múltiples brazos y requerimientos. Se analizarán barreras y facilitadores para su aplicación práctica, tales como el papel del creciente desarrollo actual de técnicas analítico-estadísticas cada vez más complejas, sobre todo en la consideración de estructuras latentes, que pueden estar actuando como pensamiento mágico sobre la calidad de las medidas. Se plantea la necesaria vuelta hacia el cultivo del arte de crear ítems, así como la aplicación de estrategias de validez de decisión mediante gold standards, de uso realmente ínfimo en la literatura, con relación al resto de fuentes propuestas de evidencias de validez. El impacto de las nuevas tecnologías, sobre todo, el Big Data y la Realidad Virtual sobre la validez de las medidas está por concretar en los tiempos venideros.

### Planificación y desarrollo de estudios de validación: una guía práctica. Isabel Benítez<sup>1</sup>, Steve Sireci<sup>2</sup>

<sup>1</sup>Universidad de Granada, <sup>2</sup>University of Massachusetts Amherst

A pesar de que la planificación y el desarrollo de estudios de validación son prácticas habituales en el contexto de la investigación, en muchas ocasiones los procesos realizados no reflejan las necesidades reales del instrumento de evaluación que se está analizando. Uno de los motivos es la falta de directrices sobre los pasos que deben seguirse para obtener evidencias de validez que apoyen el uso y las interpretaciones previstas de las

puntuaciones. Aunque existen numerosas definiciones de validez e incluso indicaciones sobre cómo obtener evidencias de validez específicas, son escasas las propuestas prácticas centradas en guiar a los investigadores en la planificación y el desarrollo de estudios de validación. Entre las guías disponibles destaca la aproximación basada en argumentos de Kane, que aporta una herramienta útil para identificar las limitaciones del instrumento, aunque no plantea ideas concretas para la búsqueda de evidencias de validez. Se presentará una guía práctica para los investigadores que deseen realizar un estudio de validación, incorporando detalles sobre cómo identificar las limitaciones del test y cómo utilizar las fuentes de evidencias disponibles. Se retoma la propuesta de Kane cuyos elementos clave son el uso y las interpretaciones previstas de las puntuaciones del test, y se proponen indicaciones e ilustraciones adicionales que pretenden facilitar la obtención de aquellas evidencias de validez de una manera comprensiva. Se presentarán los pasos propuestos, los resultados previstos y se plantearán propuestas que permitan interpretar los datos en términos de evidencias de validez. Por último, se discutirá la importancia de la definición actual de las fuentes de validez para abordar los procesos de validación de una manera reflexiva y práctica para los usuarios de los test.

**Escala de digitalización del trabajo: elaboración y validación.** Vicente González-Romá<sup>1</sup>, Inés Tomás<sup>1</sup>, Ana Hernández<sup>1</sup>, Marija Davcheva<sup>1</sup>

<sup>1</sup>Universitat de València

La digitalización del trabajo (DT, grado en el que los trabajadores utilizan tecnologías digitales para realizar sus trabajos) es un fenómeno creciente que puede afectar a los puestos de trabajo y tener consecuencias para los trabajadores. Sin embargo, son escasos los estudios empíricos sobre DT. Esta escasez puede atribuirse a la falta de instrumentos adecuados para medirla. El objetivo de este estudio es proponer una escala de DT y obtener evidencias de validez. Partiendo de la definición del constructo, se elaboró un banco inicial de 13 ítems. Un grupo de 13 expertos evaluó la idoneidad de cada ítem. Los ítems evaluados positivamente por al menos el 80% de los expertos fueron seleccionados y mejorados siguiendo sus sugerencias, resultando una escala de 9 ítems. Para estudiar las propiedades psicométricas de la escala se planificaron dos estudios con dos muestras. ESTUDIO1: La muestra estuvo compuesta por 204 trabajadores. Con una mitad de la muestra se realizó un Análisis Factorial Exploratorio. Los resultados apoyaron un modelo unifactorial. Este modelo fue validado en la otra mitad mediante un Análisis Factorial Confirmatorio. Los resultados también apoyaron el modelo unifactorial. El coeficiente alfa de Cronbach fue 0.86. Según las hipótesis de validación basadas en relaciones con otras

variables, se esperaba observar correlaciones significativas entre DT y distintas características de los puestos de trabajo: autonomía, feedback, variedad, significado, e identidad. Los resultados apoyaron parcialmente las hipótesis, con correlaciones significativas para 3 de las 5 características. En el Estudio 2 se espera replicar los resultados del primero y obtener evidencias de validez convergente y discriminante. Este estudio ejemplifica la conveniencia y los retos de integrar múltiples fuentes de evidencias.

**La validez en la aplicación de modelos psicométricos complejos: más allá del ajuste estadístico.** Francisco J. Abad<sup>1</sup>, Miguel A. Sorrel<sup>1</sup>, Pablo Nájera<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Los Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014) mantienen la idea de tipos o fuentes de evidencia de validez (contenido, procesos de respuesta, estructura interna, relaciones con otras variables, consecuencias de las pruebas). No obstante, en la visión contemporánea de la validez, se pone énfasis en que esas evidencias se refieren no al test, sino a los particulares usos e interpretaciones de las puntuaciones. En este contexto, cobra peso el marco teórico de la “validación por argumentos”, en el que se promueve una mayor especificidad para las inferencias. En la presente comunicación se abordará la potencial utilidad de ese marco teórico en tres contextos de investigación relacionados con la evaluación óptima (i.e., modelado del estilo de respuesta, test adaptativos informatizados y modelos de diagnóstico cognitivo para medir competencias). Con frecuencia se recogen evidencias de validez únicamente con relación a la estructura interna o a las relaciones con otras variables. Se ilustrará la importancia de una aproximación integral, que no olvide al sujeto psicológico, así como el alcance y el riesgo de abordar la validación desde una aproximación centrada únicamente en el ajuste estadístico. Se intentará dar respuesta a las dos cuestiones del simposio, en el marco de los contextos de investigación descritos.

**Retos de la validación de los procesos de respuesta a los ítems y preguntas de encuesta en cuestionarios “web”.**

José Luis Padilla<sup>1</sup>, Isabel Benítez<sup>1</sup>, Luis Manuel Lozano<sup>1</sup>, María del Carmen Aguilar<sup>1</sup>

<sup>1</sup>Universidad de Granada

El creciente uso de modos de administración de tests y cuestionarios diferentes a los tradicionales supone un reto al analizar las propiedades métricas de las mediciones. El cambio del formato de “papel y lápiz” y modos de administración “cara a cara” al uso de plataformas online y dispositivos móviles, requiere de nuevos métodos de validación y una visión más

comprehensiva de los procesos cognitivos y sociales de respuesta a los ítems. La metodología de encuesta puede aportar herramientas conceptuales y métodos de validación de los que puede beneficiarse la psicometría. El objetivo de la presentación es mostrar la base conceptual, tipos de evidencias de los procesos de respuesta y desafíos que el “web probing” plantea como método de validación para las mediciones aportadas por cuestionarios implementados en plataformas online. Los tipos de “probes” y sus modos de implementación, las evidencias de validez sobre la “ecología” del proceso cognitivo y social de “pregunta-y-respuesta”, y las estrategias de análisis se ilustrarán con los estudios en marcha de un proyecto de investigación para optimizar las medidas de calidad de vida y salud, y la validación de las mediciones de una escala de “eco-ansiedad”. El análisis sobre la calidad de las evidencias de validez aportadas por el “web probing” con los resultados de un estudio online con 1114 participantes, 559 de Alemania y 555 de España, que respondieron a una pregunta sobre felicidad de la octava edición de la European Social Survey. La presentación también presentará como integrar en un estudio “mixto” de validación las evidencias cualitativas aportadas por el “web probing” con el análisis de las variables sociodemográficas y los estadísticos tradicionales del análisis psicométrico.

## Mesa 1

### **Adaptación de la Escala de Carnismo 4N al contexto español: fiabilidad y estudios de validación basados en la estructura interna y en relación con otras variables.**

Claudia Suárez-Yera<sup>1</sup>, Jorge L. Ordóñez-Carrasco<sup>1</sup>, María Sánchez-Castelló<sup>1</sup>, Antonio J. Rojas Tejada<sup>1</sup>

<sup>1</sup>Universidad de Almería

La creencia de que algunos animales son más valiosos que otros (especismo) se ve influida por un sistema de creencias que legitima el matar animales para su consumo recurriendo a una serie de justificaciones (carnismo). Piazza y colaboradores crearon una escala (Escala 4N) para recoger esta variable, en la que engloban estas justificaciones en las llamadas cuatro es: comer carne es Natural, Necesario, Normal y placentero [Nice]. El objetivo de este trabajo es adaptar la Escala 4N, utilizando las directrices de la International Test Commission, al contexto español y estimar sus propiedades psicométricas. La escala se administró vía online a una muestra de 470 participantes. Se consideró la dieta como cuota de selección de participantes (248 omnívoros y 222 vegetarianos/veganos), con edades de 18-68 años, y un 68,7% de mujeres. Para obtener evidencias de validez relacionadas con la estructura interna se probaron, con análisis factorial confirmatorio, un modelo unidimensional, de cuatro dimensiones y de cuatro

dimensiones con un factor de segundo orden. La estimación de la fiabilidad de las puntuaciones de la escala fue adecuada. Como evidencias de validez en relación con otras variables se correlacionó las puntuaciones totales de la escala 4N y las variables de especismo, y actitud hacia los animales, resultando positivas y altas. También se encontraron diferencias estadísticamente significativas entre omnívoros y vegetarianos/veganos (los primeros con mayor puntuación en la escala). Los resultados obtenidos son similares a los expuestos originariamente por los autores de la escala. No obstante, consideramos que esta medida precisa mejoras, especialmente en dos ítems de la subescala Normal.

### **Propiedades Psicométricas del Cuestionario de Sentido de la Vida en Población Española.** Rocío Gómez-Molinero<sup>1,2,3</sup>, Manuel García-Sedeño<sup>1</sup>, Teresa Lozano-Antolínez<sup>1</sup>, Rocío Guil<sup>1,3,4</sup> y María Victoria Cerezo<sup>2,5</sup>

<sup>1</sup>Universidad de Cádiz, <sup>2</sup>Universidad de Educación a Distancia (UNED)

<sup>3</sup>Instituto Universitario de Investigación para el desarrollo Social Sostenible (INDESS), <sup>4</sup>Instituto de Investigación Biomédicas de Cádiz (INIBICA), <sup>5</sup>Universidad de Málaga

El sentido de la vida se ha propuesto como uno de los factores protectores más importantes para prevenir el desajuste psicológico tras las crisis. La relevancia de la construcción de sentido después de la pandemia de COVID-19 exige el desarrollo de instrumentos precisos para evaluarla. Una de las escalas más válidas y fiables para evaluar la búsqueda y existencia de sentido vital es el Cuestionario de Sentido en la Vida (MLQ) de 10 ítems desarrollado por Steger. El presente estudio se realizó para aportar evidencia psicométrica sobre el MLQ en población española, incluyendo validez y fiabilidad en base a la estructura interna y a las relaciones con la ansiedad y la depresión. La muestra estuvo compuesta por 527 españoles, siendo el 73% mujeres, que completaron una encuesta online incluyendo el MLQ y una escala de ansiedad y depresión. El MLQ comprende dos subescalas, Presencia, que evalúa el grado en que la vida se percibe como significativa y Búsqueda, que mide la motivación para descubrir el sentido vital. La estructura factorial MLQ y la validez de constructo se probaron mediante análisis factorial confirmatorio y correlación de Pearson. El análisis factorial confirmatorio apoyó la estructura de dos factores para el MLQ, con un RMSEA por debajo del valor máximo aceptable de .08 mostrando una bondad de ajuste, usando RMSEA, un ajuste aceptable para el modelo [ $\chi^2(35)=125.14$ ,  $p<.001$ , CFI=.97, TLI=.96, RMSEA=.056]. La fiabilidad de la puntuación del test se calculó mediante el omega de McDonald's obteniendo un valor de .88 tanto para presencia como para búsqueda. La evidencia de validez basada en las relaciones con otras variables, como la

ansiedad y la depresión, reveló correlaciones satisfactorias. La versión española del MLQ mostró propiedades psicométricas satisfactorias que pueden ser aplicadas en la investigación sobre el sentido de la vida en población general española.

### **Estudio de la calidad métrica de la escala de Estrés Percibido PSS-10, desde la Teoría de la Respuesta a los Ítems en una muestra de pacientes de cáncer de mama.**

María Victoria Cerezo<sup>1,2</sup>, Lorena M. Soria-Reyes<sup>1</sup>, Lucía Palacios<sup>2</sup> y Rocío Gómez-Molinero<sup>2,3,4</sup>

<sup>1</sup>Universidad de Málaga, <sup>2</sup>Universidad de Educación a Distancia (UNED), <sup>3</sup>Universidad de Cádiz, <sup>4</sup>Instituto Universitario de Investigación para el desarrollo Social Sostenible (INDESS)

Conocer y evaluar adecuadamente el estrés que padecen las pacientes de cáncer de mama puede ayudar a diseñar intervenciones psicológicas especializadas. Uno de los instrumentos más utilizados es la escala de estrés Percibido, PSS-10. Sin embargo, la mayoría de los estudios psicométricos están enfocados desde la teoría clásica de los test (TCT), siendo escasos los realizados desde la teoría de la respuesta al ítem (TRI) (Reyna et al., 2019), ninguno con pacientes de cáncer de mama. La TRI supera algunas de las limitaciones de la TCT al aportar más información sobre la calidad métrica de los ítems (Barbero et al., 2001; Muñiz, 2010). El objetivo del estudio ha sido analizar la calidad métrica de la PSS-10, desde la perspectiva de la TRI en una muestra de pacientes de cáncer de mama. Las participantes han sido 215 mujeres diagnosticadas de cáncer de mama que cumplimentaron la escala PSS-10. Se planteó el análisis desde el modelo de crédito parcial y el modelo de Rash, comprobando el supuesto de dependencia local y realizando un análisis de los ítems. Los resultados revelan que la escala PSS-10 muestra mejor ajuste al modelo de Rasch. Los parámetros  $b$  en la dimensión de Estrés Percibido oscilan entre 0.68 y 2.31 con errores de estimación comprendidos entre 0.249 y 0.285. Para la dimensión Afrontamiento los valores de  $b$  oscilan entre 1.50 y 2.68 con errores de estimación entre 0.244 y 0.307. La función de información de la dimensión Estrés Percibido, aporta mayor información para valores de estrés entre -2 y 0; para la dimensión de Afrontamiento, la función es más informativa para valores entre -3 a -1. Como conclusión, la escala PSS-10 para estas pacientes se ajusta adecuadamente al modelo de Rasch y ofrece mayor información para valores bajos tanto de estrés como de afrontamiento al estrés.

### **Análisis de perfiles latentes del Personality Inventory for DSM-5 (PID-5) y relación con síntomas psicopatológicos.**

Ana de la Rosa<sup>1</sup>, Carmen Díaz-Batanero<sup>1</sup>, Gina Rossi<sup>2</sup>, Morag Facon<sup>2</sup>, Manuel Sánchez<sup>1</sup>

<sup>1</sup>Universidad de Huelva, <sup>2</sup>Universidad Libre de Bruselas

La identificación de subtipos de perfiles de pacientes, más allá de las categorías tradicionales, permite el diseño de intervenciones clínicas aplicables a conjuntos de pacientes, desde una perspectiva transdiagnóstica. El presente trabajo analiza los subtipos de perfiles de personalidad dentro del Modelo Alternativo de Personalidad del DSM-5 y su relación con síntomas internalizantes y externalizantes. 742 participantes comunitarios fueron seleccionados aleatoriamente y 312 pacientes en centros de atención de salud mental públicos y privados. Se utilizó el Inventario de Personalidad para el DSM-5 (PID-5), para medir las facetas de la personalidad, y el IDAS-II y ESI-BF para medir síntomas internalizantes y externalizantes respectivamente. Se realizó un análisis de perfiles latentes para determinar la presencia de perfiles distintos de funcionamiento de la personalidad. Los perfiles latentes de la solución retenida se contrastaron con las variables sociodemográficas y los síntomas internalizantes y externalizantes. Una solución de 4 perfiles proporcionó el mejor ajuste a los datos. La primera clase se caracterizó como grupo con buen ajuste (44% de los datos); la segunda (36% de los participantes) puntuó alto en ansiedad, labilidad emocional, inseguridad en la separación. La tercera clase (10% de los participantes) estaba relacionada con facetas vinculadas con el trastorno esquizotípico y trastorno evitativo y la cuarta clase (13%) agrupaba a los participantes con puntuaciones más altas en facetas relacionadas la impulsividad, insensibilidad o manipulación. Los síntomas internalizantes se asociaron mayoritariamente a la tercera clase, mientras que los externalizantes se asociaron a la cuarta clase. Este trabajo está financiado por el Programa Operativo FEDER Andalucía 2014-2020 (proyecto UHU-1257470), el Ministerio de Educación (PRX21/00319) y el Ministerio de Universidades (FPU19/00144).

### **Aplicación de modelos de TRI para la formalización y la medición del proceso de estereotipia.** Pablo Sayans-Jiménez<sup>1,2</sup>, Colin W. Leach<sup>3</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería, <sup>3</sup>Barnard College, Columbia University

La medición empleando el modelo del contenido del estereotipo (SCM) ha presentado evidencias empíricas desfavorables relacionadas con la estructura interna, ausencia de invarianza ante diferentes grupos objetivo y se ha mantenido ajena a las limitaciones de la teoría clásica para reflejar los procesos de estereotipia. En este trabajo se integrarán las evidencias empíricas discordantes y teorías complementarias al SCM, junto con la aplicación de modelos de TRI, para generar nuevas propuestas teóricas y psicométricas. En el estudio 1 se analiza la asignación de nueve atributos (tres referidos a agencia, tres a benevolencia-general y tres a

benevolencia-cooperativa), en cuatro muestras (n = 550, 300, 300, 349), a cinco grupos objetivo (dos muestras hacia gitanos, una hacia bomberos y la cuarta hacia multimillonarios y personas con Síndrome de Down). Se incluyó un diferencial semántico (DS) para aportar evidencias sobre la interpretación de las puntuaciones de la dimensión principal. Se implementaron modelos bifactoriales de Rasch (mínimos cuadrados ponderados con media y varianza ajustada, parametrización theta). Los resultados muestran invarianza para los dos grupos que atribuían estereotipos a gitanos, ausencia de invarianza entre el resto de los grupos y correlaciones moderadas-altas entre la dimensión general y del DS. Los atributos se agrupan por contenido, pero su posición depende de la combinación de las características más relevantes de cada grupo. En el estudio 2 se realizó el mismo tipo de análisis con grupos simulados para controlar las características más salientes de cada grupo objetivo en dos nuevas muestras (n = 300, 442). Los resultados confirman y matizan las conclusiones del estudio 1. Se aprecia la influencia de la valencia de las características definitorias de cada grupo en las posiciones en el continuo de los atributos (e.g., grupo agente-benevolente vs. agente-inmoral). Estos resultados se explicarían por los beneficios/perjuicios que puede implicar una interacción con el grupo objetivo.

## Mesa 2

### **Análisis de la fase ofensiva en fútbol femenino a partir de entrevistas a jugadoras de élite: un enfoque de método mixto.** Rubén Maneiro<sup>1</sup>, Iván-Baragaño<sup>2</sup>

<sup>1</sup>Universidad Pontificia de Salamanca, <sup>2</sup>Universidade da Coruña

La observación indirecta es un concepto reciente en la observación sistemática. Implica en gran medida analizar material textual generado indirectamente a partir de transcripciones de grabaciones de audio de comportamiento verbal en entornos naturales (p. ej., conversaciones, discusiones grupales) o directamente a partir de narraciones (p. ej., cartas de queja, tweets, publicaciones en foros) (Anguera et al., 2018). Todos estos materiales constituyen una fuente de información para el estudio de la vida cotidiana, y están en continuo crecimiento con el surgimiento de nuevas tecnologías para el registro, difusión y almacenamiento de datos. En el presente estudio, se utilizó la observación indirecta para conocer los planteamientos de las jugadoras de la selección española de fútbol sobre las dinámicas y procesos ofensivos en el fútbol femenino de alto rendimiento. Para ello utilizamos un método mixto basado en la observación sistemática de transcripciones textuales de entrevistas realizadas a las jugadoras y entrenadores. El diseño observacional fue nomotético, puntual y multidimensional. Una vez confirmada la fiabilidad intraobservador del instrumento, realizamos un

análisis de coordenadas polares para identificar las relaciones significativas entre las valoraciones de los entrenadores y los elementos ofensivos del ataque en fútbol femenino. Como resultados, fueron registradas un total de 2410 unidades textuales. La conducta tomada como target y focal para el análisis secuencial de retardos (LSA) y análisis de coordenadas polares (PCA) respectivamente fue D2281: Vinculación al éxito en una acción ofensiva o en el juego de manera positiva. El análisis secuencial de retardos refleja que un total de 20 conductas condicionadas del instrumento de observación indirecta presentaron una asociación estadísticamente significativa a partir del valor del residuo ajustado. Las principales conclusiones de este estudio corroboran trabajos previos sobre el enorme interés y potencialidad que puede presentar la observación indirecta (dentro del paradigma mixed methods), en el estudio y análisis del fútbol.

### **Relación entre ansiedad a las matemáticas y nota obtenida en estadística, en distintos grados universitarios.** Jaume March-Llanes<sup>1</sup>, Verónica Estrada-Plana<sup>1</sup>, Núria Vita-Barrull<sup>1</sup>, Elena Lacomba-Arnau<sup>1</sup>, Jorge Moya-Higueras<sup>1</sup>

<sup>1</sup>Universitat de Lleida

En el presente estudio se ha utilizado la versión en español del cuestionario sMARS, originalmente validado por M. Isabel Núñez-Peña et al. en 2013 para comparar el patrón de respuestas en distintos grados universitarios y relacionarlos con las notas obtenidas en asignaturas de estadística. El cuestionario sMARS contiene 3 factores: (a) Ansiedad ante los exámenes de matemáticas, definida por 15 ítems que reflejan la aprehensión por tomar un examen de matemáticas o por recibir los resultados de los exámenes de matemáticas; (b) Ansiedad ante la tarea numérica, definida por 5 ítems que reflejan la ansiedad por realizar operaciones numéricas; y (c) Ansiedad por el curso de matemáticas, definida por 5 ítems que reflejan ansiedad sobre clases de matemáticas. La muestra de estudio estuvo formada por 92 alumnos de 3 grados distintos de la UdL, Biotecnología (1º curso), Biomedicina y Psicología (2º curso), matriculados en las asignaturas de Bioestadística o Estadística Inferencial. Además, otro grupo de Psicología estaba formado por alumnos matriculados en la asignatura de Metodología (1º curso). Los resultados han proporcionado evidencias de una suficiente validez de estructura interna en la muestra de la UdL. Además, han permitido sacar distintas conclusiones sobre la relación entre las notas obtenidas en las distintas asignaturas y la ansiedad hacia las matemáticas manifestada en el cuestionario. Los dos principales resultados son que, a pesar de la poca diferencia entre edades de los sujetos, la variable edad de los participantes parece influir de forma determinante en la capacidad predictiva de la ansiedad matemática sobre la nota en la asignatura de estadística. El otro resultado a

destacar es que esta relación entre ansiedad matemática y nota obtenida, se centra exclusivamente en el primer factor.

**Aplicación de Modelos de Cambio Latente para el estudio de la respuesta temprana al tratamiento en anorexia nerviosa.** Nuria Real-Brioso<sup>1</sup>, Ani Laura Ruiz-Lee<sup>1</sup>, Bronwyn C. Raykos<sup>2</sup>, David M. Erceg-Hurn<sup>2</sup>, Ricardo Olmos<sup>1</sup>, Eduardo Estrada<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Centre for Clinical Interventions

La respuesta temprana al tratamiento es el predictor más consolidado del éxito de la intervención en personas con un trastorno de la conducta alimentaria (TCA). Estudios recientes señalan la falta de investigaciones más exhaustivas de este fenómeno. Por su parte, los modelos de Cambio Latente (Latent Change Score, LCS) proporcionan una potente herramienta para estudiar la respuesta temprana debido a su capacidad para analizar dicho fenómeno como un sistema dinámico. En este trabajo analizamos dos indicadores comunes de respuesta temprana en anorexia nerviosa: el índice de masa corporal (IMC) y la puntuación global en el Eating Disorder Examination Questionnaire (EDE-Q). Éstos se midieron en una muestra de 192 personas con anorexia nerviosa durante las primeras semanas de una intervención cognitivo-conductual. Usando modelos LCS bivariados, encontramos que a) los procesos de cambio de los dos indicadores parecen ser independientes entre ellos, b) el cambio en el IMC es principalmente lineal; y c) el cambio en el EDE-Q es parcialmente lineal, y parcialmente predicho por su propia tasa de cambio en las semanas anteriores (es decir, por un componente de auto-regresión,  $\phi_Y$ , que captura la influencia del cambio latente durante el intervalo anterior ( $t-1$ ) hacia el cambio posterior ( $t$ ) en la misma variable). Discutimos las implicaciones clínicas y las consideraciones metodológicas de los hallazgos. El presente trabajo evidencia el potencial de los modelos dinámicos en la investigación de los trastornos alimentarios y los cambios que se pueden producir en el contexto de una intervención clínica.

**Obtención de un modelo de competencias académicas: integración a través de la visualización conjunta de datos.** Juan F. Luesia<sup>1</sup>, Isabel Benítez<sup>2,3</sup>, Milagrosa Sánchez-Martín<sup>1</sup>

<sup>1</sup>Universidad Loyola Andalucía, <sup>2</sup>Universidad de Granada, <sup>3</sup>Centro de Investigación Mente Cerebro y Comportamiento (CIMCYC)

Introducción. Estudiar las competencias relevantes en el éxito académico supone un objetivo central para entidades de educación superior. Aunque, tradicionalmente, los estudios cuantitativos han realizado la mayor parte de aportaciones, recientemente los

estudios cualitativos y mixtos han ganado protagonismo. No obstante, la integración de datos cuantitativos y cualitativos en estudios mixtos sigue siendo un reto para la investigación. El objetivo de este trabajo es mostrar cómo la visualización conjunta de datos facilita la interpretación de un modelo de competencias académicas obtenido mediante un diseño mixto. Método. Se diseñó un estudio "QUAN/QUAL" de dos fases (cuantitativa y cualitativa) concurrentes. En la fase cuantitativa, se evaluaron las competencias de 2.370 estudiantes y su rendimiento académico durante cuatro años. Se analizó la regresión para evaluar la capacidad predictiva de dichas competencias sobre el rendimiento. En la fase cualitativa, se desarrollaron cuatro grupos focales en los que participaron 30 expertos que debatieron sobre competencias relevantes en el éxito académico. Se analizaron y codificaron las narrativas de los expertos, identificando las competencias determinantes en el rendimiento de los estudiantes. Resultados. Se integraron los resultados a través de visualización conjunta de datos. Por un lado, hubo convergencia en cuanto a algunas variables cognitivas, como atención y ortografía, relevantes en ambas fases. Por otro lado, en la fase cualitativa se aportaron 25 nuevas competencias no cognitivas relevantes, como motivación intrínseca, estrategias de estudio o comunicación. Discusión. La visualización conjunta de datos ha permitido integrar un modelo emergente de competencias cognitivas y no cognitivas, lo que muestra la utilidad de la aproximación mixta y aporta herramientas metodológicas que podrían aplicarse en el estudio de distintos fenómenos en ciencias sociales.

**Autoconcepto y cultura: validación de medidas para la investigación transcultural.** Óscar F. García<sup>1</sup>, Celeste León-Moreno<sup>2</sup>, Iria Palacios<sup>1</sup>, Marta Alcaide<sup>3</sup> y Fernando García<sup>3</sup>

<sup>1</sup>Universidad de Valencia, <sup>2</sup>Universidad de Zaragoza, <sup>3</sup>Universidad de Valencia

El autoconcepto representa la autopercepción de un individuo formada por las experiencias con otras personas significativas, las atribuciones del propio comportamiento y los refuerzos del entorno. Entre las diferentes dimensiones del autoconcepto, suele incluirse la académica y otras no académicas (social, emocional y familiar, esta última menos estudiada). Aunque el autoconcepto ha sido ampliamente utilizado en la investigación por su importancia en el desarrollo psicosocial de niños, adolescentes y adultos, muchos estudios transculturales utilizan medidas cuya invarianza no ha sido comprobada. La validación transcultural de las medidas es especialmente relevante para la investigación en la que se comparan culturas tan diferentes como la de contextos occidentales (e.g., Europa, Estados Unidos y Australia) y orientales (e.g., China). El cuestionario Autoconcepto Forma 5 (AF5) es una de las medidas de



autoconcepto más populares y utilizadas en la práctica profesional y en investigación, cuestionario que ha sido ampliamente validado en muestras de Europa, América Latina y Estados Unidos. La validación del AF5 en China facilitaría especialmente las comparaciones transculturales entre el país asiático de mayor población (i.e., China) y las sociedades occidentales. Para la validación de la versión china del autoconcepto se aplicaron análisis de tipo factorial confirmatorio multigrupo. Los participantes fueron cerca de 2.000 adolescentes y adultos jóvenes, tanto de China como de España. En general, los resultados mostraron que la estructura factorial basada en de cinco dimensiones no relacionadas (i.e., académico, social, emocional, familiar y físico) presentaba un ajuste a los datos igual o mayor que la alternativa ortogonal y unidimensional. La versión china de la medida del autoconcepto AF5 mantuvo los pesos de los ítems en los factores, la varianza de los factores y la covarianza entre los factores en relación con la versión original en español. Esta validación supone un avance importante para la investigación transcultural del autoconcepto.

### Mesa 3

**Sobrefactorización en las escalas Likert con datos asimétricos: una comparación entre análisis factorial y teoría de respuesta al ítem.** Noelia Minaya<sup>1</sup> y Javier Revuelta<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

La evaluación psicológica suele basarse en cuestionarios con escalas de respuesta dicotómicas u ordinales. Cuando la media de los indicadores se acerca a los extremos pueden aparecer diferentes patrones de asimetría en los datos (positiva, negativa o mixta) y atenuación en el valor de las correlaciones entre los ítems. Como consecuencia, al estudiar la dimensionalidad de la escala desde un modelo de análisis factorial, es posible que se estime un número mayor del necesario de dimensiones latentes. El objetivo de este trabajo es estudiar el fenómeno de la sobrefactorización tomando como base el modelo factorial tradicional, el modelo de análisis factorial categórico basado en las correlaciones policóricas y el modelo de Samejima de la teoría de respuesta al ítem. Como criterio de ajuste se tomó el estadístico chi-cuadrado de verosimilitud. El diseño de la simulación se basa en Olsson (1979). El número de replicas por condición fue 1000. Las condiciones manipuladas son el tamaño muestral, la estructura de las cargas factoriales, la asimetría, el número de umbrales y el número de variables observadas. El número total de condiciones simuladas es 240. Tanto el modelo factorial tradicional como el modelo factorial para variables categóricas presentan tasas de error tipo I inadmisibles, siendo las condiciones de simulación con mayor tamaño del efecto

sobre el logaritmo del p valor la asimetría y el patrón de cargas factoriales. En contraste, el modelo de Samejima presenta una mayor robustez respecto a las condiciones de simulación manipuladas, así como valores de EPR más cercanos al nivel de significación utilizado.

**Modelos Doble-Explicativos Bayesianos de dos-Parámetros: Estudios Empírico y de Simulación.** Ignacio Castillejo<sup>1</sup> y Daniel Ondé<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad Complutense de Madrid

Los modelos doble-explicativos de respuesta al ítem surgen de entender los modelos de variable latente como un caso particular de la regresión multinivel generalizada. Hablamos de modelos explicativos cuando los parámetros se descomponen en regresiones con efectos aleatorios que se producen en el nivel de los ítems o de las personas y de modelos doble-explicativos si se incluyen en ambos niveles simultáneamente. Su utilidad reside en la capacidad para incorporar información adicional a los modelos psicométricos, de modo que ha servido en el estudio de la validez de procesos de respuesta, el desarrollo de algoritmos de generación automática de ítems y la mejora de la estimación de parámetros en Tests Adaptativos Informatizados. En el presente trabajo se ilustra la aplicación de una extensión doble-explicativa del modelo logístico de 2 parámetros mediante los modelos lineales mixtos y latentes generalizados (GLLMM) con estimación bayesiana utilizando el paquete brms en R. El objetivo es dar a conocer estos modelos y su aplicación en contextos de validación de instrumentos, comenzando con una introducción de las nociones básicas de los GLLMM y su relación con la Teoría de Respuesta al Ítem. Se presentan dos estudios: el primero de simulación para valorar la precisión de las estimaciones, y el segundo para mostrar una aplicación con datos reales. Se discuten ventajas y limitaciones de estos modelos en función de los resultados obtenidos.

**Avances en evaluación de estructura interna mediante Análisis Factorial Confirmatorio: Una comparación entre el modelo bifactor y el modelo bifactor S-1.** Daniel Ondé<sup>1</sup> y Jesús M. Alvarado<sup>1</sup>

<sup>1</sup>Universidad Complutense de Madrid

El problema de las evidencias de validez referida a estructura interna es bien conocido. Ocupa un lugar destacado entre los distintos aspectos a evaluar en todo proceso de elaboración de instrumentos de medición de variables o constructos psicológicos (AERA, APA, & NCME, 2014). También ha dado lugar a numerosas publicaciones que tratan de establecer, bajo expresiones generales del tipo “validación de instrumentos” o “propiedades psicométricas”, la dimensionalidad subyacente a las

respuestas de los participantes. A pesar de la importancia que se confiere a este tipo de validez, aun a día de hoy proliferan los estudios en los que las conclusiones se construyen priorizando los resultados sobre ajuste de los modelos estadísticos (generalmente, elaborados mediante análisis factorial – AF), en detrimento de una reflexión más profunda que trate de conectar dichos modelos con la teoría del constructo. A nuestro juicio, un reflejo de esta situación es el creciente interés que ha existido en las últimas décadas por aplicar modelos estadísticos como el modelo bifactor, que consiguen mantener bajo control el desajuste de modelos clásicos como el de un factor o el de factores correlacionados, pero que no contribuyen necesariamente a potenciar la reflexión de carácter más teórico (i.e., falta de conexión entre el factor general y la teoría del constructo). El objetivo de este trabajo es presentar una aplicación con datos reales del modelo bifactor y del modelo bifactor S-1 desde una aproximación confirmatoria. Los modelos S-1 se caracterizan por utilizar uno de los dominios de contenido como factor general de referencia, en lugar de especificar un factor general extra como hace bifactor. Se comparan los resultados de ambos modelos, y se discuten las principales ventajas del S-1 sobre el bifactor para conectar el modelo estadístico con la teoría subyacente en determinados contextos de evaluación.

**¿Cómo afectan las omisiones a la estimación de modelos de TRI cuando tenemos en cuenta características de los ítems y de los examinados?: Una perspectiva multinivel.** Susana Sanz<sup>1</sup>, Carmen García<sup>1</sup>, Ricardo Olmos<sup>1</sup>, Steffi Pohl<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Freie Universität Berlin

Las omisiones de los examinados en ítems de opción múltiple afectan a la estimación de los parámetros, y, dependiendo del tipo de prueba, la proporción de omisiones será más o menos elevada. Además, los mecanismos responsables de la omisión presumiblemente son no aleatorios, ya que pueden depender de características de los ítems y/o evaluados. Estos mecanismos no se suelen incorporar en la estimación de los parámetros sustantivos del modelo (el nivel de habilidad del sujeto o los parámetros de los ítems), produciéndose un problema de pérdida de datos no ignorable. Cuando esto ocurre, estos parámetros se estiman con sesgo. Por este motivo, considerar un modelo de estimación conjunta de los parámetros sustantivos y de la omisión puede ayudar a una estimación más consistente y eficiente de los parámetros. Aunque desde la perspectiva de los modelos crossed random-effects se ha aunado información relacionada con ítems y participantes, aún queda por encontrar la mejor forma de contemplar las omisiones en ellos. Por tanto, el objetivo de este trabajo es encontrar la estimación más eficiente para el nivel de rasgo de las personas y los parámetros de los ítems incluyendo el

efecto de las omisiones. Para ello, se realizará un estudio de simulación en el que se manipulará la cantidad de omisiones presentes, y la dirección del efecto de las características de los ítems y evaluados en la estimación de los parámetros del modelo, así como en su ajuste. Asimismo, se trabajará con parámetros relativos a los modelos logísticos de TRI de uno, dos y tres parámetros. Se espera encontrar que una estimación previa de la omisión incluyendo covariables que ayuden a explicarla desde los mismos modelos crossed random-effects, y su posterior inserción en el modelo general, proporcionará mejores resultados en la estimación de los parámetros sustantivos del modelo.

**Evaluación de la Dimensionalidad de Estructuras Bifactoriales Generalizadas.** Luis Eduardo Garrido<sup>1</sup>, Marcos Jiménez<sup>2</sup>, Francisco José Abad<sup>2</sup>, Eduardo García-Garzon<sup>3</sup>

<sup>1</sup>Pontificia Universidad Católica Madre y Maestra de República Dominicana, <sup>2</sup>Universidad Autónoma de Madrid, <sup>3</sup>Universidad Camilo José Cela

Actualmente carecemos de información sobre cómo evaluar de manera óptima la dimensionalidad latente de estructuras bifactoriales con uno o más factores generales, a pesar de su importancia en campos diversos como la inteligencia, la personalidad y la psicopatología. Por primera vez, comparamos el rendimiento de métodos de retención de factores tradicionales y novedosos en condiciones realistas del contexto bifactorial y bifactorial generalizado. Para estimar el número de factores de grupo, estos métodos fueron la regla de Kaiser, el criterio empírico de Kaiser, el análisis paralelo con análisis de componentes principales (PAPCA), el análisis paralelo con factorización de ejes principales y el exploratory graph analysis con el algoritmo de agrupamiento de Louvain (EGALV). Luego, estimamos el número de factores generales utilizando las puntuaciones factoriales (PF) de la solución de primer orden sugerida por los dos mejores métodos, lo que produjo las versiones de "segundo orden" PAPCA-PF y EGALV-PF. Además, también examinamos la solución multinivel directa proporcionada por EGALV. Todos los métodos fueron evaluados en una amplia simulación donde se manipularon nueve variables de interés. Los resultados indicaron que EGALV y PAPCA eran los métodos más precisos en la recuperación del número real de factores de grupo. Mientras que el primero fue sensible a cargas cruzadas elevadas, el segundo fue más sensible a factores de grupo pobremente definidos y tamaños de muestra pequeños. Con respecto a la estimación del número de factores generales, tanto PAPCA-PF como EGALV-PF mostraron una precisión casi perfecta en todas las condiciones, mientras que EGALV proporcionó estimaciones mayormente incorrectas. A pesar del desempeño general similar entre PAPCA/PAPCA-

PF y EGALV/EGALV-PF, los métodos basados en EGA fueron robustos en las condiciones más frecuentes en la práctica. Por lo tanto, destacamos la utilidad particular de EGALV y EGALV-PF para evaluar estructuras bifactoriales y bifactoriales generalizadas.

#### Mesa 4

##### **Una aplicación de la evaluación diagnóstica: Medición de competencias en la asignatura de Psicometría.**

Vicente Miguel A. Sorrel<sup>1</sup>, Pablo Nájera<sup>1</sup>, Rodrigo S. Kreichtmann<sup>1</sup>, Francisco J. Abad<sup>1</sup>, Carmen García<sup>1</sup>, Agustín Martínez-Molina<sup>1</sup>, Silvia Collado<sup>2</sup>, José Ángel Martínez-Huertas<sup>3</sup>, José David Moreno<sup>1</sup>, Ricardo Olmos<sup>1</sup>, Susana Sanz<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad de Zaragoza,

<sup>3</sup>Universidad Nacional de Educación a Distancia

Se espera que la evaluación diagnóstica de las competencias adquiera un papel destacado en la enseñanza en los próximos años. A diferencia de la evaluación sumativa, que pretende clasificar a los alumnos en un continuo, la evaluación formativa está orientada a explorar los puntos fuertes y débiles para abordarlos mediante instrucciones específicas. Además, frente al formato tradicional de puntuación consistente en la suma de ítems, el acceso cada vez más extendido a la tecnología permite estimar modelos psicométricos más ajustados al proceso de respuesta. El propósito de esta presentación es ejemplificar esta evaluación formativa mediante el uso de una familia de modelos denominados modelos de diagnóstico cognitivo (MDC). Los datos de autoevaluaciones se recogieron durante el curso 2021/2022 en la asignatura de Psicometría perteneciente al Grado de Psicología de la Universidad Autónoma de Madrid. Estas autoevaluaciones estaban diseñadas para que los/las estudiantes comprobaran su nivel de conocimiento al finalizar cada tema. La muestra estuvo formada por 366 personas y 96 preguntas. Se realizaron análisis propios de la psicometría más clásica (p.ej., dificultad, consistencia interna) y de los MDC (p.ej., ajuste, precisión de la clasificación). Se comparó la información sumativa y formativa y se exploró la capacidad predictiva frente a las pruebas de evaluación. En conclusión, se ilustra cómo es posible extraer información diagnóstica en contextos aplicados. Esta información puede ser relevante para el alumnado, que al ser consciente de sus déficits podrá llevar a cabo una revisión más guiada de los materiales, y para el personal docente, que tienen acceso rápido al rendimiento del grupo en términos de competencias, pudiendo así modificar su enseñanza en consecuencia. El proyecto unifica el acceso cada vez más extendido a la tecnología en el aula con la psicometría moderna para abordar la cuestión fundamental de la evaluación eficaz y precisa.

**Utilidad de la Teoría de Respuesta al Ítem (TRI) y la Teoría Clásica de los Tests (TCT) aplicadas al Brief Symptom Inventory (BSI) para la detección del cambio clínicamente fiable en una muestra de pacientes con patología dual.** Daniel Dacosta-Sánchez<sup>1</sup>, Marta Narváez-Camargo<sup>1</sup>, Andrea Blanc-Molina<sup>1</sup>, Lidia Torres-Rosado<sup>1</sup>, Óscar M. Lozano-Rojas<sup>1</sup>

<sup>1</sup>Universidad de Huelva

El BSI es uno de los instrumentos utilizados en la monitorizar de síntomas emocionales. Numerosos estudios han aportado evidencias psicométricas sobre las propiedades de este instrumento. No obstante, hasta la fecha, no se han encontrado estudios que hayan comparado la sensibilidad al cambio de sus puntuaciones aplicando estas dos teorías psicométricas. En este sentido, los objetivos de este trabajo son: 1) analizar las propiedades psicométricas del BSI, aplicando la TCT y la TRI, en una muestra de pacientes con patología dual; 2) comparar la sensibilidad al cambio de las puntuaciones del BSI estimadas con la TCT y la TRI. En el estudio han participado 97 pacientes diagnosticados de patología dual (trastorno por consumo de sustancias y otro trastorno mental), evaluados en los servicios de adicciones y salud mental de Huelva. El 82.5% eran hombres y la edad media de los participantes era de 41.9 (DT = 11.5) años. Los resultados reflejan adecuadas propiedades psicométricas, estimadas tanto desde la TCT como la TRI. No obstante, la aplicación de la TRI refleja una mayor capacidad para detectar pacientes con cambio clínicamente fiable.

Este trabajo está financiado por la Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía (proyecto P20\_00735), y el Ministerio de Universidades del Gobierno de España (FPU18/00490, FPU20/06606).

##### **Establecimiento de puntos de corte para determinar la disfuncionalidad de síntomas de trastornos emocionales: Resultados preliminares sobre el IDAS-II.**

Daniel Sánchez-Dacosta<sup>1</sup>, Ana de la Rosa-Cáceres<sup>1</sup>, Cinta Mancheño-Velasco<sup>1</sup>, Sandra Tejera-Centeno<sup>1</sup>, Bella González-Ponce<sup>1</sup>

<sup>1</sup>Universidad de Huelva

La aproximación Measurement-Based Care ha evidenciado mejoras en los resultados de tratamiento mediante la aplicación de instrumentos para monitorear el progreso del paciente y guiar las decisiones clínicas. Un aspecto clave en la toma de decisiones es el nivel de funcionalidad del paciente. Aunque cada síntoma tiene un impacto diferencial sobre la funcionalidad, escasos instrumentos ofrecen puntos de corte para determinar la funcionalidad de cada síntoma. El objetivo es identificar

los puntos de corte que discriminan entre puntuaciones funcionales vs. disfuncionales para síntomas de trastornos emocionales evaluados por la versión española del Inventory of Depression and Anxiety Symptoms-II (IDAS-II). Participaron 1072 adultos de la población general seleccionados mediante muestreo aleatorio entre más de 155.000 panelistas de España, y 318 pacientes de los servicios públicos de salud mental de Huelva. El 53.3% (n = 741) de los participantes eran mujeres. Las edades oscilaron entre 18 y 80 años (M = 43.12; DT = 14.76). Para establecer los puntos de corte se empleó el criterio C establecido por Jacobson & Truax (1991). Los puntos de corte para todos los síntomas evaluados por el IDAS-II mostraron una adecuada discriminación entre las puntuaciones de muestras funcionales (población general) y disfuncionales (pacientes). Aunque estos resultados preliminares muestran evidencias favorables sobre la adecuada discriminación de los puntos de corte, se precisa continuar aumentando el tamaño muestral de los pacientes.

Este trabajo está financiado por el Ministerio de Ciencia e Innovación (proyecto PID2020-116187RB-I00), y el Ministerio de Universidades del Gobierno de España (FPU18/00490, FPU19/01413, FPU19/00144).

**Aplicación del protocolo de los 10 pasos para la construcción de un test que evalúe conducta suicida en la adolescencia.** Adriana Díez-Gómez del Casal<sup>1</sup>, Alicia Pérez de Albéniz<sup>1</sup> y Eduardo Fonseca-Pedrero<sup>1</sup>

<sup>1</sup>Universidad de la Rioja

**Introducción:** La conducta suicida supone un problema de salud pública a nivel mundial. Entre la juventud, la muerte por suicidio constituye la primera causa de muerte no natural. Disponer de herramientas construidas y validadas para esta población en nuestro país siguiendo los estándares de calidad supone un hito indispensable para el desarrollo de la investigación en conducta suicida. **Objetivo:** validar una herramienta para población adolescente que permita la evaluación de la conducta suicida y cumpla con las propiedades psicométricas adecuadas basadas en evidencias empíricas. **Método:** se sometió a evaluación la herramienta diseñada siguiendo los diez pasos para la construcción de un test, propuestos por Muñiz y Fonseca-Pedrero (2019). **Resultados:** Las puntuaciones de dicha herramienta mostraron adecuados niveles de fiabilidad. El análisis factorial sugirió un modelo de dos factores, uno general de conducta suicida y otros tres factores específicos (Ideación Suicida, Comunicación Suicida y Acto/Planificación Suicida).

**Conclusiones:** la herramienta de evaluación ha sido construida siguiendo los estándares propuestos por la literatura científica y se erige como una herramienta validada para el estudio de la conducta suicida en población adolescente en España. La conducta suicida

requiere de instrumentos que aporten evidencias empíricas de fiabilidad y validez y permitan, de la misma forma, un análisis más ajustado a la realidad entre los más jóvenes.

## Mesa 5

**Un test estadístico basado en simulación intensiva para evaluar la bondad de ajuste en modelos factoriales exploratorios.** Urbano Lorenzo-Seva<sup>1</sup>

<sup>1</sup>Universitat Rovira i Virgili

Con la finalidad de informar sobre la bondad del ajuste de un modelo factorial confirmatorio, es habitual inspeccionar índices de ajuste basados en el test estadístico chi-cuadrado, el cual, generalmente, es una función de la magnitud de los residuales cuadráticos. Así, por ejemplo, (a) los índices Comparative Fit Index (CFI) o Non-Normed Fit Index (NNFI), que informan sobre el ajuste comparativo de la solución propuesta con respecto al modelo nulo de independencia; y (b) el índice Root Mean Square Error of Approximation (RMSEA), que informa sobre el ajuste relativo del modelo con respecto a su complejidad, se derivan directamente del estadístico chi-cuadrado. En los últimos años, esta práctica se ha extendido al análisis factorial exploratorio, también conocido como análisis factorial no restringido. Ahora bien, la disponibilidad de estos índices está condicionada a la utilización de un procedimiento de extracción factorial cuya base estadística permita el cálculo previo de un estadístico que se distribuya como chi-cuadrado bajo la hipótesis nula (H0). Para solucionar esta limitación, proponemos un procedimiento empírico basado en simulación intensiva que permite calcular dicho estadístico convenientemente escalado respecto al modelo factorial que se espera en la población bajo H0. Nuestro procedimiento permite entonces calcular los índices CFI, NNFI y RMSEA cuando el método de extracción factorial ha sido, por ejemplo, Ejes Principales o Minimum Rank Factor Analysis. Por otra parte, el procedimiento permite también informar sobre la potencia del análisis realizado en base al índice RMSEA. La utilización del procedimiento propuesto se ilustra mediante el análisis de un caso real.

**Considerando la incertidumbre de los parámetros en modelos de diagnóstico cognitivo.** Rodrigo Schames-Kreitchmann<sup>1</sup>, Jimmy de la Torre<sup>2</sup>, Miguel A. Sorrel<sup>1</sup>, Pablo Nájera<sup>1</sup>, & Francisco J. Abad<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>The University of Hong Kong

**Introducción:** Los modelos de diagnóstico cognitivo se utilizan en el ámbito educativo, clínico o de recursos humanos para clasificar a evaluados con respecto a

atributos discretos, buscando identificar fortalezas y debilidades, para así poder ofrecer entrenamiento/tratamiento personalizado. Una correcta estimación de las probabilidades posteriores de las clasificaciones sirve de base para la correcta interpretación de la evaluación. Por ejemplo, en las posteriores se basan los estimadores MAP y EAP de perfil de atributos, bien como sus errores típicos e indicadores de fiabilidad. En este sentido, las posteriores se calculan tradicionalmente utilizando estimaciones puntuales para los parámetros del modelo como aproximaciones a sus valores poblacionales. Siempre que exista incertidumbre acerca de estos parámetros, usar estimaciones puntuales de dichos parámetros como valores poblacionales resultará en distribuciones posteriores demasiado extremas, llevando a una infraestimación de los errores típicos y sobrestimación de la fiabilidad. Así, se presenta un procedimiento de imputación múltiple (IM) para integrar los parámetros del modelo en la estimación de las posteriores. Método: Se ha llevado a cabo un estudio de simulación para investigar el procedimiento de IM propuesto. A efectos prácticos se evalúa su impacto sobre la estimación de la fiabilidad. Como línea de comparación, se incluyen las fiabilidades verdaderas, bien como aquellas calculadas tomando las posteriores usando las estimaciones puntuales de parámetros del modelo (método tradicional). Se han manipulado cinco factores: la estructura de atributos, el modelo generador y estimado (DINA y G-DINA), la longitud del test, el tamaño de la muestra y la calidad de los ítems. Resultados: Especialmente en condiciones con menores muestras, tests cortos o peores ítems, se han sobrestimado las fiabilidades al usar posteriores basadas en las estimaciones puntuales de los parámetros. Por otro lado, las posteriores calculadas con el procedimiento de IM proporcionaron estimaciones más precisas de la fiabilidad.

**Uso de modelos crossed random-effects para evaluar el efecto de Ninguna de las Anteriores y los rasgos de personalidad en ítems de opción múltiple.** Susana Sanz<sup>1</sup>, Carmen García<sup>1</sup>, Ricardo Olmos<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Escribir ítems con buenas propiedades psicométricas requiere esfuerzo y, en ocasiones, se utilizan opciones generales que son sencillas de crear, como “Ninguna de las anteriores” (NOTA, por sus siglas en inglés). Normalmente se desaconseja este tipo de alternativas porque se considera que empeoran las propiedades psicométricas de los ítems, pero la investigación previa es contradictoria. Igualmente, se conoce que la personalidad afecta al rendimiento académico, pero la investigación realizada hasta el momento no ha sido concluyente. Para testar estos efectos en contextos universitarios, se ha desarrollado un inventario de conceptos de estadística (SCI, por sus siglas en inglés) manipulando la aparición de

NOTA y su uso como opción correcta o distractor, y recogiendo información sobre los rasgos de personalidad de los evaluados. La muestra consistió en 449 estudiantes de Psicología, que respondieron una de las tres formas posibles del SCI. Posteriormente, se aplicó un modelo de crossed random-effects para modelar la probabilidad de acertar el ítem usando como covariables NOTA, los rasgos de personalidad y otras características del ítem o los evaluados. La aparición de NOTA afectó negativamente a la probabilidad de acertar los ítems, aunque estos efectos no fueron iguales para todos los evaluados. Algunos rasgos de personalidad y el conocimiento previo en asignaturas similares a la estadística también resultaron significativos. Al aplicar un LLTM, teniendo en cuenta también la dificultad del enunciado, se encontraron resultados similares en cuanto al uso de NOTA. Como conclusión, se encuentra que el uso de NOTA aumenta significativamente la dificultad de los ítems, aunque no todos los evaluados manifiestan el mismo comportamiento cuando tienen que lidiar con este tipo de opciones. Igualmente, la personalidad de los evaluados afecta a cómo responden a los ítems tipo test, por lo que es relevante intentar minimizar estas diferencias creando un entorno de evaluación justo.

**Extrayendo la fiabilidad y el ajuste de evaluaciones diagnósticas en contextos de baja muestra.** Vicente Pablo Nájera<sup>1</sup>, Francisco José Abad<sup>1</sup>, Chia-Yi Chiu<sup>2</sup>, Miguel Ángel Sorrel<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Rutgers, The State University of New Jersey

Los modelos de diagnóstico cognitivo (MDC) son modelos de respuesta al ítem en los que las variables latentes, en vez de ser continuas, son discretas. Esto posibilita realizar evaluaciones altamente multidimensionales que aportan información diagnóstica en ámbitos educativos o psicológicos. Tradicionalmente, los MDC han sido desarrollados y evaluados en contextos de alto tamaño muestral con el fin de facilitar la estimación de los múltiples parámetros que suelen estar involucrados en el modelo. Más recientemente, MDC no paramétricos han sido propuestos para reducir la dependencia al tamaño muestral. Estos modelos son capaces de aportar clasificaciones precisas incluso con tamaños muestrales de 30 participantes, lo que los convierte en opciones atractivas para realizar evaluaciones educativas a nivel del aula. Una limitación de los MDC no paramétricos es que, al prescindir de la estimación de los parámetros de los ítems, no es posible calcular su verosimilitud, necesaria para evaluar el ajuste, ni las probabilidades posteriores, necesarias para evaluar la fiabilidad. Por tanto, la aplicación de estos modelos para realizar evaluaciones diagnósticas en contextos de baja muestra no deja de ser, actualmente, un (prometedor) acto de fe. En este trabajo, presentamos el desarrollo de una extensión de los MDC no paramétricos que, manteniendo

su parsimonia, permite derivar su verosimilitud y las probabilidades posteriores. Mediante una serie de estudios de simulación, se muestra la utilidad de esta extensión para recuperar el ajuste y la fiabilidad de los modelos no paramétricos. Esta propuesta tiende un puente entre los MDC paramétricos y no paramétricos, combinando las ventajas de los primeros (índices de ajuste y fiabilidad) y los segundos (mayor precisión en la clasificación en contextos de baja muestra), permitiendo así la realización de evaluaciones diagnósticas rigurosas y válidas en contextos escolares o clínicos.

**R\_MORGANA: Un paquete en R para detectar residuales correlacionados en soluciones factoriales exploratorias.** Ana Hernández-Dorado<sup>1</sup>, Urbano Lorenzo-Seva<sup>1</sup>, PereJoan Ferrando<sup>1</sup> y Andreu Vigil-Colet<sup>1</sup>

<sup>1</sup>Universitat Rovira i Virgili

El desarrollo de métodos de detección de residuales correlacionados (o dobles) en el análisis de factorial (AF) tiene actualmente un claro interés. Sin querer entrar en el debate de la pertinencia de correlacionar residuales en los modelos de AF, omitir la presencia de estos dobles, cuando están presentes, tendrá dos posibles consecuencias (1) un mal ajuste del modelo a los datos y (2) sesgos en las estimaciones de los parámetros (principalmente los pesos). MORGANA, es un procedimiento desarrollado por Ferrando et al., 2022, basado en el concepto de “Cambio Esperado de Parámetro” (EPC, Saris et al, 1987) y que tiene como objeto detectar la presencia y magnitud de los residuales correlacionados. En concreto, el índice EREC cuantifica la cantidad de especificación incorrecta en la correlación residual misma que es inducida al forzarla a cero, mientras que ENIDE indica en qué medida la especificación incorrecta se propaga y produce sesgos en las estimaciones de los pesos. EREC y ENIDE se implementaron en el programa FACTOR (Ferrando & Lorenzo-Seva, 2017) y ofrecieron resultados más eficientes que los procedimientos tradicionales. La aportación que se hace aquí es, en primer lugar, implementar MORGANA en R para aquellos usuarios más acostumbrados a este lenguaje de programación. En segundo lugar, el funcionamiento de la implementación con respecto al de otros dos procedimientos clásicos se comparará utilizando un ejemplo empírico.

## Mesa 6

**Identificación de las prácticas parentales en el espacio euclídeo bidimensional.** Fernando García<sup>1</sup>, Marta Alcaide<sup>1</sup>, Óscar F. García<sup>1</sup>

<sup>1</sup>Universidad de Valencia

Las diferentes prácticas socializadoras de los padres se ordenan teóricamente en la literatura sobre dos dimensiones ortogonales: afecto y severidad. Precisamente una cuestión ampliamente discutida es la ordenación de las prácticas parentales sobre el espacio euclídeo, lo que implica realizar los cálculos para poder identificar en qué cuadrante se sitúa cada una de las prácticas de los padres. El objetivo de este trabajo es analizar mediante dos técnicas de análisis (correlaciones e intervalos de confianza) las covariaciones entre las prácticas parentales medidas por diferentes estudios representativos localizándolas sobre el modelo de dos dimensiones teóricamente ortogonales (i.e., independientes) y considerando, a su vez, las relaciones implícitas que mantienen entre sí las diferentes prácticas parentales y los ejes del llamado modelo teórico bidimensional de la socialización parental. También se analiza el tamaño del efecto. Los resultados indicaron que la representación geométrica de la colinearidad frente a ortogonalidad permite, a partir de los resultados estadísticos de cualquier estudio publicado, ubicar las diferentes prácticas parentales en el espacio euclídeo bidimensional. Por el contrario, una seria deficiencia en la literatura es que las prácticas parentales únicamente han medido una parte del modelo sin considerar teóricamente las otras. Parte del debate sobre los resultados de socialización de los padres y su contribución al desarrollo de los hijos, con resultados discrepantes y contradictorios, podría deberse a deficiencias del modelo teórico bidimensional que sostiene las investigaciones. Un sistema de contraste de las bases empíricas del modelo teórico bidimensional podría ayudar a superar estas discrepancias y contradicciones de los resultados empíricos.

**¿PTSD? PFA, ESEM y EFA aplicados a 127 muestras representativas de personas expuestas a un terremoto.**

Marcelo Leiva Bianchi<sup>1</sup>, Cristian Cáceres<sup>2</sup>, Carlos Serrano<sup>1</sup> Pablo Medina<sup>1</sup>, Francisco Ahumada<sup>1</sup>

<sup>1</sup>Universidad de Talca, <sup>2</sup>Universidad de las Américas

Objetivo: desde su primera definición, el número de dimensiones del trastorno de estrés postraumático (PTSD) ha aumentado (eg., de uno a siete). Esto podría deberse a dos razones metodológicas. Primero, los estudios no seleccionan muestras representativas de la población y no consideran la magnitud de la exposición al mismo evento. En segundo lugar, utilizan análisis de datos no robustos, lo que aumenta el número de dimensiones de forma artificial. El objetivo fue determinar cuántas dimensiones tiene el PTSD, cuando se controlan ambas fuentes de error. Método: de 25.400 participantes representativos de una misma población expuesta al mismo terremoto y evaluados con la Escala de Trauma de Davidson, se obtuvo 127 muestras de 200 participantes aproximadamente. Cada muestra fue

clasificada según la magnitud de Mercalli de la provincia de la cual fue extraída. Se aplicó análisis factoriales paralelos (PFA), modelos exploratorios de ecuaciones estructurales (ESEM) y análisis factorial exploratorio clásico (EFA) a cada muestra. Resultados: Hubo diferencias significativas en el ajuste y número de dimensiones entre la magnitud alta y baja de Mercalli. En condiciones de alta exposición se obtuvo una dimensión según PFA y tres según EFA; ESEM indicó mejores ajustes y menor cantidad de ítems eliminados en condiciones de alto Mercalli. El principal hallazgo fue que el PTSD tiene una dimensión única que incluye todos sus síntomas. También que el método utilizado cambia el número de dimensiones, lo mismo que el grado de exposición al evento. Conclusiones: Se argumenta la necesidad de utilizar métodos robustos de análisis para evaluar datos de muestras expuestas homogéneamente a un mismo evento potencialmente traumático.

**La opinión de los estudiantes de psicología sobre el uso de los test.** Marcelino Cuesta<sup>1</sup>, Mará Dolores Hidalgo Montesinos<sup>2</sup>, Eduardo García-Cueto<sup>1</sup>

<sup>1</sup>Universidad de Oviedo, <sup>2</sup>Universidad de Murcia

No cabe duda de la importancia que pueden llegar a tener los test en la vida de las personas. A día de hoy, los test están cada vez más presentes en múltiples circunstancias de las vidas de las personas. De su uso puede depender desde la evaluación académica, hasta diagnósticos y procesos de selección. Es evidente la importancia de hacer un buen uso de estos instrumentos, conocer sus propiedades y limitaciones sobre todo por aquellos profesionales que han de utilizarlos. Así pues, resulta de sumo interés conocer la opinión de los diferentes colectivos del mundo de la psicología respecto a estas herramientas. En los años 2000, 2010 y 2020 desde la Comisión de Test del Consejo General de la Psicología de España se promovieron sendos estudios para conocer la opinión de los profesionales españoles, en 2021 se realizó una réplica del estudio entre profesores universitarios. Ahora se trata de cerrar el círculo: En este trabajo, se recoge la opinión sobre los test de los estudiantes de psicología de diversas facultades españolas, tanto de grado como de master. Se establecen comparaciones en función del curso, asignaturas cursadas, etc. Dado que los test parece que también se han adueñado del ámbito de la evaluación académica, sobre todo en el ámbito universitario, otro de los objetivos de esta investigación será conocer la frecuencia de uso de los exámenes “tipo test” en los estudios de psicología y la percepción que de algunas de sus características tienen nuestros estudiantes.

**Evaluación del estrés en pacientes con cáncer de mama: propuesta de puntos de corte del PSS-10.** María Victoria

Cerezo<sup>1,2</sup>, Lorena M. Soria-Reyes<sup>1</sup>, Lucía Palacios<sup>2</sup>, Rocío Gómez-Molinero<sup>2,3,4</sup>

<sup>1</sup>Universidad de Málaga, <sup>2</sup>Universidad de Educación a Distancia, <sup>3</sup>Universidad de Cádiz, <sup>4</sup>Instituto Universitario de Investigación para el Desarrollo Social Sostenible (INDESS)

Un diagnóstico de cáncer de mama puede ser una experiencia traumática para algunas pacientes que puede generar altos niveles de estrés (Dooley et al., 2017), reportando aproximadamente el 40% de las pacientes, síntomas de angustia emocional (Cerezo et al., 2020). Por tanto, evaluar los niveles de estrés es vital para desarrollar intervenciones que favorezcan el ajuste psicológico a la enfermedad. El objetivo de este estudio fue desarrollar puntos de corte en la Escala Breve de Estrés Percibido (PSS-10; Cohen et al., 1983; Cohen & Williamson, 1988) para probar la capacidad discriminativa de esta prueba diagnóstica. Para ello, se evaluó el área bajo la curva ROC (AUC) con el objetivo de estimar la prevalencia de la enfermedad en estudio, reflejando qué tan buena es la escala para discriminar pacientes con y sin estrés en todo el rango de posibles puntos de corte. Participaron 215 mujeres españolas diagnosticadas de cáncer de mama que cumplieron las escalas PSS-10 y DASS-21. Se realizó un análisis ROC utilizando el índice de Youden para encontrar las mejores puntuaciones de corte para determinar la sensibilidad y mayor especificidad del instrumento. Las puntuaciones de las participantes se clasificaron siguiendo los puntos de corte de la subescala de estrés del DASS-21 (Lovibond & Lovibond, 1995), que clasifica las puntuaciones por encima de 15 como estrés leve. Respecto al nivel de estrés leve, el análisis de la curva ROC del PSS-10 ofrece un punto de corte de 22.5 puntos, con un AUC del 89.2% (84.3%-94.1%), mostrando 33 pacientes por encima de esta puntuación. En conclusión, el punto de corte generado a partir del PSS-10 para estas pacientes puede ayudar a los profesionales sanitarios a evaluar los niveles de estrés, siendo útil para determinar aspectos clave de sus intervenciones que mejoren el ajuste psicológico de las mujeres con cáncer de mama.

**Avances y limitaciones en la medición de personalidad a través de modelos de espacio vectorial.** José Á. Martínez-Huertas<sup>1</sup>, Ricardo Olmos<sup>2</sup>, José David Moreno<sup>2</sup>, Alejandro Martínez-Mingo<sup>2</sup>, Guillermo Jorge-Botana<sup>3</sup>

<sup>1</sup>Universidad Nacional de Educación a Distancia, <sup>2</sup>Universidad Autónoma de Madrid, <sup>3</sup>Universidad Complutense de Madrid

Distintos indicadores psicolingüísticos han sido relacionados con constructos psicológicos como los rasgos de personalidad a través de modelos computacionales del lenguaje. En este estudio, exploramos el uso de los modelos de espacio vectorial para la evaluación de respuestas construidas y la relación de sus indicadores computacionales con los rasgos de personalidad del Big Five (apertura mental,

responsabilidad, extraversión, amabilidad, y neuroticismo). Una muestra de 643 estudiantes universitarios respondió a distintas pruebas de respuesta construida y a tests de personalidad estandarizados que evalúan los cinco grandes. En este estudio, proponemos la generación de subespacios específicos para cada respuesta construida que sean capaces de capturar diferencias entre los distintos rasgos de personalidad. Esta metodología fue puesta a prueba con respuestas construidas que consisten tanto en una única prueba (e.g., una autodescripción) como en múltiples pruebas (e.g., las respuestas escritas a entrevistas estructuradas, o a tests proyectivos). Parece que la utilización de múltiples pruebas de respuesta construida puede asemejarse más a la evaluación de un test clásico de personalidad que está compuesto por múltiples indicadores o ítems gracias a la consideración del modelo de medida. Aunque los resultados muestran la utilidad de estos subespacios específicos para capturar parte de la varianza asociada a los rasgos de personalidad del Big Five, se encontraron tamaños del efecto bajos y relativamente baja generalizabilidad en muestras de validación. En nuestra opinión, estos resultados se asemejan a los resultados encontrados en otros indicadores indirectos de personalidad. Este trabajo ha sido financiado por las ayudas AEMCCO-Julio Olea a jóvenes investigadores.

## Mesa 7

### **Potencia predictiva del residual ajustado en las matrices de transición.** Manuel Morales-Ortiz<sup>1</sup>

<sup>1</sup>Universidad de Sevilla

El análisis de secuencias de conducta es una técnica habitual dentro de la psicología. Para realizar este análisis se suele realizar alguna prueba estadística sobre las matrices de transición obtenidas. Para extraer la información contenida en dichas matrices de transición suele utilizarse algún estadístico. Una de las posibilidades existentes es el estadístico residual ajustado. El objetivo de este trabajo fue realizar un estudio de simulación para determinar las posibilidades del residual ajustado para extraer la información de las matrices de transición. METODO. Se diseñó un experimento de Monte Carlo donde se calculó el residual ajustado para cada una de las casillas de las matrices de transición. Se estudió como afectó al cálculo del residual ajustado el tamaño de la secuencia (20, 50 y 100), el número de categorías (4 y 8) y el tipo de dependencia existente en la matriz de transición (sin dependencia, dependencia en una categoría y dependencia en 2 categorías). Se obtuvieron 5000 matrices de transición para cada una de las condiciones. Las matrices fueron generadas con el programa R. RESULTADOS. Los resultados mostraron que el porcentaje de residuales significativos en las matrices independientes estuvo alrededor del 5%. El porcentaje de

patrones de dependencia detectados aumentó a medida que aumentaba el tamaño de la muestra. Hubo patrones de dependencia no establecidos en las matrices originales a medida que aumentó el tamaño de la secuencia. Se discuten los resultados.

### **¿Mi modelo estadístico es generalizable? Estimación directa del error de generalización mediante técnicas de validación cruzada.** Diego Iglesias<sup>1</sup>, Miguel Sorrel<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

En el proceso de evaluación de la bondad de ajuste de un modelo estadístico, un índice que no penaliza la complejidad del modelo no es robusto frente al fenómeno del sobreajuste, por lo que ofrece estimaciones sesgadas (optimistas) del error de generalización (Err) del modelo. Dada la necesidad de desarrollar modelos predictivos complejos (sensibles al sobreajuste), la comunidad del machine learning ha prestado mucha atención a la validación cruzada (CV), un método que permite realizar una estimación directa de Err. A pesar de su aparente sencillez, el comportamiento de CV es complejo, pues depende del contexto en el que se aplique. El objetivo de este estudio consiste en estudiar el funcionamiento de CV como estimador de Err en modelos estadísticos típicamente utilizados en Psicología como los modelos lineales generalizados. Para ello, se ha llevado a cabo un estudio de simulación Monte Carlo en el que se ha generado una población de observaciones según un modelo de regresión. Variando la ratio señal/ruido entendida como porcentaje de varianza explicada y el tamaño de la muestra extraída de la población sobre la que se estiman diversos modelos que varían en complejidad, se ha analizado el sesgo en la estimación de Err de índices calculados mediante distintas técnicas de validación cruzada (MSE hold out, leave-one-out, 10-folds, repeated 10-folds) vs aquellos que no penalizan la complejidad del modelo (MSE). Según los resultados obtenidos, MSE leave-one-out, 10-folds y repeated 10-folds son los estadísticos que realizan estimaciones insesgadas de Err. Dado que todo modelo ajustado en una muestra se adapta a las condiciones de la misma, se deben utilizar técnicas estadísticas robustas al sobreajuste y que permitan estimar de manera insesgada la generalizabilidad de los resultados obtenidos en futuras observaciones de la población objetivo.

### **Evaluación de la calidad metodológica en tiempos de COVID-19: desarrollo y aplicación de una “checklist” a estudios sobre intervenciones psicológicas online.**

Cristina Rodríguez-Prada<sup>1</sup>, Luíís Morís<sup>1</sup>, Miguel A. Vadillo<sup>2</sup>, Salvador Soto-Faraco<sup>1</sup>, Miguel Burgeleta<sup>3</sup>

<sup>1</sup>Center for Brain and Cognition, Universitat Pompeu Fabra,

<sup>2</sup>Universidad Autónoma de Madrid; <sup>3</sup>Universitat de Barcelona



El impacto negativo de la pandemia COVID-19 sobre el bienestar psicológico ha propiciado la proliferación de intervenciones psicológicas online, así como de investigaciones orientadas a la probar su efectividad. Sin embargo, se desconoce si la calidad científica de esta evidencia podría haberse visto comprometida durante la pandemia. La meta-investigación es la disciplina encargada de analizar las prácticas científicas con el fin de introducir mejoras progresivas en el funcionamiento las instituciones científicas y en la calidad de los resultados de investigación respecto a cuestiones como los métodos, la comunicación de resultados, la reproducibilidad o la evaluación de prácticas del ámbito investigador. Entre algunas herramientas que utilizan se encuentran las 'checklists'. El objetivo de este trabajo es diagnosticar la calidad de la evidencia que apoya las intervenciones psicológicas online publicadas tras la irrupción del COVID-19 y compararla con un conjunto de artículos control publicados antes de la pandemia. Para ello, hemos desarrollado una checklist para evaluar la calidad y comparar los resultados en ambas muestras. Esta checklist recoge diferentes aspectos de los estudios: diseño, enmascaramiento, análisis de datos, replicabilidad y comunicación de los resultados. Nuestros resultados señalan diferencias entre ambos grupos de artículos. Por ejemplo, las investigaciones realizadas durante la pandemia tienden a (1) tener un menor tiempo de aceptación en las revistas, (2) no utilizar pre-registros, (3) prescindir de la aleatorización de los participantes; (4) publicar en mayor medida en acceso abierto. Sin embargo, la tendencia a no compartir los datos recogidos es común para ambos tipos de artículos. Se discuten las posibles implicaciones prácticas de estos hallazgos respecto a la evidencia generada en Psicología Clínica.

**Análisis de los factores asociados a la evaluación de la toma de decisiones ante dilemas morales reales y de laboratorio.** Belén Carrascal-Caputto<sup>1</sup>, José Luis Padilla<sup>1,2</sup>, Isabel Benítez<sup>1,2</sup>

<sup>1</sup>Universidad de Granada, <sup>2</sup>Centro de Investigación Mente Cerebro y Comportamiento (CIMCYC)

**Introducción:** A pesar de que los dilemas sacrificiales han sido la herramienta principal en el estudio de la toma de decisiones ante dilemas morales, la evidencia empírica sugiere que las decisiones tomadas ante dilemas reales predicen mejor el comportamiento moral que las situaciones artificiales. Entre las limitaciones de los dilemas de laboratorio se destaca su falta de verosimilitud con respecto a los dilemas que ocurren en la vida cotidiana. El objetivo de este estudio es analizar los factores que provocan diferencias entre la toma de decisiones ante los estímulos experimentales y ante los dilemas morales reales. **Método:** 61 participantes completaron un cuestionario web donde respondieron a un dilema moral sacrificial y describieron un dilema moral experimentado. Además, los participantes respondieron

varias preguntas abiertas en las que explicaron los motivos de sus decisiones en cada uno de los dilemas. **Resultados:** Ante el dilema sacrificial las personas fueron más consecuencialistas, eligiendo la alternativa que les permitía salvar a un mayor número de personas. Fueron más deontológicas ante los dilemas de su vida real, tomando decisiones basadas en las normas sociales o en las normas asociadas al puesto de trabajo. Resaltaron las características de los implicados y el grado de cercanía en la toma de decisiones en ambos escenarios. Sin embargo, las decisiones tomadas ante los dilemas cotidianos reflejaron una mayor presencia de las responsabilidades asociadas al puesto de trabajo. **Conclusiones:** Este estudio evidencia la idoneidad de las preguntas abiertas para observar la falta de convergencia entre los procesos de respuesta realizados ante los dilemas de laboratorio y ante los dilemas de la vida real, y muestra la necesidad de generar dilemas sacrificiales más realistas que mejoren la evaluación.

## Mesa 8

**Desarrollo de la escala de Actitudes Policiales hacia la Intervención en Violencia de género y obtención de evidencias de validez: Aproximación desde los métodos mixtos.** Celia Serrano-Montilla<sup>1</sup>, Luis-Manuel Lozano<sup>2</sup>, José-Luis Padilla<sup>2</sup>

<sup>1</sup>Universidad de Córdoba, <sup>2</sup>Universidad de Granada

El objetivo del presente trabajo es presentar la aplicación de un diseño secuencial exploratorio de métodos mixtos para el desarrollo de escalas y análisis de validación propuesto por Zhou (2019). Concretamente, al constructo de actitudes policiales hacia la intervención en violencia de género dado que no existen instrumentos disponibles adaptados al contexto lingüístico, cultural y legislativo español para su medida. Tres fases serán abordadas. Una primera fase cualitativa en la que se define y delimita el constructo (definición semántica y sintáctica) a partir de la revisión sistemática de la literatura, grupos focales con agentes y un juicio de expertos. Una fase de desarrollo del instrumento que incluye la generación de los ítems y su revisión. Finalmente, una última fase cuantitativa en la que se obtienen evidencias del funcionamiento psicométrico del instrumento y se obtienen evidencias de validez basadas en el contenido del test, en la estructura interna y en la relación con otras variables. A lo largo de las fases se emplearon procedimientos para obtener datos cuantitativos y cualitativos que ayudaron a establecer una definición semántica de cuatro componentes (disposición a la mínima implicación policial, tolerancia hacia la violencia de género, disposición comprometida con la intervención en violencia de género, y comprensión de la violencia de género), así como sus comportamiento, indicadores y definición sintáctica (por

ej., experiencia profesional del agente, daños físicos de la víctima). Tras ello, se obtuvieron evidencias de validez basadas en el contenido del test, obteniéndose para un total de 32 ítems, así como evidencias basadas en la estructura interna y en relación con otras variables. Los pasos recomendados no solo enfatizan la utilidad del diseño, sino que también se enfocan en técnicas específicas de unión de datos cualitativos y cuantitativos, y validación para generar instrumento de calidad.

**Explorando el papel mediador de las regulaciones motivacionales en la relación entre la imagen corporal y nivel de actividad física: una aproximación mediante un modelo de ecuaciones estructurales.** Sergio Navas-León<sup>1</sup>, Milagrosa Sánchez-Martín<sup>1</sup>, Ana Tajadura-Jiménez<sup>2</sup>, Mercedes Borda-Más<sup>3</sup>, María Luisa Avaregues-Navarro<sup>3</sup>, Luis Morales Márquez<sup>1</sup>

<sup>1</sup>Universidad Loyola Andalucía, <sup>2</sup>Universidad Carlos III de Madrid, <sup>3</sup>Universidad de Sevilla

Objetivo: Los efectos de la imagen corporal (IC) sobre el nivel de actividad física (AF) son escasamente estudiados y los mecanismos que subyacen esta relación rara vez se exploran. Por tanto, este estudio tiene como objetivo explorar el papel mediador de las regulaciones motivacionales en la relación entre la IC y la AF. Diseño/Método: Se realizó un estudio transversal. 1012 participantes fueron seleccionados mediante muestreo aleatorio estratificado. La IC se evaluó con el Multidimensional Body Self-Relations Questionnaire, la regulación motivacional con el Behavioural Regulation Exercise Questionnaire y el nivel de AF con el International Physical Activity Questionnaire. Se empleó un modelo de ecuaciones estructurales. Resultados: El modelo teórico mostró buenos índices de ajuste [CFI= 0.925; TLI= 0.917; RMSEA= 0.050 (IC 90 % = 0.048-0.053); SRMR= 0.067]. La evaluación de la apariencia mostró un efecto directo negativo sobre la AF ( $\beta = -0.85$ ,  $p < 0.01$ ) y un efecto indirecto positivo a través de las motivaciones intrínseca ( $\beta = 0.29$ ,  $p < 0.01$ ), integrada ( $\beta = 0.22$ ,  $p < 0.01$ ) e introyectada ( $\beta = 0.20$ ,  $p < 0.01$ ). La evaluación de la forma física mostró un efecto directo positivo sobre la AF ( $\beta = 2.90$ ,  $p < 0.01$ ) y un efecto indirecto negativo a través de las motivaciones intrínseca ( $\beta = -0.42$ ,  $p < 0.01$ ), integrada ( $\beta = -0.56$ ,  $p < 0.01$ ) e introyectada ( $\beta = -0.54$ ,  $p < 0.01$ ). Conclusiones: Estos resultados respaldan la integración de las regulaciones motivacionales como procesos subyacentes entre la IC y AF. Los hallazgos pueden ayudar a sentar las bases para futuras políticas innovadoras que busquen promover una AF saludable.

**Nueva aproximación a la evaluación de la personalidad en animales: primer estudio piloto con osos marinos.** Ignacio Miguel Pardillo<sup>1</sup>, Ricardo Olmos<sup>1</sup>, Susana Sánchez<sup>1</sup>, Ana Fidalgo<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

El estudio del bienestar de los animales criados en cautividad pasa directamente por la evaluación de su personalidad, en tanto que su bienestar es consecuencia directa de la interacción entre el conjunto de conductas estables de los mismos y las circunstancias donde se encuentran. Pese a la importancia de esta rama del conocimiento, existen serias dudas sobre qué procedimientos de análisis emplear para validar los modelos de personalidad en animales. El principal problema es que los datos de este tipo de estudios parten de pocos individuos, pero con múltiples medidas repetidas recogidas a partir del registro de conductas observadas. La aplicación de Análisis Factorial Exploratorio acompañado del pack "Little Jiffy" con este tipo de datos lleva a la obtención de pesos sobreestimados, un número de dimensiones cuestionable y conclusiones, por lo tanto, difícilmente replicables. Existen alternativas de análisis interesantes como el análisis clúster jerárquico que no impone una distribución de probabilidad a los datos, el análisis factorial multinivel o la aproximación de modelos regresión multinivel que descompone la variabilidad de las conductas observadas en diferentes fuentes (variabilidad intrasujeto, entre sujetos o entre distintos periodos del día). Es por esto que la investigación que llevamos a cabo tiene como objetivos: A) el estudio de la personalidad en osos marinos; B) la consolidación de una nueva metodología fiable y válida, que agrupe los métodos de codificación de comportamientos y evaluación de tendencias y que emplee herramientas de análisis de datos capaces de evaluar la variabilidad inter e intraindividual con un tamaño muestral tan reducido; C) evaluar el efecto de ciertas variables sobre el bienestar de los osos marinos criados en cautividad, como la incorporación de exhibiciones a su rutina, el tamaño de las instalaciones o el momento del día.

**Aplicación del análisis de T-patterns en la construcción de un instrumento de observación para la formación en educación temprana.** Haizea Belza<sup>1</sup>, Mariona Portell<sup>2</sup>, Elena Herrán<sup>1</sup>, M. Teresa Anguera<sup>3</sup>

<sup>1</sup>Universidad del País Vasco/Euskal Herriko Unibertsitatea, <sup>2</sup>Universitat Autònoma de Barcelona, <sup>3</sup>Universidad de Barcelona

Existe un importante vacío en cuanto a la investigación de estrategias para la formación y la evaluación de las profesionales de la educación temprana (0-3 años), en el desempeño de actividades cotidianas que se adecuen a las características específicas de esta etapa. El objetivo de esta comunicación es presentar un instrumento de observación para dicha formación y evaluación basada en la educación Pikler. Se propone la conversión en sistema de categorías del formato de campo Dar de desayunar en la Escuela Infantil Emmi Pikler, elaborado para captar y explicar la conducta de las educadoras de esta institución.

Una vez realizada la revisión de los catálogos de conducta de cada sub-dimensión del formato para poder establecer un sistema exhaustivo y mutuamente excluyente, se modifican los registros obtenidos de la observación sistemática de dos educadoras expertas en la educación Pikler. Ambas bases de datos, la original y la modificada, son sometidas a un análisis de T-patterns (TPA) con el software THEME. Con TPA se obtiene un acercamiento multivariante para la detección y descripción de secuencias recurrentes de eventos conductuales. Se comparan los patrones temporales obtenidos con las categorías previas y con la agrupación actual, que pasa de un nivel más molecular a uno más molar; asimismo, se comparan los patrones entre educadoras. Los resultados evidencian la utilidad del instrumento de observación para la detección de patrones y esclarecen aspectos relevantes de las coreografías seguidas por estas educadoras, proporcionando con ello material sustancial para generar un plan formativo, que incluya estos patrones como técnica de aprendizaje por modelado.

## Mesa 9

**Evaluación del uso problemático de internet: Adaptación al euskera del Generalized Problematic Internet Use Scale – 2 (GPIUS-2).** Olatz Goñi-Balentiaga<sup>1</sup>, Jone Aliri<sup>1</sup>, Nekane Balluerka<sup>1</sup>, Arantxa Gorostiaga<sup>1</sup>

<sup>1</sup>Universidad del País Vasco UPV/EHU

Diversas definiciones del uso problemático de internet incluyen un uso inadecuado de internet que causa angustia, consume mucho tiempo, impide el funcionamiento en áreas importantes de la vida y se mantiene a pesar de estas consecuencias negativas. En Europa se estima que entre el 1% y el 18% de las personas que utilizan internet realizan un uso problemático de esta herramienta. Dado que este uso puede generar problemas psicológicos y sociales, su evaluación es de gran importancia. El Generalized Problematic Internet Use Scale – 2 (GPIUS-2; Caplan, 2010) es uno de los escasos instrumentos que mide el uso problemático de internet en base a la teoría cognitivo conductual y uno de los que presenta mejores propiedades psicométricas. El objetivo del presente trabajo consistió en adaptar al euskera el GPIUS-2 y en examinar sus propiedades psicométricas. Se utilizaron dos muestras, una de adultos (Muestra 1) compuesta por 252 participantes (56,7% mujeres) de 18 a 62 años (M=28,7 y DT=10,21), y otra de adolescentes (Muestra 2) compuesta por 1.265 participantes (53,7% mujeres) de 11 a 17 años (M=14,44 y DT=1,69). En la Muestra 1, los datos se recogieron online con una batería que incluía el GPIUS-2 y las versiones españolas del IAT y del DASS21. En la Muestra 2, la recogida se realizó de manera presencial con una batería que incluía el GPIUS-2 y el BASC. Los resultados del AFC respaldan una estructura factorial de 4 componentes de

primer orden, uno de segundo orden y una dimensión total, con índices de fiabilidad adecuados y buena validez convergente en relación al IAT. Asimismo, se obtuvieron evidencias de validez de relación con variables como la Depresión, la Ansiedad y el Estrés. Se concluye que el GPIUS-2 es un instrumento útil para evaluar el uso problemático de internet en población vascoparlante.

**Propiedades psicométricas del Cuestionario sobre expectativas de los efectos del alcohol (EQ-SF) en una muestra comunitaria de jóvenes adultos.** Bella González-Ponce<sup>1</sup>, Nehemías Romero-Pérez<sup>1</sup>, Alberto Parrado-González<sup>1</sup>, Pepe Carmona-Márquez<sup>1</sup>, Fermín Fernández-Calderón<sup>1</sup>

<sup>1</sup>Universidad de Huelva

Las expectativas sobre los efectos del alcohol han demostrado predecir el consumo actual y futuro de alcohol. Entre los instrumentos más utilizados para evaluar esta variable se encuentra el Expectancy Questionnaire, que incluyen expectativas positivas y negativas. La versión corta (EQ-SF) ha mostrado adecuadas propiedades psicométricas en una muestra de adolescentes y una de adultos. El objetivo del presente estudio es analizar las propiedades psicométricas del EQ-SF en una muestra comunitaria de jóvenes adultos. La muestra estuvo compuesta por 360 jóvenes adultos (edad: M = 21.15; DT = 2.23) que cumplimentaron la escala compuesta por 24 ítems organizados en ocho dimensiones. Los coeficientes de fiabilidad alfa de Cronbach de las subescalas fueron adecuados, oscilando entre .898 (subescala social positiva) y .691 (subescala social negativa). Aplicando la prueba de la diferencia entre chi-cuadrados de Satorra-Bentler se encontró que el modelo de ocho factores correlacionados ajustaba significativamente mejor que el modelo de dos factores de segundo orden,  $SB\chi^2(19) = 71.83, p < .001$ . Además, ese modelo de ocho factores correlacionados mostraba también un ajuste aceptable a los datos,  $SB\chi^2(224) = 405.22, p < .001, CFI = .943, TLI = .930, RMSEA = .047, SRMR = .050$ . Por tanto, se encontraron evidencias de validez de la estructura interna del EQ-SF. Los resultados del presente estudio muestran que el EQ-SF posee adecuadas propiedades psicométricas para medir las expectativas de los efectos del alcohol en jóvenes adultos. Este trabajo está financiado por I+D+i Biomédica y Ciencias de la Salud en Andalucía (PI-0503-2018, PI: Fermín Fernández Calderón), y por el Ministerio de Universidades del Gobierno de España (FPU19/01413).

**Satisfacción con la Vida en Estudiantes Universitarios Españoles tras el COVID-19: Invarianza factorial de la escala SWLS.** Rocío Gómez-Molinero<sup>1,2,3</sup>, Manuel García-Sedeño<sup>1</sup>, Teresa Lozano-Antolínez<sup>1</sup>, Rocío Guil<sup>1,3,4</sup>, María Victoria Cerezo<sup>2,5</sup>

<sup>1</sup>Universidad de Cádiz, <sup>2</sup>Universidad de Educación a Distancia (UNED), <sup>3</sup>Instituto Universitario de Investigación para el desarrollo Social Sostenible (INDESS), <sup>4</sup>Instituto de Investigación Biomédica de Cádiz (INIBICA), <sup>5</sup>Universidad de Málaga

La pandemia provocada por el COVID-19 ha traído cambios en la vida académica de muchos universitarios principalmente por el impacto en la participación y actividad social en un momento vital que es clave para el desarrollo de la persona y pudiendo afectar a su bienestar. La satisfacción con la vida se considera la parte cognitiva del bienestar y se ha definido como el juicio de valor que se hace de la diferencia entre las expectativas y logros actuales de cada persona a nivel global, es decir, la evaluación que cada persona hace de su vida valorándola como un todo (Diener et al., 1999; Diener et al., 2013; Pavot y Diener, 2008). El objetivo de este estudio fue determinar los niveles de satisfacción con la vida de los estudiantes universitarios españoles tras la pandemia, comprobando las propiedades psicométricas y la invariancia factorial de la escala SWLS. La muestra estuvo compuesta por 379 estudiantes universitarios de toda España que cumplieron los cuestionarios SWLS, resiliencia e inteligencia emocional percibida. Los resultados sugirieron una adecuada estructura unidimensional de la SWLS, mostrando, además, una invariancia factorial entre sexos. Los índices de ajuste de invariancia fueron adecuados, mostrando un RMSEA de 0.025, con intervalos entre [0.00-0.063], SRMR de 0.026, y CFI y TLI, ambos de 0.99. Estos resultados indicaron que existe invariancia factorial entre estudiantes universitarios hombres y mujeres. Además, el análisis mostró un coeficiente omega de McDonald's de 0.87. La validez basada en las relaciones con otras variables, en este caso inteligencia emocional percibida y resiliencia, mostró correlaciones satisfactorias de 0.50 con reparación emocional y 0.63 con resiliencia. En conclusión, la SWLS parece ser una herramienta de evaluación adecuada para comprender mejor el juicio cognitivo sobre el bienestar subjetivo de los estudiantes españoles después de la pandemia.

**Validación de una serie de medidas implícitas y de autoinforme sobre interés por las crías para el estudio evolutivo de la prosocialidad.** Ignacio Miguel Pardillo<sup>1</sup>, José Manuel Caperos<sup>2</sup>, Denisa Cristina Lupu<sup>1</sup>, Claudia Rodríguez-Ruiz<sup>1</sup>, Beatriz Gil-Martí<sup>1,3</sup>, Ana Fidalgo<sup>1</sup>Nereida Bueno-Guerra<sup>2</sup>, Susana Sánchez<sup>1</sup>; Enrique Turiégano<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad Pontificia de Comillas, <sup>3</sup>Instituto Cajal - CSIC

El estudio de los mecanismos que sustentan la crianza cooperativa es clave para la comprensión evolutiva de la conducta cooperativa y prosocial de la especie humana. Dentro de estos mecanismos, resultan de vital interés aquellos que facilitan la orientación de la atención de los

adultos hacia los estímulos propios de las crías. Las pruebas de autoinforme suelen ser las más empleadas en las investigaciones sobre diferencias individuales en la motivación e interés hacia los recién nacidos; sin embargo, ninguno de los cuestionarios desarrollados hasta el momento ha sido traducido al español. Además, las respuestas a este tipo de pruebas pueden verse afectadas por distintos sesgos, pudiendo ser más adecuado el empleo de medidas implícitas. El objetivo de esta investigación es la validación al castellano la Escala de Actitud hacia los Bebés (ABS), así como la evaluación de la relación entre esta medida y dos tareas implícitas, una atencional (Dot-Probe) y otra actitudinal (ST-IAT). Un total de 792 participantes (553 mujeres; 20,00 ± 3,27 años) respondieron a una versión traducida del ABS junto a algunas variables criterio, y de estos, 269 realizaron las tareas implícitas. Se realizó un AFE en 512 respuestas, mostrando un ajuste adecuado del ABS a un modelo de 4 factores, confirmado mediante un AFC en 253 respuestas. Los 4 factores presentaron una consistencia interna adecuada, así como las medidas implícitas mostraron también un adecuado funcionamiento. Tanto la medida de autoinforme como las tareas implícitas mostraron una relación con el criterio sexo. Las medidas relativas a las escalas del autoinforme presentan una relación moderada con las dos medidas implícitas, que sin embargo no presentan relaciones entre ellas. Concluimos la utilidad del autoinforme ABS traducido al castellano, así como de las pruebas implícitas Dot-Probe y ST-IAT, para la medición de diferencias individuales en el interés adulto hacia los recién nacidos.

**Mitos sobre la ciberviolencia sexual contra las mujeres: un análisis temático y de contenido de las reacciones sociales a denuncias en Twitter.** Rocío Vizcaíno-Cuenca<sup>1</sup>, Mónica Romero-Sánchez<sup>1</sup>, Hugo Carretero-Dios<sup>1</sup>

<sup>1</sup>Universidad de Granada

Introducción: La violencia sexual que las mujeres sufren en los espacios online representa un fenómeno poco estudiado. A pesar de los indudables beneficios de las redes sociales, las características de los entornos online (i.e., anonimato, amplia audiencia, rapidez, etc.) también han ocasionado consecuencias negativas. Entre ellas, el presente trabajo analiza la "ciberviolencia sexual" contra las mujeres, así como la cultura sexista online que justifica dicha violencia. En concreto, nos interesa conocer en profundidad la percepción social sobre este tipo de violencia para en última instancia aislar las áreas de contenido que servirían para definir los mitos sobre la ciberviolencia sexual contra las mujeres. Para ello, y siguiendo investigaciones previas que han destacado la relevancia de la información intercambiada en redes sociales (i.e., Twitter), en este estudio se realiza un análisis cualitativo de reacciones sociales a denuncias en Twitter. Método: Primero, se han seleccionado 18 denuncias sobre ciberviolencia sexual contra mujeres

publicadas en Twitter empleando para ello el paquete de extracción de datos rtweet implementado en el software estadístico R. Segundo, a partir de las denuncias, se han extraído un total de 4,048 reacciones utilizando el software Octoparse. Finalmente, a partir de estas reacciones, se ha hecho un análisis temático y de contenido siguiendo la propuesta de Clarke and Braun (2018). Resultados: Los resultados del análisis muestran que existen actitudes que justifican la ciberviolencia sexual contra las mujeres a través de: 1) minimizar/negar el problema, 2) culpabilizar a la víctima, 3) exonerar al perpetrador y 4) factores socioculturales. Conclusión: Estos resultados posibilitan una comprensión más profunda de los mitos sobre la ciberviolencia sexual contra las mujeres, y muestran que la cultura de la violencia sexual contra las mujeres persiste y alcanza también a los espacios online.

## Mesa 10

### **Análisis exploratorio de las secuencias de conducta mediante gráficas polares.** Manuel Morales-Ortiz<sup>1</sup>

<sup>1</sup>Universidad de Sevilla

El análisis de secuencias de conducta es una técnica habitual dentro de la psicología. Para realizar este análisis se suele realizar alguna prueba estadística sobre las matrices de transición obtenidas, utilizándose algún estadístico. Por lo general, se ha propuesto el residual ajustado para representar los patrones de dependencia mediante gráficos polares. Sin embargo, este estadístico presenta algunos problemas para describir la información contenida en las matrices de transición, siendo necesario la búsqueda de estadísticos alternativos. El objetivo de este trabajo fue realizar un estudio de simulación para determinar las posibilidades de distintas medidas para obtener la información contenida en las matrices de transición. METODO. Se diseñó un experimento de Monte Carlo donde se calcularon distintas medidas para cada una de las casillas de las matrices de transición. Se estudió como afectó al cálculo de una serie de estadísticos el tamaño de la secuencia (20, 50 y 100), el número de categorías (4 y 8) y el tipo de dependencia existente en la matriz de transición (sin dependencia, dependencia en una categoría y dependencia en 2 categorías). Se obtuvieron 5000 matrices de transición para cada una de las condiciones. Las matrices fueron generadas con el programa R. RESULTADOS. Los resultados indicaron que los estadísticos alternativos al residual ajustado presentaron una buena sensibilidad para detectar los patrones de dependencia incluso cuando los tamaños muestrales fueron pequeños. El resultado fue independiente del número de categorías.

### **Observaciones sobre el Análisis Multinivel de Potenciales Evocados con Muestras Pequeñas.** Juan C. Oliver-Rodríguez<sup>1</sup>, Mirjam Moerbeek<sup>2</sup>

<sup>1</sup>Universitat Jaume I, <sup>2</sup>Utrecht University

Los estudios de neuroimagen se realizan con cierta frecuencia con muestras pequeñas y razones entre la señal y el ruido o actividad neural subyacente no muy elevadas. En estos casos, el análisis multinivel produce a veces estimaciones negativas de varianzas que por definición son positivas y que por tanto representan valores no permisibles. Estos en ocasiones no son incluidos en los informes. La probabilidad de que esto ocurra cuando el valor real de la varianza es positiva aumenta cuando disminuyen el número de participantes, el número de observaciones por participante, y la razón entre su valor real y el residual. Una segunda consecuencia del uso de muestras pequeñas es la aparición de sesgos en las estimaciones de los componentes de varianza en los niveles superiores, aunque el número aconsejado de participantes varía según estudios. Dada la diversidad de las características de los campos de aplicación de los procedimientos multinivel así como el coste del número de participantes en los experimentos de neuroimagen, es aconsejable una planificación informada previa del diseño de investigación en cada caso. El objetivo del presente trabajo es analizar e interpretar los resultados con muestras pequeñas de una simulación cuyos parámetros fueron extraídos de un estudio de percepción facial con potenciales evocados, así como la sugerencia de algunas recomendaciones.

### **Cambio clínicamente significativo de síntomas de trastornos emocionales en una muestra de pacientes de salud mental: Resultados preliminares.** Ana de la Rosa<sup>1</sup>, Juan José Mancheño<sup>2</sup>, Matilde Alonso<sup>2</sup>, Maria Dolores Rodríguez<sup>3</sup>, María Teresa Macías<sup>3</sup>

<sup>1</sup>Universidad Autónoma de Huelva, <sup>2</sup>Hospital Universitario Juan Ramón Jiménez, Huelva, <sup>3</sup>Distrito Sanitario Huelva-Costa UGC Los Rosales, Huelva

La literatura sobre la evaluación de resultados en salud generalmente se ha centrado en la evaluación del cambio estadísticamente significativo. Sin embargo, un cambio estadísticamente significativo no implica que se haya producido un cambio clínicamente significativo, esto es, que la sintomatología del paciente haya retornado a un nivel funcional tras la intervención. Resulta, pues, crucial analizar ambos aspectos y examinar cuántos pacientes se han recuperado, han mejorado, se han mantenido estables, o han empeorado tras la intervención. Los objetivos del estudio son: 1) analizar el nivel funcional de los síntomas de un grupo de pacientes al inicio del tratamiento; 2) examinar si existen diferencias entre las puntuaciones basales y de seguimiento (a los tres meses), atendiendo al cambio estadísticamente significativo y el

cambio clínicamente significativo. Participaron 63 pacientes que recibían atención psicológica en un centro de salud mental público de Huelva. El 70% (n = 44) eran mujeres. Las edades oscilaron entre 19 y 79 años (M = 45.38; SD = 16.86). Todos los participantes completaron the Inventory of Depression and Anxiety Symptoms (IDAS-II), que evalúa síntomas de depresión, ansiedad y trastorno bipolar. Para la mayoría de síntomas, más del 50% de pacientes mostraron niveles disfuncionales al inicio de la intervención. Se observaron cambios estadísticamente significativos para la mayoría de los síntomas. El porcentaje de pacientes que mostraron un cambio clínicamente significativo alcanzó el 34.21% para Disforia. Dado el carácter preliminar de los resultados, se precisa aumentar el tamaño muestral de pacientes. Este trabajo está financiado por el Ministerio de Ciencia e Innovación (proyecto PID2020-116187RB-I00), y el Ministerio de Universidades (FPU19/00144).

**Significación clínica, tamaño del efecto y potencia: los errores de tipo I y tipo II.** Marta Alcaide<sup>1</sup>, Fernando García<sup>1</sup>, Óscar F. García<sup>1</sup>

<sup>1</sup>Universidad de Valencia

Significación clínica implica generalmente que el tratamiento distancie a la persona de la población disfuncional o que la incluya en la población funcional. Esta definición se puede operativizar considerando: por un lado, el cálculo del intervalo de confianza de la población funcional, por otro lado, el intervalo de confianza de la población disfuncional, y, finalmente, el punto de corte para el cambio clínicamente significativo. Estos valores se pueden obtener a partir de la media de los grupos experimental y de control del pretest, la media de la población normal que presenta un buen funcionamiento y la desviación estándar del grupo control, de la población normal y del grupo experimental previo al tratamiento. Es necesario considerar diferentes criterios como la normalidad o el solapamiento o no de los intervalos de confianza de la población funcional y la disfuncional en cada caso. La consideración de los intervalos de confianza de la variable clínica tratada, más allá de las pruebas de significación o simplemente comparar las puntuaciones de un sujeto o grupo de sujetos que ha recibido el tratamiento presenta una manera más clara de valorar no sólo si se ha detectado un cambio, también determinar que este cambio sea clínicamente significativo.

**Análisis de supervivencia con riesgos competitivos: una aplicación práctica.** Juan José Montaña<sup>1</sup>, Elena Gervilla<sup>1</sup>, Rafael Jiménez<sup>1</sup>, Berta Cajal<sup>1</sup>, Alfonso Palmer<sup>1</sup>, Albert Sesé<sup>1</sup>

<sup>1</sup>Universitat de les Illes Balears

El estimador de Kaplan-Meier (1958) y el modelo de regresión de Cox (1972) constituyen los procedimientos más ampliamente utilizados para describir y modelar, respectivamente, el tiempo transcurrido hasta la ocurrencia de un evento de interés en el marco del análisis de supervivencia en Ciencias del Comportamiento y de la Salud. No obstante, tales procedimientos no son adecuados ante la presencia de riesgos competitivos, esto es, eventos que impiden el evento de interés, debido a que son tratados como datos censurados. Mientras que el dato censurado se limita a impedir la observación del evento de interés, un evento competitivo impide que el evento de interés se produzca por completo. En este trabajo se presenta el estimador de la incidencia acumulada (Marubini y Valsecchi, 1995) y el modelo de regresión de riesgos competitivos (Fine y Gray, 1999) como una alternativa más adecuada ante la presencia de riesgos competitivos. Para ilustrar estos procedimientos se utilizan los datos de la población de 2869 mujeres que fueron diagnosticadas de cáncer de mama entre 2006 y 2012 con un seguimiento de su estatus vital hasta 2018 procedentes del Registro de Cáncer de Mallorca. Además del seguimiento del estatus vital de las pacientes, también se han registrado una serie de indicadores clínicos que actuarán como variables explicativas en los modelos de regresión. El evento de interés es el fallecimiento por cáncer de mama, mientras que el evento competitivo es el fallecimiento por otras causas. La realización de un estudio comparativo entre los procedimientos estándar y los basados en riesgos competitivos mediante la utilización del paquete estadístico STATA, permite ilustrar la mejor adecuación de estos últimos cuando se tienen registrados eventos competitivos.

## Mesa 11

**Localización y distribución de los ítems de una medida de actitudes hacia comportamientos sexuales en mujer heterosexuales y bisexuales mediante un modelo de Rasch.** Bella González-Ponce<sup>1</sup>, Daniel Dacosta-Sánchez<sup>1</sup>, Alberto Parrado-González<sup>1</sup>, Cinta Mancheño-Velasco<sup>1</sup>, Andrea Blanc-Molina<sup>1</sup>

<sup>1</sup>Universidad de Huelva

La medición de las actitudes sexuales se ha realizado principalmente a través de escalas, siendo la TCT la que ha sustentado la mayoría de estos instrumentos. Entre los escasos instrumentos de actitudes sexuales apoyados bajo otros modelos de medida se encuentra una escala de actitudes hacia comportamientos sexuales, la cual incluye comportamientos específicos contextualizados. En un estudio se aportó evidencias de validez del modelo teórico en el que se sustenta esta escala a través de un mapa de personas e ítems proporcionado por un modelo de Rasch y en otro se relacionaron las puntuaciones de la

escala en logits con otras variables. En dichos estudios todas las personas o la mayoría eran heterosexuales. Sin embargo, podría ser que la localización y distribución de los ítems en el continuo actitudinal fuera diferente según la orientación sexual. El objetivo fue analizar la localización y distribución de los ítems de la escala de actitudes hacia comportamientos sexuales en mujeres heterosexuales y bisexuales y comprobar la existencia de DIF. Participaron 897 mujeres entre 18 y 35 años que contestaron dicha escala. Los resultados muestran que la localización y distribución de los ítems es similar en mujeres heterosexuales y bisexuales. Los ítems referentes a comportamientos sexuales online con pareja ocasional son los que más peso presentan en la medida frente a los comportamientos sexuales frecuentes con pareja estable y la masturbación en solitario que son los que menos peso presentan. Únicamente en el ítem referente al trío existe una diferencia de puntuaciones logits ligeramente superior a 0.5 entre las heterosexuales y bisexuales. Por tanto, en general, la probabilidad de respuesta de las mujeres a los ítems depende de su actitud y no de su orientación sexual.

**Aplicación de un PCA, EFA, ESEM-t y un CFA en datos de neuroimagen (MRI): explorando diferentes aproximaciones.** Elena Lacomba<sup>1</sup>, Agustín Martínez<sup>2</sup>, Alfonso Roberto Barrós<sup>1</sup>

<sup>1</sup>Universitat Jaume I, <sup>2</sup>Universidad Autónoma de Madrid

Son frecuentes los estudios correlacionales con datos de imágenes de resonancia magnética (MRI), p.ej., entre variabilidad volumétrica regional y algunas funciones cerebrales emocionales o cognitivas clásicas. Sin embargo, a pesar del importante papel del Análisis Factorial en la historia del conocimiento científico, no se ha aplicado suficientemente a datos MRI. Los estudios existentes han utilizado enfoques de reducción de datos distintos del Análisis Factorial Exploratorio (EFA), p.ej., el Análisis de Componentes Principales (PCA), o el Análisis factorial Confirmatorio (CFA). Utilizar un PCA suponiendo que se trata de un análisis factorial, es un error a veces potenciado por el abuso de las opciones de análisis de datos que aparecen por defecto en algunos programas estadísticos (p. ej., SPSS), y otras veces inconsciente por falta de conocimiento técnico. Además, se suma el mal uso del Análisis Factorial Confirmatorio (CFA), cuyas restricciones deberían emplearse para responder hipótesis más específicas. Las propuestas exploratorias factoriales (p.ej., EFA o ESEM) proporcionan estimaciones más flexibles (menos restringidas en sus parámetros) y, deberían ser las que se propongan inicialmente en estrategias de validación ante datos tan novedosos y complejos como los de MRI. Se modelaron datos de MRI de diferentes maneras (PCA, ACF y EFA/ESEM) en función de la Teoría de la Sensibilidad al Refuerzo (RST; modelo teórico de diferencias individuales de acercamiento y evitación sustentado por tres sistemas neuropsicológicos:

el Sistema de Activación Conductual (SAC); el Sistema de Inhibición del Conductual (SIC); el sistema Lucha-Huida-Paralización (FFFS). Los modelos fueron aplicados sobre un conjunto de datos de morfometría basada en el vóxel en una muestra de 300 participantes voluntarios (172 hombres; edad media 24.23, DT = 7.12; rango de edad 18-49 años). Con la intención de alcanzar una perspectiva de todo el cerebro con una clara definición de las regiones cerebrales y evitar posibles superposiciones, utilizamos la parcelación cerebral del Atlas Neuromorphometrics. Los modelos fueron valorados correspondientemente en función de los índices de ajuste resultantes (i.e.,  $\chi^2/df$ , CFI, TLI, RMSEA y SRMR). Los índices mostraron: (a) mejores valores de ajuste en las aproximaciones ESEM-t ( $\chi^2/df = 5.12$ , CFI = .913, TLI = .878, RMSEA = .117 y SRMR = .03) y EFA ( $\chi^2/df = 5.98$ , CFI = .881, TLI = .852, RMSEA = .129 y SRMR = .047); (b) que en las aproximaciones CFA ( $\chi^2/df = 5.38$ , CFI = .885, TLI = .87, RMSEA = .121 y SRMR = .063) y el PCA ( $\chi^2/df = 33.63$ , KMO = .944); (c) cargas cruzadas en el PCA; (d) elevadas correlaciones factoriales (CFA) y (e) una alta correlación entre componentes (PCA) y factores (CFA). Estos resultados sugieren que los procedimientos EFA y ESEM son una forma prometedora de modelar datos en el contexto MRI. Sin embargo, se requiere más investigación que siga utilizando aproximaciones factoriales con diferentes atlas y diferentes estrategias de parcelamiento.

**¿Cómo medir y modelar la consciencia en tareas de búsqueda visual?** Alicia Franco-Martínez<sup>1</sup>, Francisco Vicente-Conesa<sup>1</sup>, Miguel Vadillo<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Numerosos estudios sugieren que la mayor parte de los procesos psicológicos básicos pueden operar inconscientemente. Una limitación importante de esta literatura es que los investigadores suelen partir de supuestos cuestionables sobre la medición de la consciencia. El objetivo de este trabajo es doble. En primer lugar, repasaremos brevemente algunas de las medidas que se han diseñado para medir la consciencia y valorar su impacto en el desempeño de los participantes en diferentes tareas experimentales, prestando especial atención a tareas de búsqueda visual. En segundo lugar, contestaremos a la pregunta de cómo modelar el comportamiento de los participantes en estas tareas a partir de las variables dependientes (tiempos de respuesta y medidas de consciencia) e independientes (condiciones experimentales de la tarea) que tenemos a nuestra disposición. Explorar estos modelos abre la puerta a la simulación de datos experimentales, primer paso para evaluar críticamente distintas asunciones sobre cómo funciona la consciencia en los procesos involucrados en la tarea. En este trabajo, contamos con tres muestras de 80 participantes cada una, para cada uno de los cuales disponemos de 720 ensayos en la "additional singleton task". Tras explorar

diferentes modelos que conectan los tiempos de respuesta con tres medidas de consciencia diferentes (pero anidadas entre sí), comparamos su bondad de ajuste y viabilidad teórica. La comunicación pretende ser una introducción al papel de la psicometría en la psicología experimental y, al mismo tiempo, una ilustración práctica y aplicada de la misma al enigmático caso de la consciencia.

**Análisis de la influencia de los tribunales calificadoros en las pruebas de evaluación del Bachillerato para el acceso a la universidad.** Alejandro Veas<sup>1</sup>, José-Antonio López-Pina<sup>1</sup>, Elena Govorova<sup>2</sup>, Isabel Benítez<sup>3</sup>, Rubén Fernández-Alonso<sup>2</sup>

<sup>1</sup>Universidad de Murcia, <sup>2</sup>Universidad de Oviedo, <sup>3</sup>Universidad de Granada

Los procesos de evaluación constituyen una herramienta fundamental para medir el impacto de las políticas educativas. En las últimas décadas, diversos estudios han tratado de medir la calidad de las pruebas a nivel nacional e internacional, empleando diversas técnicas estadísticas para asegurar la equivalencia de puntuaciones y la medición objetiva del constructo. Dichos análisis también deben considerar la posible influencia de los examinadores o jueces, ya que pueden sesgar la distribución de puntuaciones del alumnado. Partiendo del enfoque de comparabilidad de constructo, la presente comunicación mide el impacto de los tribunales de evaluación de la evaluación del Bachillerato para el acceso a la universidad (EBAU), una de las pruebas más importantes del sistema educativo español. Partiendo de los avances y limitaciones de estas pruebas, se aplica un modelo de Rasch de facetas de las pruebas EBAU realizadas en la Comunidad Valenciana en las convocatorias de junio de 2018, incluyéndose 1000 estudiantes y hasta 54 tribunales calificadoros distribuidos entre las tres provincias (Alicante, Valencia y Castellón). En términos generales, los resultados muestran un ajuste óptimo de los datos al modelo con respecto a los tribunales, lo que indica una buena aplicación de los criterios de evaluación de las pruebas, independientemente del tipo de contenido o cuestiones planteadas. No obstante, se aprecian desajustes en algunas de las pruebas, así como un desajuste claro en el ajuste de las categorías de calificación (1-10) asociado a la falta de representatividad de las calificaciones más bajas. Ante estos resultados, se discuten las limitaciones y se proponen estrategias de mejora para la mayor eficacia de estas pruebas.

**Análisis gráfico exploratorio y análisis de variable única en las versiones original y española del BAPQ.** Carla Pérez-Guerra<sup>1</sup>, Marta Godoy-Giménez<sup>1,2</sup>, Stefan Anca<sup>1</sup>,

Andrés Soler-Martínez<sup>1</sup>, Regina Caparrós-Armand<sup>1</sup>, Pablo Sayans<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería

El Broad Autism Phenotype Questionnaire (BAPQ) fue el primer test desarrollado para evaluar el fenotipo ampliado del autismo (FAA). A pesar de ser actualmente uno de los test de FAA más traducidos y aplicados, su estructura interna se ha mostrado inestable, tanto en la versión original como en sus múltiples adaptaciones. Un estudio reciente ha mostrado que el control de la redundancia de alguno de sus ítems en la versión original podría contribuir a mejorar las evidencias de validez relacionadas con la estructura interna de sus puntuaciones. En este estudio se pretenden replicar dichos resultados en dos muestras comunitarias, una inglesa y una española (n = 533 y 970) empleando análisis gráfico exploratorio y análisis de varianza única. Empleando el paquete EGAnet se estimaron sendas redes empíricas (EBICglasso y correlaciones de Spearman). En el caso de la muestra española, se identificaron cuatro comunidades (consistencia estructural: 30%, 47%, 99% y 94%) y solapamiento entre dos ítems, mientras que en la muestra inglesa se encontraron tres comunidades (consistencia estructural: 91%, 69% y 65%) y solapamiento entre tres pares de ítems (uno de ellos el mismo que en la muestra española). En la versión española, siete ítems mostraron una estabilidad menor al 70% mientras que en la inglesa solo se observó en tres. Se reestimaron ambas redes tras fusionar los ítems redundantes en cada versión resultando en cuatro comunidades cada una (consistencia estructural: 68%, 98%, 90% y 37% en la muestra española y 99%, 21%, 94%, y 99% en la muestra inglesa). Seis ítems mostraron una estabilidad menor al 70% en la muestra española y cinco en la inglesa. Estos resultados, unidos a la desactualización en la operacionalización del FAA suponen un incentivo para proponer nuevas herramientas de medida capaces de mejorar la medición de este constructo.

## Mesa 12

**Estructura bifactor en modelos de espacio vectorial: Destilando el método Inbuilt Rubric.** José Á. Martínez-Huertas<sup>1</sup>, Guillermo Jorge-Botana<sup>2</sup>, Ricardo Olmos<sup>3</sup>, José A. León<sup>3</sup>

<sup>1</sup>Universidad Nacional de Educación a Distancia, <sup>2</sup>Universidad Complutense de Madrid, <sup>3</sup>Universidad Autónoma de Madrid

La evaluación de respuestas construidas, como los resúmenes de estudiantes, puede llevarse a cabo a través de modelos computacionales. Pero, para ello, es necesario validar los indicadores de los constructos que se miden computacionalmente. En este estudio,



presentamos la validación de estructuras bifactor en modelos de espacio vectorial a través del método Inbuilt Rubric. El método Inbuilt Rubric fue diseñado originalmente para generar rúbricas en modelos de espacio vectorial cuyo fin es evaluar o captar conceptos concretos presentes en los textos. En este trabajo se ha añadido un procedimiento para captar también nivel de conocimiento general, lo que da lugar a una estructura bifactor clásica (conocimiento general y conocimiento de contenidos concretos). Las puntuaciones generadas por este método crean lo que hemos llamado espacio vectorial jerárquico. Para ilustrar la validación de este método y esta estructura bifactor, se utilizaron 615 resúmenes de tres textos distintos generados por 205 estudiantes. Concretamente, se aplicaron modelos de ecuaciones estructurales para evaluar la validez convergente y discriminante de las puntuaciones que arroja la rúbrica del modelo computacional. Se utilizaron las evaluaciones de jueces como criterio de validez. Nos gustaría resaltar dos conclusiones principales: (1) Existen evidencias de validez convergente y discriminante en puntuaciones computacionales sin que sea necesario utilizar miles de muestras preevaluadas como sucede en modelos supervisados, y (2) los espacios vectoriales jerárquicos, como el que genera el método Inbuilt Rubric bifactor, pueden incrementar la validez de las puntuaciones computacionales a través de la evaluación de conceptos generales y específicos a la vez. En este trabajo se presentan algunos de los resultados publicados recientemente en <https://doi.org/10.3758/s13428-021-01764-6>. En el artículo original, también se proporciona código de R para aplicar el método Inbuilt Rubric en sus versiones clásicas y bifactor.

**Estimando curvas de adquisición léxica en la infancia mediante un modelo de bayesiano de TRI.** Gonzalo García-Castro<sup>1</sup>, Alicia Franco-Martínez<sup>2</sup>, Cristina Rodríguez-Prada<sup>1</sup>, Ignacio Castillejo<sup>2</sup>, Nuria Sebastian-Galles<sup>1</sup>

<sup>1</sup>Universitat Pompeu Fabra, <sup>2</sup>Universidad Autónoma de Madrid

El segundo año de vida es la etapa de la vida de una persona en la que su vocabulario aumenta más rápidamente: los bebés de estas edades pasan de conocer 15 palabras a conocer más de 400. No todas las palabras que se adquieren en este periodo se aprenden con la misma facilidad. Esta variabilidad resulta de la interacción entre las características de la propia palabra (ej., frecuencia léxica) y las características de la persona y su entorno (e.j., edad, perfil lingüístico). Existe un gran interés en desarrollar una metodología con la que estimar la edad de adquisición de una palabra en función de estas características. El objetivo de este estudio es modelar el proceso de adquisición de vocabulario en edades tempranas, concretamente en el contexto del bilingüismo catalán-castellano. Hemos recogido datos de vocabulario de 435 niños/as de entre 10 y 34 meses edad residentes en el Área Metropolitana de Barcelona. Este cuestionario

(MultiLex) consiste en un inventario de ~400 palabras en catalán y castellano en el que las personas cuidadoras respondieron si su niño/a "No comprende ni dice", "Comprende" o "Comprende y dice" cada palabra. Hemos aplicado una extensión de la Teoría de Respuesta al Ítem (específicamente un modelo bayesiano multinivel de regresión logística ordinal) para modelar las curvas de comprensión y producción de cada palabra a través de la edad de los participantes. Además, el modelo incluye predictores sobre las propiedades de las palabras y las personas que responden que pueden servir de ejemplo práctico para otros contextos aplicados. En esta comunicación presentamos la base de datos (alojada en el paquete de R multilex), el modelo, así como la curva de adquisición estimada de cada palabra visualizada mediante una aplicación Shiny interactiva de libre acceso.

**¿Por qué seguimos empecinados en utilizar el análisis factorial confirmatorio cuando hay mejores alternativas?** Juan Ramón Barrada<sup>1</sup>

<sup>1</sup>Universidad de Zaragoza

Las aplicaciones comunes del análisis factorial confirmatorio parten de un supuesto de escasa verosimilitud: las cargas factoriales secundarias son nulas en la población. Esta es técnica de elección en muchos estudios incluso en ocasiones en las que puede esperarse claramente que este supuesto no es defendible. Violaciones de este supuesto conllevan que los parámetros recuperados estén distorsionados, por lo que las decisiones a tomar a partir de estos (tanto teóricas como aplicadas) pueden ser erróneas. Estas interpretaciones equivocadas pueden no ser detectadas mediante los índices de ajuste globales habituales (CFI, TLI o RMSEA). Los modelos exploratorios de ecuaciones estructurales permiten estimar simultáneamente cargas principales, secundarias e incluir unicidades correlacionadas, esto es, son menos restrictivos. Ilustraremos cómo los modelos confirmatorios pueden llevar a errores a través de la un estudio empírico o un estudio de simulación. En el estudio empírico se reanalizan datos recientes de la Triangular Love Scale (n = 7,467; 26 países), con la que resultados previos ya apuntaban a problemas de cargas cruzadas. Se ve cómo mediante el análisis confirmatorio no se detectan problemas importantes y cambian las conclusiones. En el estudio de simulación se muestra cómo los índices de ajuste globales toleran distorsiones importantes en los modelos recuperados. Se discutirán posibles motivos e implicaciones.

**Una revisión sistemática "mixta" para abordar el análisis de la invarianza y el "gender DIF" en medidas educativas.** M. Carmen Navarro-González<sup>1</sup>, José-Luis Padilla<sup>1</sup>, Isabel Benítez<sup>1</sup>

<sup>1</sup>Universidad de Granada

Aportar evidencias de validez que apoyen las interpretaciones comparativas de las diferencias entre chicos y chicas en estudios internacionales como el PISA (e.g., OECD, 2019), es esencial para entenderlas e intervenir sobre ellas. Para apoyar interpretaciones comparativas “justas” de las puntuaciones de los tests, resulta necesario descartar la posibilidad de que esas diferencias sean producto de un artefacto del proceso de medida (AERA et al., 2014). El análisis del funcionamiento diferencial de los ítems (DIF) puede ayudar a lograr ese objetivo. Sin embargo, las técnicas de detección de DIF “tradicionales” no tienen en cuenta las posibles explicaciones contextuales del DIF (Zumbo, 2007). La investigación mixta ayuda a superar esta limitación, integrando resultados cuantitativos de los análisis de DIF con información cualitativa sobre cómo los elementos del contexto de evaluación pueden influir diferencialmente en el proceso de respuesta a los ítems (Padilla et al., 2018). El objetivo de este trabajo es mostrar cómo una “revisión sistemática mixta” puede aportar una “overview” sobre las técnicas de detección de DIF y los métodos cualitativos que pueden aplicarse para estudiar el “gender gap” en los dominios cognitivos y no-cognitivos de estudios de evaluación educativa en estudiantes de entre 12 y 18 años. Para realizar la revisión sistemática mixta se han seguido las recomendaciones PRISMA (Shamseer et al., 2014). Para adaptar la metodología de la revisión sistemática a una aproximación mixta, hemos seguido y adaptado “frameworks” y herramientas diseñadas para la investigación cualitativa y mixta, como SPIDER (Cooke et al., 2012), QATSDD (Sirriyeh et al., 2012) como herramienta de critical appraisal, y GRADE-CERQual (Lewin et al., 2015).

**Guía de informe de un análisis factorial exploratorio: Revisión bibliográfica y aplicación.** David Paniagua<sup>1</sup>, Raimundo Aguayo-Estremera<sup>1</sup>, Irene Ruiz Martín<sup>1</sup>, Marcos Romero Suarez<sup>1</sup>, María Olivares Fernández<sup>1</sup>, Jesús Alvarado<sup>1</sup>

<sup>1</sup>Universidad Complutense de Madrid

Introducción: En el campo del análisis factorial se han realizado diversas evaluaciones acerca de la adscripción de los estudios de validación a los decálogos establecidos por revistas y expertos en Psicometría. No obstante, estas normas varían entre las diferentes propuestas. Esta situación ha podido conllevar resultados diferentes acerca de las evaluaciones. Es por esto que el desarrollo de una única herramienta, con normas consensuadas por expertos y exigidas por revistas, sería útil para poder comparar estas evaluaciones. El objetivo de este estudio es: (a) realizar una propuesta de unificación de las diversas normas usadas hasta ahora en una única herramienta; (b) aplicarla a una muestra de artículos

publicados en el último año en revistas españolas indexadas en el Journal Citation Reports (JCR); y (c) comparar los resultados de la nueva herramienta con los obtenidos por los estudios anteriores. Método: Se extrajo una muestra de artículos de las revistas españolas indexadas en el JCR en 2022. La nueva herramienta se usó para evaluar el grado de adscripción de los artículos publicados a las normas y para comparar los resultados con los obtenidos en anteriores estudios. Resultados: Siguen existiendo diferencias entre los decálogos psicométricos recogidos en la literatura y en las revistas. Se ha desarrollado una herramienta de evaluación de estudios de validación, basada en la unificación de las diversas normas propuestas hasta el momento. Pese a existir decálogos publicados en revistas y recomendados por expertos, no todos los estudios de validación se ajustan a sus normas. En esta línea, la aplicación de la nueva herramienta a una muestra de estudios publicados en 2022 sugiere que los estudios de validación no siguen las normas propuestas. Discusión: En nuestra opinión, estos resultados muestran la necesidad de redoblar el esfuerzo de difusión de prácticas adecuadas en la realización del análisis factorial.

## Sesión de Pósteres 1

**Funcionamiento diferencial de la versión española para adultos de la Escala de Responsividad Social (SRS-II) en dos formatos de aplicación: diferencias entre las aplicaciones papel-lápiz y computarizada.** Marta Godoy-Giménez<sup>1,2</sup>, Stefan Anca<sup>1</sup>, Andrés Soler-Martínez<sup>1</sup>, Carla Pérez-Guerra<sup>1</sup>, Regina Caparrós-Armand<sup>1</sup>, Pablo Sayans-Jiménez<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería

La Escala de Responsividad Social (SRS-II) evalúa el trastorno del espectro autista en adultos sin retraso mental. Actualmente, no se conocen las propiedades psicométricas de su aplicación computarizada. Nuestro objetivo fue estudiarlas en las versiones lápiz-papel y online y el funcionamiento diferencial de los ítems (DIF) entre versiones. 349 universitarios completaron la versión lápiz-papel (edades 18-52 años; muestreo accidental) y 363 personas de la población general la online (edades 18-66 años; muestreo por cuotas). Aplicamos modelos de escalas de clasificación de Rasch para analizar sus propiedades psicométricas (e.g., ajuste de los ítems, fiabilidad de las escalas, severidad de los ítems) y estudiar el DIF. Se observó unidimensionalidad en cuatro de las subescalas de ambas versiones. Comunicación mostraba una posible fuente de varianza secundaria y se excluyó del análisis. Para el resto de las subescalas, solo dos ítems de Motivación online y uno de Motivación lápiz-papel mostraron ligeros problemas de

ajuste. En el escalamiento conjunto de personas e ítems, la media de los ítems estaba más de una desviación típica arriba de la de las personas en todas las subescalas. Varios ítems mostraron DIF uniforme (en Conciencia Social uno más severo online y otro en lápiz-papel; en Cognición Social tres más severos en lápiz-papel frente a uno online; y uno en Motivación Social y uno Patrones de Comportamiento Restringidos y Repetitivos ambos con mayor severidad en lápiz-papel). Se concluyó que se mantienen propiedades psicométricas adecuadas del SRS-II online. La diferencia en términos de severidad entre la media y la distribución de los ítems y las personas en el continuo advierte que los ítems expresan una severidad mucho más alta que la observada en la población general. La mayoría de los ítems se mantuvieron invariantes entre versiones. Se proporcionarán directrices para interpretar las puntuaciones en las escalas online en función de los resultados.

**Oxford Utilitarianism Scale en tiempos de pandemia por COVID-19 en España. Análisis factorial confirmatorio y diferencias por género.** Marcos Pascual-Soler<sup>1,2</sup>, José Berríos-Riquelme<sup>3</sup>, Raquel Gómez-Frías<sup>4</sup>, Dolores Frías-Navarro<sup>4</sup>

<sup>1</sup>ESIC Business & Marketing School, <sup>2</sup>ESIC University, <sup>3</sup>Universidad de Tarapacá, <sup>4</sup>Universidad de Valencia

La pandemia por COVID-19 ha supuesto una situación de alerta sanitaria mundial y un desafío importante para estudiar los dilemas morales en ese contexto de incertidumbre y amenaza para la supervivencia. Cuestiones como la escasez de los recursos médicos y la toma de decisiones sobre a quién asignarlos (por ejemplo, los ventiladores respiratorios) o el uso de los datos privados para rastrear los contagios por el virus fueron decisiones que afectaron al razonamiento moral y abrieron el debate social y político. Nuestra investigación se enmarca en el primer mes del confinamiento obligatorio en España por COVID-19 y tiene como objetivo medir el constructo del utilitarismo como explicación de actuación ante nueve dilemas morales. El utilitarismo plantea la utilidad máxima de la actuación moral, destacando que el resultado adverso afecte al menor número de personas, aunque sus consecuencias sean muy graves. La "Escala de Utilitarismo de Oxford" (Oxford Utilitarianism Scale, OUS; Kahane y cols., 2018) utiliza la perspectiva psicológica del utilitarismo en los dilemas morales para plantear un modelo de 2 dimensiones que mide sus aspectos positivos (sub-escala de beneficencia imparcial, imparcial beneficence, IB) y sus aspectos negativos (sub-escala de daño instrumental, instrumental harm, IH). La muestra está formada por 1047 participantes españoles, 364 se identifican como hombres, 678 como mujeres y 5 manifiestan tener otro género (0.5%). La media de edad es de 39.26 años (DT = 16.23). Se ha realizado la

adaptación de la escala al castellano y se presentan los resultados de su estructura teórica bifactorial (análisis factorial confirmatorio). La calidad de los resultados de consistencia interna es óptima para las dos dimensiones y el análisis factorial confirma la estructura de las dos dimensiones teóricas propuestas por los autores. El estudio se completa con el análisis de las diferencias por género donde la literatura plantea que los hombres prefieren los argumentos utilitarios frente a las mujeres, especialmente en dilemas morales personales donde el daño requiere fuerza física. Nuestros resultados con la escala Oxford señalan que los hombres obtienen una puntuación más alta en la dimensión negativa de daño instrumental mientras que las puntuaciones de las mujeres son más altas en beneficencia imparcial.

**Estructura factorial de la Escala de Adicción a Internet en población española.** F. Javier García-Castro<sup>1</sup>, Rebecca Bendayan<sup>2</sup>, María J. Blanca<sup>1</sup>

<sup>1</sup>Universidad de Málaga, <sup>2</sup>King's College London

El uso de internet se ha incrementado recientemente. De esta forma, también se ha incrementado la dependencia y su uso patológico, lo que ha sido definido como adicción a internet. Debido al impacto que esta adicción puede tener en la salud mental, se debe contar con herramientas adecuadas para su evaluación. La Escala de Adicción a Internet (EAI) contiene 6 ítems, con 5 opciones de respuesta tipo Likert (1-nunca; 5-siempre), obteniéndose una puntuación entre 6 y 30. Las puntuaciones elevadas indican una mayor adicción a internet. En la versión original se obtuvo una estructura de un único factor. Este instrumento no ha sido adaptado al español, por lo que el objetivo del presente trabajo es el de adaptar y aportar evidencias de la validez basada en la estructura interna del EAI. Participaron 700 adultos españoles con edades comprendidas entre 18 y 68 años. Se ha llevado a cabo un análisis factorial confirmatorio, ajustando un modelo de un solo factor. Los resultados muestran un ajuste satisfactorio según los índices de ajuste del modelo, aportando evidencias empíricas de su validez basada en su estructura interna. Este resultado es consistente con los resultados obtenidos con la escala original.

**Interpretar diagramas de flecha: de la formalización del algoritmo a la comprensión del operador humano.** Ana Hernando<sup>1</sup>, Antonio Lucas-Alba<sup>1</sup>, Andrés Sebastián Lombas<sup>1</sup>

<sup>1</sup>Universidad de Zaragoza

El objeto de nuestro trabajo es determinar los procesos de razonamiento que permiten comprender signos internacionales complejos de tráfico, cuyas premisas integran elementos de distinta índole (topónimos,

pictogramas, números, flechas) pero no se basan en el lenguaje verbal. En línea con las investigaciones sobre el razonamiento espacial asumimos que el conductor “comprende” en la medida en que elabora un modelo mental correcto a partir de unas premisas dadas. Sin embargo, la mayoría de las investigaciones previas se basan en problemas construidos sobre premisas verbales, como “el tenedor está a la izquierda del plato; la cuchara está a la derecha del plato; ¿se encuentra el tenedor (necesariamente) a la izquierda de la cuchara?”. Este planteamiento no se ajusta a los estreñimientos de nuestro trabajo (¿qué es posible representar en signos electrónicos?) ni a nuestros objetivos, cuyo foco principal es la localización cualitativa de eventos variables (obras, accidente, niebla) con respecto a una población. Se trata de mensajes que incorporan tres elementos: un topónimo, un pictograma y una flecha, y la tarea del conductor es determinar dónde se encuentra el evento. El uso de la flecha es conveniente por distintos motivos: resulta familiar al conductor, es versátil, fácil de exhibir en signos electrónicos y universal. Sin embargo, la flecha es esquemática y abstracta, y su significado cambia en función de los elementos con los que interactúa en el mensaje. Algunos investigadores se han preocupado por la formalización de los potenciales significados de la flecha (Eco, 1975; Tversky, 2001; Kurata y Egenhofer, 2005). En este póster vamos a aprovechar una de esas formalizaciones sintácticas de la flecha y los diagramas en los que puede intervenir (Kurata y Egenhofer, 2008) para analizar la correspondencia entre los significados esperados de ciertas combinaciones y nuestros propios resultados (Hernando et al., 2022; Hernando et al., en preparación).

**Elaboración de un cuestionario para evaluar las creencias en personas con Ortorexia Nerviosa.** Miriam Pitarch<sup>1</sup>, Juan Ramón Barrada<sup>1</sup>, María Roncero<sup>2</sup>

<sup>1</sup>Universidad de Zaragoza, <sup>2</sup>Universitat de Valencia

La ortorexia se compone de dos dimensiones: por un lado, la ortorexia saludable (OrSa), no considerada como patológica y caracterizada por un interés, comportamiento y alimentación saludable que forma parte de la propia identidad; por otro lado, la ortorexia nerviosa (OrNe), considerada la vertiente patológica y caracterizada por una obsesión por la dieta sana, donde la persona presenta una serie de restricciones progresivas, pensamientos obsesivos, conductas compulsivas y autocastigos, llegando a generar un deterioro en la calidad de vida de la persona. Dada a la relevancia que tiene la parte cognitiva en el comportamiento humano, así como teniendo en cuenta algunas de las características de la OrNe como, por ejemplo, los pensamientos obsesivos y la obsesión por la dieta, es relevante conocer con mayor profundidad esta parte. Por ello se ha desarrollado una escala de creencias

sobre alimentación saludable, con la finalidad de conocer como son los pensamientos de las personas que puntúan alto en OrNe, así como las posibles diferencias entre ambas dimensiones de ortorexia. Para la elaboración de la escala de creencias, se elaboraron en un banco inicial de 72 ítems de los cuales. Tras ser administrado a una muestra de 419 participantes junto con medidas de restricción alimentaria, se retuvieron 21 ítems. Esta escala evalúa tres dimensiones de creencias: inmoralidad de la comida no saludable, la (sobre)valoración de la alimentación y el temor a la pérdida de control con los alimentos no saludables. La dimensión más relacionada con OrNe es la de miedo a la pérdida de control, mientras que la (sobre)valoración de la alimentación es más solapada con OrSa.

**Validación de un test de eficiencia de la comprensión lectora para estudiantes universitarios españoles.**

Ladislao Salmerón<sup>1</sup>, Lidia Altamura<sup>1</sup>, M. Carmen Blanco<sup>2</sup>, Laura Gil<sup>1</sup>, Sandra Montagud<sup>2</sup>, Cristina Vargas<sup>1</sup>

<sup>1</sup>ERI Lectura, ERI Lectura, Universitat de València, <sup>2</sup>Universidad de Zaragoza

A pesar de la importancia de la habilidad lectora en adultos, su investigación resulta limitada en países de habla española por la ausencia de un instrumento válido que evalúe la comprensión lectora de población adulta en este idioma. Por ello, nuestro objetivo fue adaptar al español y validar la prueba Deep Cloze de Elbro y Jensen (2017). 453 estudiantes de las universidades de Valencia y Zaragoza completaron la versión adaptada al español de Deep Cloze. Se comprobó su validez criterial con pruebas de comprensión lectora adicionales en subgrupos muestrales y con la calificación de un examen final llevado a cabo cuatro meses después de la realización de la prueba. Los resultados indican que, aunque el Deep Cloze presenta una buena consistencia interna, las correlaciones con otras pruebas de comprensión lectora y calificaciones académicas son en su mayoría bajas. La prueba Deep Cloze cubre una laguna importante al permitir la evaluación de la comprensión lectora de estudiantes universitarios españoles. Adicionalmente, presenta importantes ventajas, como su simplicidad y tiempo de aplicación, y puede utilizarse con propósitos científicos. Finalmente, se discute la necesidad de futuras investigaciones para evaluar su validez como herramienta diagnóstica, que pudiera emplearse para dar apoyo específico a adultos con dificultades lectoras.

**Validación de la Escala de Evaluación de las Competencias en Psicometría.** Laura Galiana<sup>1</sup>, Laura Badenes-Ribera<sup>1</sup>, Sara Martínez-Gregorio<sup>1</sup>, Irene Fernández<sup>1</sup>, M. Castillo Fuentes<sup>1</sup>

<sup>1</sup>Universidad de Valencia

Introducción: Las asignaturas relacionadas con la estadística suponen un reto para los estudiantes de Grado, generando baja motivación entre el alumnado que se enfrenta a ellas. Una de estas asignaturas es la Psicometría, presente en el Grado de Psicología. En este contexto, la investigación sobre iniciativas desarrolladas desde las instituciones educativas para la mejora aumentar de la motivación es cada vez más amplia, y habitualmente tiene por fin último la mejora de las competencias del alumnado. Sin embargo, son escasas las escalas que evalúan las competencias. Objetivos: El objetivo de este trabajo es presentar evidencias de validez y fiabilidad de la Escala de Evaluación de las Competencias en Psicometría. Método: Se trata de un estudio con diseño de panel, en el que participaron 180 estudiantes de Psicología, que cursaban la asignatura de Psicometría. Se utilizaron los datos de la primera ola (septiembre 2021). Además de las competencias en Psicometría, se encuestó a los alumnos sobre su motivación académica. Los análisis incluyeron estadísticos descriptivos, un análisis factorial confirmatorio, correlaciones de Pearson para obtener evidencia de las relaciones test-criterio y estimaciones de fiabilidad. Resultados: El ajuste del análisis factorial confirmatorio que hipotetizaba una estructura de un solo factor a la base de los seis ítems de la escala fue adecuado:  $\chi^2(9) = 32.225$  ( $p < .001$ ), CRI = .971, RMSEA = .120[.077,.166], SRMR = .039. Las puntuaciones en la escala correlacionaron de forma estadísticamente significativa y positiva con las dimensiones de motivación intrínseca. Las estimaciones de fiabilidad también fueron adecuadas, con un alfa de Cronbach de .828 y un omega de McDonald de .869. Conclusión: La Escala de Evaluación de las Competencias en Psicometría ofreció evidencias de un adecuado funcionamiento psicométrico, pudiendo ser utilizada para evaluar la competencia del alumnado en esta materia.

**Adaptación española del cuestionario Dark Triad of Personality at Work (TOP).** María Solar Sastre<sup>1</sup>, David Arribas Águila<sup>1</sup>

<sup>1</sup>Hogrefe TEA Ediciones

Los estudios en la vertiente aplicada de la Psicología del Trabajo y de las Organizaciones se han centrado durante años en las características “positivas” de la personalidad, esto es, en lo que las personas manifiestan cuando tratan de dar lo mejor de sí mismas. Desde esta perspectiva, el modelo más frecuentemente citado es el de los Cinco Grandes factores, una orientación centrada en lo positivo que, como tal, reduce el alcance y la capacidad de generalización de la investigación y la práctica profesional. En la literatura reciente se aprecia un creciente interés en la búsqueda de otras variables a incluir en los modelos del desempeño profesional que permitan suplir o complementar al modelo de los Cinco

Grandes. Es en este contexto donde surge el TOP, una herramienta internacional que evalúa la triada oscura de la personalidad en el ámbito laboral. Se trata de un grupo de rasgos de la personalidad formado por el narcisismo, el maquiavelismo y la psicopatía subclínica, los cuales son bien conocidos individualmente desde la Psicología Clínica pero que profesionalmente han empezado a verse como estilos interpersonales problemáticos y negativos para el desempeño laboral. Se presentarán las principales características psicométricas de la adaptación española del TOP a partir del análisis de una muestra superior a los 1.000 casos representativa de la población general profesionalmente en activo. Además, se darán a conocer algunas evidencias de fiabilidad y validez que se han examinado durante la adaptación.

**Meta-análisis de generalización de fiabilidad del Maslach Burnout Inventory: comparación de los enfoques univariado y MASEM.** Raimundo Aguayo-Estremera<sup>1</sup>, Desirée Blázquez-Rincón<sup>2</sup>

<sup>1</sup>Universidad Complutense de Madrid, <sup>2</sup>Universidad de Murcia

Objetivo: Recientemente se ha propuesto utilizar el meta-análisis de modelos de ecuaciones estructurales (MASEM, por sus siglas en inglés) para estudiar la fiabilidad de una escala. Su principal ventaja frente al enfoque univariado reside en el análisis de la estructura interna de la escala y la elección de un índice de fiabilidad adecuado. Sin embargo, el alcance de sus resultados puede verse limitado por la imposibilidad de recabar las suficientes matrices de correlaciones inter-items. El objetivo del presente trabajo es replicar un estudio de generalización de la fiabilidad (GF) del Maslach Burnout Inventory (MBI) de 2011, realizado desde el enfoque univariado, usando las técnicas MASEM. Método/Diseño: Los autores de correspondencia de los estudios incluidos en el meta-análisis de 2011 fueron contactados por correo electrónico. Se utilizaron diferentes técnicas MASEM sobre las matrices recabadas para estudiar la estructura interna del MBI. Se ajustó un modelo de tres factores correlacionados siguiendo la propuesta de las autoras originales, se eligió un índice de fiabilidad apropiado y se calculó su promedio junto con el intervalo de confianza. Resultados: De 51 estudios primarios incluidos en el meta-análisis de 2011, tan solo se recabaron las matrices de 6 estudios (11,8%), uno de los cuales quedó descartado por el empleo de una versión reducida del MBI. Se obtuvo una matriz combinada siguiendo un modelo de efecto fijo, dado el reducido número de matrices. El modelo puesto a prueba no consiguió alcanzar índices de ajuste aceptables, obteniéndose una fiabilidad promedio inferior a la del meta-análisis univariado. Conclusiones: A pesar de sus ventajas, el enfoque MASEM como estrategia para replicar y mejorar un estudio GF puede verse limitado por la imposibilidad

de recabar las matrices de correlaciones que sirven para realizar las estimaciones combinadas, especialmente con estudios primarios relativamente antiguos.

**¿Ítems continuos o dicotómicos?: Una aproximación de diferencias individuales basada en el estilo de pensamiento holístico-analítico.** Blanca Requero<sup>1</sup>, Manuel Martín-Fernández<sup>1</sup>, Adrián Salinas<sup>2</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>IE University

En el presente estudio se examina el efecto moderador de los estilos de pensamiento holístico vs. analítico en el impacto del formato de respuesta de los ítems (continuo vs. dicotómico) sobre la experiencia subjetiva de facilidad a la hora de responder los mismos. En el estudio 1, se midió el estilo de pensamiento (holístico o analítico) de los participantes y se les asignó aleatoriamente a una condición de formato de respuesta continuo vs. dicotómico. Para esta variable independiente del formato de respuesta de los ítems se usó un paradigma de manejo de impresiones en el que se presentó a un candidato y se manipuló si las actitudes hacia dicho candidato se respondían en un formato continuo (cómo de favorable era la actitud de 1 a 7) o dicotómico (actitud favorable o desfavorable). Después se preguntó la facilidad con la que los participantes habían respondido los ítems de la manipulación del formato de respuesta. Los resultados demostraron que los participantes con un estilo de pensamiento más holístico percibían más facilidad al responder ítems continuos en comparación con ítems dicotómicos. Por el contrario, los participantes con un estilo de pensamiento más analítico percibían más facilidad al responder ítems dicotómicos en comparación con ítems continuos. En un estudio 2, se indujo el estilo de pensamiento con una tarea de procesamiento cognitivo (holístico vs. analítico) y se manipuló el formato de respuesta de los ítems (continuo vs. dicotómico) adaptando dicho formato en un test de personalidad. Se concluye que las diferencias individuales (en este caso, en el estilo de pensamiento) pueden resultar un factor relevante a la hora de responder ítems con distinto tipo de formato de respuesta.

**Invarianza factorial exacta y aproximativa de las actitudes hacia la violencia de género en 12 países latinoamericanos.** Manuel Martín-Fernández<sup>1</sup>, Ricardo Rey-Sáez<sup>2</sup>, Enrique Gracia<sup>3</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad Complutense de Madrid, <sup>3</sup>Universidad de Valencia

La violencia de género es un problema de salud pública de proporciones globales. Es el tipo de violencia que las mujeres sufren más habitualmente y conlleva graves repercusiones para su bienestar físico y psicológico. El

análisis de los datos disponibles de las grandes encuestas internacionales sobre este problema, como los de la multiple indicator cluster surveys (MICS), supone un desafío metodológico ya que estas encuestas incluyen una gran cantidad de países y presentan ítems dicotómicos con distribuciones de respuesta marcadamente asimétricas. Las técnicas habituales para evaluar la invarianza factorial exacta como el análisis factorial confirmatorio multigrupo (MG-CFA) pueden dar problemas bajo estas condiciones, lo que dificultaría la realización de comparaciones válidas entre varios países. La optimización por alignment (Asparouhov y Muthén, 2014) es una técnica de invarianza factorial aproximativa en la que se permite cierto grado de variabilidad entre los parámetros estructurales del modelo configural entre los diferentes grupos, lo que podría suponer una solución cuando no se consigue alcanzar la invarianza factorial escalar con este tipo de datos. El objetivo de este trabajo es poner a prueba la invarianza factorial de los datos del MICS sobre actitudes hacia la violencia de género en 12 países latinoamericanos, comparando los resultados de la aproximación clásica mediante MG-CFA con los obtenidos mediante la optimización alignment. Los resultados de este trabajo suponen una ilustración empírica del análisis de la invarianza factorial de grandes encuestas mediante la técnica de optimización alignment.

**Construcción de ítems de vocabulario: ¿diversidad contextual o frecuencia de las palabras?** Jennifer Pérez-Sánchez<sup>1</sup>, Ana R. Delgado<sup>1</sup>

<sup>1</sup>Universidad de Salamanca

Como parte de la estrategia de validación de las puntuaciones de los test de vocabulario es recomendable emplear fuentes objetivas, p.ej., datos procedentes de un corpus que permita conocer la frecuencia de uso de las palabras, pues esta es un buen predictor de la eficiencia del procesamiento. Una interpretación plausible del efecto de la frecuencia de las palabras señala que es la diversidad contextual (y no la frecuencia de las palabras) el mejor predictor de la eficiencia del procesamiento. Si fuera así, la construcción de ítems a partir de corpus tendría que considerar el número de documentos en que aparecen las palabras y no su frecuencia. Se ha contrastado esta hipótesis a partir de los datos de dificultad de 40 ítems de un test de vocabulario emocional y de las frecuencias de casos y documentos de los enunciados de los ítems en CORPES XXI. La correlación de la dificultad de los ítems con el número de documentos no es mayor que con el número de casos ni desde el punto de vista estadístico ni desde el sustantivo. Se concluye que considerar el número de documentos en lugar de la frecuencia de las palabras para construir los ítems no serviría para mejorar la validez de las puntuaciones.

**Desarrollo de la Escala de Preocupación por el Contagio de COVID-19 (CoCo10 + 1): por contagiarse y por contagiar.** Marcos Pascual-Soler<sup>1,2</sup>, Raquel Gómez-Frías<sup>3</sup>, José Berríos-Riquelme<sup>4</sup>, Christian Montañez-Fernández<sup>3</sup>, Dolores Frías-Navarro<sup>3</sup>

<sup>1</sup>ESIC Business & Marketing School, <sup>2</sup>ESIC University,

<sup>3</sup>Universidad de Valencia, <sup>4</sup>Universidad de Tarapacá

Uno de los efectos de la pandemia por COVID-19 ha sido el gran número de publicaciones que se ha realizado sobre dicha problemática. Concretamente, el análisis de la salud mental y la conducta de la población (seguir las normas de higiene, distancia social y aislamiento) forman parte destacada del estudio no farmacológico de la pandemia. En general, y especialmente en este tipo de situaciones donde es urgente llevar a cabo estudios y medir diversas variables, es importante disponer de instrumentos breves que midan adecuadamente los constructos. El objetivo del estudio era desarrollar una escala para medir la preocupación por el contagio por COVID-19 ya que los primeros informes y artículos publicados señalaban un nivel muy alto de preocupación por las consecuencias tan dramáticas que suponía contagiarse. Y ese grado de preocupación tenía predictores claves como el género (más alta en las mujeres) o la percepción de riesgos de salud. Después de tres investigaciones (2020, 2021 y 2022), se ha desarrollado una escala de 10 ítems que mide la Preocupación por el Contagio de COVID-19 (CoCo10 + 1), formada por 2 sub-escalas: preocupación por el contagio personal y preocupación por contagiar a los demás. La escala se completa con un ítem independiente de preocupación general por la COVID-19. La muestra está formada por 155 participantes residentes en España en enero de 2022 (38.4% hombres y 61.6% mujeres) con una media de edad 44.8 años (DT=18.4). Los resultados del análisis factorial confirman la estructura teórica de los dos factores, con valores adecuados para los índices de ajuste. La consistencia interna de las dos sub-escalas es óptima. Y la evidencia de validez se obtiene con el análisis de las diferencias por género, percepción de tener variables de riesgo de salud y riesgo por trabajo, conductas de evitación del contagio (higiene, distancia social y aislamiento). Además, las puntuaciones de la Escala CoCo10 + 1 mantienen relaciones con las variables de actitudes hacia las vacunas en general y hacia las vacunas COVID-19. Los ítems de este instrumento pueden ser adaptados a otro tipo de virus o circunstancia que implique la percepción de preocupación por el contagio.

**Mindfulness en Educación: Una nueva técnica y estrategia para la sana convivencia Escolar en Escuelas**

**de República Dominicana.** Rodolfo Canela-Fajardo<sup>1,2</sup>, Alberto Amutio-Kareaga<sup>1</sup>

<sup>1</sup>Universidad del País Vasco UPV/EHU, <sup>2</sup>Ministerio de Educación de la República Dominicana

El objetivo general del estudio fue analizar el impacto de un programa basado en Mindfulness o Atención Plena en el profesorado y alumnado de educación secundaria en el Distrito Educativo 10-02 en Santo Domingo, República Dominicana. Se seleccionó una muestra no Probabilística por conveniencia de n=323 alumnos/as, de 4to y 5to de Secundaria, de los cuales 127 participaron en el grupo experimental y 196 en el grupo control (Edad media=17 años, masculino 35% y femenino 65%), y n=174 profesores/as; 73 grupo experimental y 53 grupo control (Edad media=37 años, masculino 32.2% y 67.8% femenino). Se empleó un diseño cuasi experimental y método de comparación de grupos, con un grupo control y un grupo experimental, medidas pre-test y post-test. Los datos fueron analizados en su totalidad con el software SPSS 24.0. Los resultados evidencian una mejora significativa ( $p < 0,005$ ) en el Grupo Experimental del alumnado en Mindfulness (MAAS) ( $t(126)=9,66$ ;  $p=0,001$ ), con un efecto muy grande (d de Cohen=1,16). También en Clima de aula (EBCC) ( $t(126)=7,39$ ;  $p=0,002$ ), con un efecto medio (d de Cohen=0,46). Un gran cambio positivo en competencias socioemocionales (BarOnICE) ( $t(126)=11,33$ ;  $p=0,001$ ), con un efecto grande (d de Cohen=0,83), y un resultado moderado o leve en Empatía (TECA) ( $t(126)=4,22$ ;  $p=0,001$ ), con un efecto pequeño (d de Cohen=0,23). En cuanto al profesorado los resultados evidencian un cambio significativo en todas las variables; Mindfulness (FFMQ) ( $t(69)=18,84$ ;  $p=0,001$ ), con un efecto muy grande (d de Cohen=1,71). Gestión de estrés (ED-6) ( $t(69)=23,68$ ;  $p=0,002$ ), con un efecto muy grande (d de Cohen=1,19). Competencias socioemocionales (ECSD) ( $t(69)=26,52$ ;  $p=0,002$ ) con un efecto muy grande (d de Cohen=1,98). Clima de aula (EBCC) ( $t(69)=19,58$ ;  $p=0,001$ ) con un efecto medio (d de Cohen=0,78). En Empatía (IRI), que se observa un cambio leve moderado ( $t(69)=6,0$ ;  $p=0,001$ ) aunque con un efecto mediano (d de Cohen=0,53). Por lo tanto, a la luz de estos resultados, puede afirmarse que la intervención en mindfulness en educación resulta útil y eficaz a la hora de realizar una intervención psicoeducativa dirigida a la mejora de los niveles de Empatía, Competencias socioemocionales, gestión del estrés y Clima de aula en estudiantes y profesores en Santo Domingo, República Dominicana.

**Programas Psico-socio-educativos en Justicia Juvenil: Revisión Sistemática de las características metodológicas y su eficacia.** Mery Estefanía Buestan-Játiva<sup>1,2,3</sup>, Elena Ortega-Campos<sup>1,2,3</sup>, Leticia de la Fuente-

Sánchez<sup>1,2,3</sup>, Flor Zaldívar-Basurto<sup>1,2,3</sup>, Juan García-García<sup>1,2,3</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud (CEINSA), <sup>3</sup>Proyecto ref. P18-RT.1469 financiado por la Consejería de transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía (España) Y fondos FEDE de la Unión Europea

La conducta antisocial penada realizada durante la adolescencia ha sido objeto de atención por parte de la sociedad a lo largo de la historia. Los menores infractores que son juzgados por el sistema de Justicia Juvenil español se someten, en el marco de la medida judicial, a programas psico-socio-educativos adecuados a los factores de riesgo y necesidades que presentan. La finalidad del sistema de Justicia Juvenil, así como de los programas que se aplican durante las medidas sancionadoras-educativas es la reinserción de los menores. El objetivo de este trabajo es realizar una revisión sistemática sobre la eficacia de los programas psico-socio-educativos que se aplican a los menores infractores que pasan por el sistema de Justicia Juvenil español, siguiendo las recomendaciones PRISMA. La búsqueda de estudios se ha realizado en bases de datos electrónicas (Web of Science, Scopus, Science Direct, Redalyc, PubMed, PsycINFO y Dialnet) y se han revisado revistas científicas y otros recursos especializados en Justicia Juvenil. La búsqueda se llevó a cabo durante los meses de noviembre y diciembre de 2021. Las palabras clave utilizadas en español fueron: Justicia Juvenil, Menores infractores, Jóvenes Infractores, Programas, Evaluación y sus equivalentes en inglés. Se han incluido en este trabajo los estudios en los que se evaluaba alguna medida de la eficacia de algún programa psico-socio-educativo aplicado a menores infractores que estuvieran cumpliendo una medida sancionadora-educativa dentro del sistema de Justicia Juvenil español en el momento de aplicación del programa. Un total de 19 trabajos han cumplido los criterios de inclusión, siendo la reincidencia del menor el indicador más frecuente de eficacia del programa. Entre los resultados encontrados, destaca la escasez de programas aplicados en Justicia Juvenil con evaluación de su eficacia, siendo necesario la realización de evaluaciones de seguimiento post-intervención.

#### **Funcionamiento diferencial de los ítems del inventario de personalidad NEO-versión reducida en dos formatos de aplicación: aplicaciones lápiz-papel y computarizada.**

Carla Pérez-Guerra<sup>1</sup>, Ángel García-Pérez<sup>1</sup>, Andrés Soler-Martínez<sup>1</sup>, Antonio González-Rodríguez<sup>1,2</sup>, Marta Godoy-Giménez<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería

A pesar de haberse adaptado a múltiples idiomas y formatos, las propiedades psicométricas de la aplicación online del inventario de personalidad NEO – Versión reducida (NEO-FFI) en castellano no han sido documentadas. En este estudio, se comparó el funcionamiento diferencial de los ítems (DIF) del inventario NEO-FFI entre las versiones lápiz-papel y computarizada. Una muestra española de 349 estudiantes universitarios completó el NEO-FFI en lápiz-papel (edades 18-52 años; muestreo accidental) mientras que 363 personas españolas de la población general completaron la versión online (edades 18-66 años; muestreo intencional por cuotas). Se aplicaron modelos de escalas de clasificación de Rasch para analizar sus propiedades psicométricas (ajuste de ítems, fiabilidad de las escalas, severidad de ítems, adecuación de opciones de respuesta) y estudiar el DIF. Los resultados mostraron unidimensionalidad en tres dimensiones para ambos formatos. Extraversión y Responsabilidad presentaron una posible fuente de varianza secundaria que impidió profundizar en el análisis. Se observó un ligero desajuste en un ítem de Amabilidad online, en uno de Neuroticismo lápiz-papel y en uno de Apertura lápiz-papel. Las medias y distribuciones de los ítems y las personas en los continuos mostraban posiciones similares. Solo tres ítems de Neuroticismo presentaron ausencia de DIF. El resto de los ítems mostraron DIF con tamaños entre moderados (3 ítems) y grandes (29 ítems). Estos resultados impiden la comparación directa de las puntuaciones de las versiones papel y lápiz y online del NEO-FFI. Se proporcionarán directrices para la interpretación de las puntuaciones en las escalas online en función de los resultados obtenidos.

#### **Análisis bajo la Teoría de Respuesta al Ítem del Cuestionario de Autoconcepto (CAG).** José Augusto Miralles Sacristán<sup>1</sup>, Ana Martínez Dorado<sup>1</sup>

<sup>1</sup>Giunti Psychometrics S.L.

El Cuestionario de Autoconcepto (CAG) es una herramienta de evaluación educativa publicada por Giunti Psychometrics S.L que contiene escalas de cinco tipos de autoconcepto y una de sensación de control. En 2022 se publicó su tercera edición, con baremos actualizados y nuevos análisis con una muestra (n=1659) de uso real de la prueba. Es particularmente interesante el hecho de que, por primera vez, se ha analizado bajo el paradigma de la Teoría de Respuesta al Ítem (TRI). Se evaluó su idoneidad para este análisis (unidimensionalidad) mediante dos Análisis Factoriales Confirmatorios (AFC) en el que se puso a prueba un modelo de segundo orden y se realizó una validación cruzada (RMSEA=.038;.04; SRMR=.057;.057; CFI=.952;.948; TLI=.949;.945). Se analizó su ajuste al Modelo de Respuesta Graduada (MRG) de Samejima



(1997) de los ítems de las escalas mediante el índice S-X2 de Orlando y Thissen (2000) obteniendo resultados mixtos en alguna de ellas. Se analizó el comportamiento diferencial de los ítems (DIF) según el sexo de los participantes mediante el método de la regresión logística y la comparación de modelos. Se halló que un ítem de la escala de autoconcepto familiar presentaba DIF de tipo no uniforme ( $Pr(c_{21} 3, 2) = .003$ ;  $Pr(c_{22} 3, 1) = .001$ ). Se observó que en este ítem las mujeres tenían, significativamente, una probabilidad mayor de elegir opciones de respuesta altas que los hombres a niveles de rasgo ( $\theta$ ) bajos. Por último, se analizaron las funciones de información y error del test, hallando que se trataba de un test más informativo en niveles de rasgo bajos ( $-4 - 0$   $\theta$ s), pero que mantiene una buena discriminación a lo largo de todo el espectro del rasgo debido a su bajo error.

**Interpretación de resultados estadísticos: pensamiento dicotómico, de estimación y meta-analítico.** Raimundo Aguayo-Estremera<sup>1</sup>, Iván Sánchez-Iglesias<sup>1</sup>, David Paniagua<sup>1</sup>, Alejandro Miguel-Alvaro<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

**Introducción:** Los contrastes de hipótesis basados en la significación de la hipótesis nula (NHST, por sus siglas en inglés) y las técnicas de estimación son dos procedimientos inferenciales habituales en Psicología. Desde el área de la Statistical Cognition se ha argumentado que implican procesos cognitivos distintos. Mientras que con los NHST la forma de razonar es de tipo dicotómico, con las técnicas de estimación, como los intervalos de confianza (IC), el razonamiento se centra en la magnitud del efecto. El objetivo de la presente investigación es evaluar si hay diferencias en la forma de razonar sobre resultados estadísticos en función de la información presentada (NHST, IC o meta-análisis). **Método/Diseño:** Se usó un diseño unifactorial intrasujetos, con una muestra de estudiantes de Psicología a los que se les planteó una investigación ficticia: dos estudios evaluaban la eficacia de un tratamiento para mejorar la calidad del sueño. Los estudiantes contestaron si ambos estudios sugerían resultados consistentes (a favor del tratamiento), ambiguos o inconsistentes. Realizaron la tarea tres veces, una por cada condición del factor: (a) resultados del NHST; (b) gráficos con IC; (c) resultados de una síntesis meta-analítica. **Resultados:** Las respuestas variaron en función de la información presentada. A pesar de que los resultados de la investigación ficticia indicaban que el tratamiento era eficaz, tanto con el NHST como con los IC la conclusión más frecuente fue "resultados ambiguos". No obstante, el porcentaje de respuestas en "resultados concluyentes" fue mayor con los IC. Con el meta-análisis, el 80% aproximadamente indicó "resultados consistentes". **Conclusiones:** Los resultados

sugieren que los estudiantes no fueron capaces de aprovechar la información adicional de los IC y superar la aparente contradicción mostrada con el NHST. La síntesis meta-analítica consiguió reducir la incertidumbre en torno a los resultados.

**Propiedades Psicométricas de la Escala de Inteligencia Emocional (TMMS-24) en mujeres diagnosticadas de cáncer de mama.** Lorena M. Soria-Reyes<sup>1</sup>, María Victoria Cerezo<sup>1,2</sup>

<sup>1</sup>Universidad de Málaga, <sup>2</sup>Universidad de Educación a Distancia (UNED)

El cáncer de mama es una enfermedad que influye en el bienestar de las personas diagnosticadas, aunque muchas lo afrontan adaptativamente. Entre las variables que influyen en este afrontamiento destacan la inteligencia emocional (Guil et al., 2020). El objetivo del estudio ha sido comprobar las propiedades psicométricas de la versión breve española Trait Meta Mood Scale (TMMS-24; Fernández-Berrocal et al., 2004) que mide inteligencia emocional, basada en la escala original de Salovey et al. (1995). Dicha escala está comprendida por tres subescalas: atención emocional, grado de percepción acerca de los pensamientos y sentimientos; claridad, capacidad percibida sobre el entendimiento de los sentimientos, y reparación, creencias sobre el manejo de las emociones negativas y focalización en las emociones positivas. Una muestra de 103 mujeres diagnosticadas de cáncer de mama cumplieron la escala TMMS-24 y las siguientes escalas: autoestima de Rosenberg, resiliencia (CD-RISC-10; Connor y Davison, 2003) y un ítem único de felicidad. Su estructura factorial se comprobó mediante análisis factorial confirmatorio y la evidencia de validez mediante correlaciones de Pearson. Siguiendo la base teórica, el análisis factorial confirmatorio realizado con los tres factores de la TMMS-24, ofreció un RMSEA de 0.99, mostrando una bondad de ajuste mejorable, aunque otros índices de ajuste como el CFI = 0.93 y TLI = 0.92, han sido aceptables. La fiabilidad se calculó mediante el alfa de Cronbach obteniendo 0.88 para atención emocional y para claridad, y 0.91 para reparación. La evidencia de validez basada en las relaciones con los factores claridad y reparación emocional han sido positivas y significativas con felicidad, resiliencia y autoestima, mientras que la atención emocional no ha correlacionado con ninguna de ellas. En conclusión, la escala TMMS-24 mostró que sus propiedades psicométricas son adecuadas en una muestra de mujeres diagnosticadas de cáncer de mama, aunque conviene ampliar la muestra para seguir mejorando su aplicabilidad clínica.

**Propiedades psicométricas empleando modelos de Rasch de la adaptación computarizada de la versión española de la Escala de Soledad de Jong Gierveld.** Carla Pérez-Guerra<sup>1</sup>, Stefan Anca<sup>1</sup>, Andrés Soler-Martínez<sup>1</sup>, Regina Caparrós-Armand<sup>1</sup>, Antonio González-Rodríguez<sup>1,2</sup>, Marta Godoy-Giménez<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería

La Escala de Soledad de Jong Gierveld (DJGLS) es la escala de soledad en adultos más utilizada en Europa. Desde su primera adaptación al castellano, varios estudios han señalado que es una escala mayoritariamente unidimensional con buena fiabilidad y evidencias de validez en relación con otras variables. En este estudio se aplicó la DJGLS de manera online y se estudió la dimensionalidad de las puntuaciones de la escala y si sus propiedades psicométricas se mantenían en formato computarizado haciendo uso de modelos de Rasch (ajuste de los ítems, fiabilidad de la escala, severidad de los ítems, adecuación de las opciones de respuesta). Una muestra de 363 personas (18-66 años; muestreo incidental por cuotas) completó la versión española de la DJGLS a través de un enlace. Si bien se observó una posible fuente secundaria de varianza (varianza explicada por la medida: 42%; varianza explicada por los ítems 19.7%; varianza explicada por el primer contraste: 12.3% [eigenvalue: 2,32]), se concluye que la escala es esencialmente unidimensional. La fuente secundaria se corresponde con un efecto de método generado por los ítems invertidos. La correlación desatenuada de este clúster de ítems con el principal fue .95. Se señalaron adecuadas propiedades psicométricas (ningún ítem mostró problemas de ajuste; la distribución de los ítems y las personas en el continuo mostró ítems distribuidos a lo largo del continuo; la media de los ítems estaba por encima de la media de las personas; las categorías de respuesta funcionaban adecuadamente). Estos resultados se alinean con estudios previos donde también se concluye que la DJGLS es esencialmente una escala unidimensional con una fuente secundaria de varianza debida a un efecto de la redacción de los ítems. Las propiedades psicométricas de la escala se mantienen en su aplicación online.

**Desarrollo de un agente conversacional para el aprendizaje y la evaluación de contenidos de metodología de la investigación y estadística.** Jose Manuel Caperos<sup>1</sup>, Jose Ángel Martínez-Huertas<sup>1,3</sup>, Carlos Martínez-de-Ibarreta<sup>1</sup>, Ricardo Olmos<sup>2</sup>, Guillermo de-Jorge-y-Botana<sup>4</sup>

<sup>1</sup>Universidad Pontificia Comilla, <sup>2</sup>Universidad Autónoma de Madrid, <sup>3</sup>Universidad Nacional a Distancia, <sup>4</sup>Universidad Complutense de Madrid

La metodología de la investigación y la estadística son materias con creciente importancia en la formación de alumnos en casi todas las áreas de conocimiento. Sin embargo, habitualmente son asignaturas consideradas difíciles y en muchos casos poco valoradas, especialmente en aquellas titulaciones donde el alumnado ve distante su vinculación con la práctica profesional. De hecho, las asignaturas con una carga alta de estadística, como el análisis de datos, la psicometría o la econometría, se relacionan con un elevado estrés en el alumnado (lo que se conoce como statistics anxiety). Los agentes conversacionales son sistemas automáticos que responden de forma inteligente a las demandas de los usuarios permitiendo una interacción fluida en forma de conversación mediante texto escrito o voz. El uso de estos sistemas está generalizado a distintos dominios por ser un formato amigable, flexible y sencillo en el que resolver dudas e intercambiar información. En los contextos académicos se utiliza fundamentalmente para la auto y heteroevaluación y para el aprendizaje. Presentamos un piloto de chatbot que permite conversar sobre contenidos de metodología de la investigación y estadística permitiendo la profundización en el aprendizaje, y la evaluación del propio conocimiento.

**Adaptación de la escala Self-Efficacy for Appropriate Medication Use Scale (SEAMS).** Marta Godoy-Giménez<sup>1,2</sup>, Ángel Pérez-García<sup>1,2</sup>, Stefan Anca<sup>1</sup>, Regina Caparrós-Armand<sup>1</sup>, Fernando Cañadas<sup>1,2</sup>, Angeles F. Estévez<sup>1,2</sup>, Pablo Sayans-Jiménez<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería

La falta de adherencia al tratamiento se encuentra influida por factores como la percepción de autoeficacia sobre los cuidados. La Escala sobre la Percepción de la Autoeficacia para la Adherencia al Tratamiento (SEAMS) es una escala de 13 ítems que fue diseñada con el objetivo de que el personal sanitario conociera y comprendiera los problemas y asuntos personales que podrían estar dificultando la adherencia al tratamiento del paciente. El objetivo de este estudio fue realizar la traducción y adaptación cultural de la SEAMS al castellano. Se llevó a cabo un protocolo de traducción y adaptación siguiendo las directrices de la Comisión internacional de test. El proceso implicó las siguientes fases: 1) dos traducciones ciegas inglés-castellano de los ítems; 2) juicio de expertos para conciliar las versiones de ambas traducciones y proceder con la adaptación cultural; 3) un estudio piloto con una muestra de 11 participantes mayores de 60 años (población objetivo final) para recoger posibles incidencias y modificar aquellos aspectos que fuesen necesarios; 4) una retro-traducción ciega castellano-inglés; 5) estudio comparativo de las traducciones y retro-traducciones

por un tercer traductor bilingüe (donde además se incluyó la discusión sobre las adaptaciones culturales); y 6) juicio de expertos final para valorar la similitud entre las versiones original y adaptada. En estos momentos la escala está siendo aplicada a una muestra de personas polimedicadas mayores de 60 años. Se ofrecerán los ítems y las instrucciones finales, así como los datos resultantes del análisis de las propiedades psicométricas de la versión en castellano de la SEAMS. Para ello, se aplicarán modelos de escalas de clasificación de Rasch. La idoneidad de las propiedades de la escala determinará si puede ser empleada para comprobar la eficacia de una App para la mejora de la adherencia al tratamiento y las creencias de autoeficacia sobre auto-cuidados en personas mayores polimedicadas.

**Adaptación al castellano de las escalas de creencia en un mundo justo de Dalbert.** Andrés Soler-Martínez<sup>1</sup>, Marta Godoy-Giménez<sup>1,2</sup>, Stefan Anca<sup>1</sup>, Regina Caparrós-Armand<sup>1</sup>, Carla Pérez-Guerra<sup>1</sup>, Pablo Sayans-Jiménez<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería

La Creencia en un Mundo Justo (CMJ) se define como la necesidad básica de creer que vivimos en un mundo donde cada uno obtiene lo que merece. Se relaciona con los procesos de adaptación, logro de objetivos a largo plazo y bienestar subjetivo. Ante situaciones de injusticia, la necesidad de restablecer esta creencia se ha relacionado con fenómenos graves como la derogación de la víctima y la autoinculpación. A pesar de ser un constructo ampliamente utilizado en el ámbito internacional, se carece de medidas en castellano. El objetivo de este estudio es llevar a cabo la adaptación de las escalas general y personal de CMJ propuestas por Dalbert. Para ello se contó con la colaboración de tres traductores profesionales (dos nativos de EE.UU y uno español). Dos de ellos realizaron sendas traducciones de los ítems de la escala al castellano y, tras un juicio de expertos (en el que participó la autora de la escala original), el tercero tradujo las propuestas de ítems al inglés. Durante el juicio de expertos, en contra del criterio de la autora, se propuso eliminar parte del contenido de algunos ítems (“generalmente”, “suelo”, “la mayoría”, etc.). Para comprobar el efecto de esta supresión y explorar las propiedades psicométricas de las dos escalas se aplicaron los ítems con el contenido original y con el contenido suprimido en una muestra (n = 182) y se emplearon modelos de escalas de clasificación de Rasch. Los resultados muestran que la omisión de contenido generó que la media de severidad de los ítems superase la media de la muestra, en contraste con la propuesta original. Además, los indicadores de ajuste permitieron detectar un ítem cuya redacción debe mejorarse. En general, las escalas presentan buen funcionamiento, pero

el sobreajuste de algunos ítems apunta a una excesiva redundancia de contenido.

**Validación de una escala de motivación hacia la lectura en alumnado de educación primaria.** Amelia Mañá<sup>1</sup>, Cristina Vargas<sup>1</sup>, Mario Romero<sup>1</sup>, Lidia Altamura<sup>1</sup>

<sup>1</sup>Universidad de Valencia

Junto con el aprendizaje de la lectura es importante el desarrollo de una actitud y motivación favorable que promueva buenos hábitos lectores. La Teoría de la Autodeterminación (SDT, Ryan y Deci, 2019), basada en la necesidad de sentirnos responsables de nuestras propias acciones, ha servido de marco para el estudio de la motivación académica en general y hacia la lectura en particular (de Naeghel et al., 2012). La SDT distingue entre motivación intrínseca (autodeterminada), integrada, identificada, introyectada y extrínseca (heterodeterminada). Trabajos previos han desarrollado escalas para su medición, pero aún no disponemos de una versión para alumnos españoles de los últimos años de educación Primaria. Guay et al. (2010) diseñaron la Elementary School Motivation Scale (ESMS) para medir, entre otras, la motivación hacia la lectura en alumnado canadiense, equivalente a los cursos de 1º a 3º de primaria. Chanal y Guay (2015) también adaptaron dicha escala para alumnado de 5º de primaria. En el contexto español, De Sixte et al. (2020) utilizaron la ESMS en estudiantes de 1º a 3º, sin embargo, no existe aún ninguna prueba de estas características para población española de 4º y 5º. El objetivo de este trabajo es validar la versión española de la ESMS sobre lectura de Chanal y Guay et al. (2015), para alumnado de 4º y 5º de EP. Se ha planificado la recogida de datos de aproximadamente 2000 alumnos (actualmente ya se han evaluado a 1102, de 9 centros educativos (6 concertados/privados y 3 públicos) de la provincia de Valencia). En el congreso se presentarán las propiedades psicométricas de la adaptación por medio de análisis factoriales exploratorios y confirmatorios, así mismo, se valorará la fiabilidad de la escala para cada curso.

**Versión adaptada de la Escala de Socialización Étnica Familiar: propiedades psicométricas aplicando un modelo politómico de Rasch.** Jorge L. Ordóñez-Carrasco, María Sánchez-Castelló<sup>1</sup>, Claudia Suárez-Yera<sup>1</sup>, Antonio J. Rojas Tejada<sup>1</sup>

<sup>1</sup>Universidad de Almería

El incremento de adolescentes y jóvenes de origen inmigrante en España en los últimos años se ha visto reflejado en un aumento de estudios psicosociales centrados en los procesos migratorios de esta población. Por ello es de gran relevancia contar con buenas medidas de las principales variables en este campo, como es la

socialización étnica familiar (i.e., los esfuerzos de padres, madres y otros miembros de la familia para exponer a los jóvenes a los valores y comportamientos de su cultura de origen). El objetivo de este trabajo fue analizar las propiedades psicométricas de la adaptación al contexto español de la versión revisada de la Escala de Socialización Étnica Familiar. Los participantes fueron 670 adolescentes de origen inmigrante (350 de origen marroquí, 136 de origen rumano y 184 de origen ecuatoriano), con edades entre los 12 y 19 años ( $M = 15,17$ ;  $SD = 1,40$ ), y un 56,10% de chicas. Se aplicó un modelo politómico de Rasch (i.e., Rating Scale Model) a las puntuaciones obtenidas en los tres grupos de adolescentes, y se analizó el funcionamiento diferencial de los ítems (DIF). A la vista de los resultados, los índices de ajuste de los datos al modelo (infit y outfit) fueron adecuados en las tres muestras de adolescentes. Las propiedades psicométricas se pueden considerar adecuadas en los tres grupos tomadas de forma independiente. Sin embargo, se observó un funcionamiento diferencial (DIF) en seis de los 12 ítems, esto es, presentaban diferencias en función del grupo etnocultural no debidas a impacto. En definitiva, si bien las propiedades psicométricas son apropiadas, la existencia de DIF en la mitad de los ítems de la escala revela un potencial sesgo que debe ser atendido.

**Monitorización del cambio de significado basado en alineación de espacios semántico-vectoriales de noticias de distintos años.** MGuillermo Jorge-Botana<sup>1</sup>, José Ángel Martínez-Huertas<sup>2</sup>, José Javier Olivas<sup>3</sup>, Ricardo Olmos Albacete<sup>4</sup>, Alejandro Martínez-Mingo<sup>5</sup>

<sup>1</sup>Universidad Complutense de Madrid, <sup>2</sup>Universidad Nacional de Educación a Distancia, <sup>3</sup>London School of Economics and Political Science, <sup>4</sup>Universidad Autónoma de Madrid, <sup>5</sup>Instituto de Empresa

El cambio de significado de las palabras es una información de interés para medir el efecto que tienen en los juicios sobre ellas y la comprensión de los marcos en los que participan. Tradicionalmente, la tarea de monitorizar las palabras en los medios se hace por medio de índices basados en la frecuencia. Se calcula la frecuencia de la palabra en una muestra textual correspondiente a un año y se compara con los demás años. Este procedimiento es frutífero, porque proporciona información de la importancia de la palabra en distintos años. Sin embargo, no considera el cambio de significado a lo largo de los años. En este estudio se proponen índices sensibles tanto a la importancia temporal de las palabras como a su cambio de significado. Estos índices son posibles gracias a tener un espacio semántico vectorial de muestras textuales de cada año y realizar con ellos un alineamiento mediante una rotación Procrustes para calcular medidas de similitud entre vectores de distintos espacios. Las

muestras son todas las noticias aparecidas desde el año 2012 al 2019 en El Mundo y El País. Esto genera una base de 70.000 palabras con valor en los siguientes índices: saliencia de la palabra en cada año, diversidad semántica en cada año, continuidad de significado de un año para otro, resonancia del significado de los años anteriores en el último año, fragmentación del significado a lo largo de los años, además de la representación en forma de nube de palabras de su contorno semántico en cada uno de los años. Mostramos como ejemplo los índices de relevancia y cambio de significado para algunas palabras relacionadas con el “procés”, acaecido justo en los años en que se recogen las muestras y tomado como ocasión en la que consignar variaciones en los marcos de significado conceptual.

**Optimización de la estimación de tamaños del efecto a través de p-uniform: comparación de estadísticos.** Juan I. Durán<sup>1</sup>, Manuel Suero<sup>2</sup>, Juan Botella<sup>2</sup>

<sup>1</sup>Universidad a Distancia de Madrid, <sup>2</sup>Universidad Autónoma de Madrid

La preferencia por publicar estudios con resultados significativos tiene como consecuencia que las estimaciones metaanalíticas del tamaño del efecto sean sobreestimadas. Entre los métodos propuestos para detectar y corregir este efecto se encuentra p-uniform, que recurre a: (1) los valores p de los estudios primarios con resultados significativos; y (2) que, en ausencia de sesgo de publicación, la distribución de dichos valores p es uniforme. Durán, Suero y Botella (2021) propusieron una serie de modificaciones al procedimiento original con las que optimizar las estimaciones obtenidas con él. Las principales modificaciones consisten en recurrir a una distribución t de Student no centrada para calcular los valores p y en la transformación de esos valores al intervalo [0 - 1] recurriendo a esa misma distribución. En dicha propuesta, los valores p se resumen utilizando el estadístico de Fisher, con distribución muestral conocida bajo distribución uniforme de los valores p. Se propuso estimar el valor del tamaño del efecto como aquel al que corresponde un valor para el estadístico de Fisher igual a la mediana de la distribución correspondiente, restando un factor de corrección. Dado que existen otros estadísticos, como el de Irwin-Hall, para valorar el ajuste de una distribución a un modelo uniforme, en el presente trabajo los implementamos en el método de p-uniform optimizado y los ponemos a prueba en un estudio de simulación. En dicho estudio se valoran como factores el tamaño muestral, el número de estudios primarios incluidos en el metaanálisis y la magnitud del tamaño del efecto. Obtenidas las estimaciones con las implementaciones correspondientes a cada estadístico, presentamos resultados de la comparación de sus rendimientos según las condiciones.

**Traducción, adaptación cultural y validación al español del Teenagers' Quality of Life (T-QoL) para adolescentes con enfermedades de la piel.** Laura Alonso-Naranjo<sup>1</sup>, Emilio Berna-Rico<sup>2</sup>, Carlota Abbad-Jaime de Aragón<sup>2</sup>, Raúl Castañeda-Vozmediano<sup>3</sup>, Lola Prieto-López<sup>3</sup>, Ana Isabel Sánchez-Moya<sup>1</sup>, Cristina Pérez-Hortet<sup>1</sup>, Álvaro González-Cantero<sup>2,3</sup>

<sup>1</sup>Hospital Universitario de Toledo, <sup>2</sup>Hospital Universitario Ramón y Cajal, <sup>3</sup>Universidad Francisco de Vitoria

**Introducción.** El test de calidad de vida de los adolescentes (T-QoL) es un cuestionario específico utilizado para cuantificar el impacto de las enfermedades de la piel en la calidad de vida (QoL) de los mismos. A día de hoy, no se dispone de una versión validada en español por lo que el objetivo de esta investigación es traducir, adaptar culturalmente y validar el T-QoL al castellano. **Métodos.** La traducción y adaptación cultural se llevó a cabo en base a las directrices ISPOR (Professional Society for Health Economics and Outcomes Research), realizando un estudio prospectivo con 133 pacientes para su validación. La validez fue estudiada con respecto a su estructura factorial, por medio de Análisis Factorial Confirmatorio, y su convergencia con otras medidas objetivas y validadas. La fiabilidad se analizó en términos de consistencia interna y test-retest. **Resultados.** La validez convergente quedó demostrada por una correlación significativa con la calidad de vida dermatológica (DLQI) y la calidad de vida dermatológica infantil (CDLQI) ( $r=0,75$ ) y con la pregunta global (GQ) sobre la gravedad de la enfermedad autoevaluada ( $r=0,63$ ). Los resultados del análisis factorial confirmatorio fueron consistentes con los encontrados por los autores de la prueba original. Los indicadores de fiabilidad revelan valores pronunciadamente altos, como el  $\alpha$  de Cronbach= $0,89$ , el índice Lambda 6 de Guttman ( $G6$ )= $0,91$  y el Omega  $\omega=0,91$ . El test-retest mostró una alta estabilidad temporal en las respuestas ( $ICC=0,85$ ). **Conclusión.** El T-QoL versión española es un cuestionario válido y fiable para la evaluación de la calidad de vida de adolescentes hispanohablantes con enfermedades de la piel.

**Adaptación del Informant Questionnaire on Cognitive Decline in the Elderly en la encuesta Survey of Health, Aging and Retirement in Europe.** Irene Fernández<sup>1</sup>, Arne Bethmann<sup>2</sup>, Jose M. Tomás<sup>1</sup>

<sup>1</sup>Universidad de Valencia, <sup>2</sup>Technical University of Munich / Munich Economics of Aging, Max Planck Institute for Social Law and Social Policy

**Introducción.** El incremento de casos de demencia en los últimos años ha dado lugar al fomento de diagnósticos precoces. La escala Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) se ha convertido en una herramienta popular para detectar demencias. Además, encuestas poblacionales como el Health and Retirement Study han empleado también este instrumento. Recientemente, la encuesta Survey of Health, Aging and Retirement in Europe (SHARE) realizó una adaptación del IQCODE. **Objetivo.** Examinar la validez factorial, fiabilidad y validez de criterio de esta adaptación del IQCODE. Además, poner a prueba la invarianza de medida por género. **Método.** Los datos provienen de la ola 8 de SHARE. La muestra estuvo compuesta por 1059 personas (Medad = 79.26, DTedad = 10.53), de las cuales el 48.2% eran mujeres. La adaptación de la escala de SHARE incluyó 7 ítems, utilizó dos años para cambios en habilidades mentales, y empleó una escala de respuesta de tres anclajes. Se empleó Análisis Factorial Confirmatorio (AFC) para probar la estructura de un factor, así como la invarianza de medida por género a nivel configural, métrico, escalar y estricto. La fiabilidad se estimó empleando el coeficiente omega, y la validez de criterio comparando la puntuación media del IQCODE entre individuos con y sin diagnóstico de Alzheimer. **Resultados.** La estructura de un factor mostró un buen ajuste a los datos:  $c2(14)=218.92$ ,  $p < .001$ , CFI= .995, SRMR= .028. La consistencia interna fue adecuada,  $w = .94$ , y se encontró evidencia de validez de criterio,  $t(1009.04)= 20.61$ ,  $p < .001$ ,  $d = 1.16$ . Asimismo, se estableció la invarianza de medida completa. **Conclusiones.** Esta adaptación del IQCODE mostró propiedades psicométricas satisfactorias, lo que garantiza su uso como herramienta adecuada para la detección del deterioro cognitivo en el contexto de encuestas poblacionales. Se discuten las limitaciones y las directrices a futuro.

**Grupos focales presenciales versus online: Una revisión sistemática.** Chuen Ann Chai<sup>1</sup>, Alba Aza<sup>1</sup>, Georgina Guilera<sup>1</sup>, Juana Gómez-Benito<sup>1</sup>, Maite Barriosa<sup>1</sup>

<sup>1</sup>Universitat de Barcelona

**Introducción/objetivos.** En el contexto actual cada vez es más frecuente la utilización de las nuevas tecnologías en el desarrollo de la investigación, como es el caso de los grupos focales (GF) online. Sin embargo, sus particularidades y posibles diferencias en los resultados con respecto a los GF presenciales no están aún esclarecidas. Por ello, el objetivo de la presente revisión sistemática es analizar estas diferencias y determinar bajo qué condiciones es mejor seleccionar GF online o presenciales. **Metodología.** Se ha llevado a cabo una revisión sistemática utilizando la base de datos Scopus con los descriptores de "focus group" AND ("in-person" OR "conventional" OR "traditional" OR "face-to-face")

AND (“online” OR “virtual” OR “internet”). Los criterios de inclusión fueron: estudios que (1) incluyeran datos de al menos un GF online y otro presencial y (2) estuvieran publicados hasta abril de 2022. Se excluyeron todos aquellos que no eran artículos científicos y que no estaban publicados en castellano o inglés. Resultados. En total 84 artículos cumplieron los criterios de inclusión/exclusión. La publicación más reciente data del año 1998, pero el año con mayor número de publicaciones es 2021 (18%). Mayoritariamente, las publicaciones estaban vinculadas al área de las ciencias sociales (45%) y medicina (39%), pero también en menor medida a economía (9,5%), enfermería (9,5%) y psicología (8,3%). Los resultados muestran diferencias en la cantidad y extensión de las contribuciones de los participantes, en la conexión emocional, la diversidad de temas planteados y en el surgimiento de roles asimétricos, entre otras. Discusión. Estos resultados evidencian importantes diferencias entre ambos procedimientos, por lo que es imprescindible profundizar y ampliar en esta búsqueda para disponer de información actualizada y sistemática que le dé un mayor rigor científico al uso de la metodología online. Esto permitirá, asimismo, disponer de guías prácticas de utilización de ambas metodologías.

**Evaluación de la recuperación en salud mental: Adaptación española de la Maryland Assessment of Recovery Scale (MARS-12).** Arantxa Gorostiaga<sup>1</sup>, Nekane Balluerka<sup>1</sup>, Jone Aliri<sup>1</sup>, Ana González-Pinto<sup>1</sup>

<sup>1</sup>Universidad del País Vasco UPV/EHU

Desde el modelo de la recuperación en salud mental, el bienestar supone la posibilidad de desarrollar una vida satisfactoria que responda a las expectativas y valores de la persona y que permita realizar contribuciones a la sociedad independientemente de la ausencia o presencia de síntomas. En el presente trabajo se proporcionan los primeros resultados acerca de las propiedades psicométricas de la versión española de la MARS-12 (Drapalski et al., 2012), un instrumento que evalúa la recuperación en base a dicho modelo. En la traducción de la escala participaron seis personas (cuatro expertas en el modelo de recuperación y en la elaboración de instrumentos, una traductora profesional y una filóloga). En la fase inicial del estudio, se llevaron a cabo entrevistas cognitivas con 8 personas de la población objetivo y se consultó con 12 expertos/as en salud mental. La versión obtenida en dicha fase, se aplicó junto con la batería necesaria para la validación a una muestra compuesta por 86 personas (44,2% mujeres) que llevan padeciendo un trastorno mental severo durante un periodo mínimo de 2 años, usuarias de servicios de salud mental del País Vasco, con una edad comprendida entre los 19 y los 83 años (M=47,12; DT=13,36). El instrumento presentó una

estructura unidimensional (CFI=0,981; TLI=0,977) y una adecuada consistencia interna (alfa de Cronbach=0,94). La correlación entre la puntuación obtenida en el MARS-12 y la obtenida en el Cuestionario sobre el Proceso de Recuperación (QPR-ESP; Neil et al., 2009; adaptación española de Goodman-Casanova et al., 2021) fue de 0,85, lo que indicó una buena validez convergente. En conjunto, los primeros resultados derivados del proceso de adaptación de la MARS-12 a la lengua y cultura española permiten afirmar que posee adecuadas propiedades psicométricas. En los próximos meses se pretende ampliar la muestra y aportar nuevas evidencias de validez de la escala.

**¿Cómo controlo el efecto de la práctica en mis resultados? El índice de cambio fiable como método para interpretar el cambio entre evaluaciones.** Belén Ruiz-Fernández<sup>1</sup>

<sup>1</sup>Hogrefe TEA Ediciones

En el ámbito clínico las evaluaciones repetidas son muy habituales para comprobar el efecto de las intervenciones o para valorar el curso de las dificultades identificadas (p. ej., en los trastornos del neurodesarrollo, demencias...). Al comparar las puntuaciones de una misma persona en diferentes momentos es habitual encontrar diferencias –aunque no haya mediado una intervención– y resulta complicado saber cuándo estas son el reflejo de un cambio “real”. La incorporación de métodos estadísticos para analizar estas diferencias ayudará a los clínicos a interpretarlas mejor y establecer cuándo son fiables y clínicamente relevantes. Para poner a prueba diferentes métodos se utilizó la muestra de tipificación del d2-R (n = 2.984), de donde se extrajo aleatoriamente una submuestra (n = 249; edad media = 31,35 años; Dt = 22,02; 53,5 % mujeres) que fue evaluada con el Test de Atención d2-R en dos ocasiones, con un intervalo temporal entre ellas de 10 días, y sin que hubiera mediado ningún tipo de intervención o tratamiento entre ambas. Se aplicaron y compararon diferentes procedimientos propuestos en la literatura para la valoración del cambio entre evaluaciones. Los distintos métodos aplicados mostraron diferentes valores para la clasificación del cambio, siendo más precisos aquellos que contemplan en su cálculo más fuentes de variación (efecto de aprendizaje, fiabilidad de la medida, variabilidad de las puntuaciones pre y post...). La disponibilidad de los valores de cambio fiable de los test y cuestionarios utilizados en las evaluaciones de seguimiento permite valorar mejor el curso de los puntos fuertes y débiles identificados en el perfil de la persona y su respuesta a las intervenciones.

**Imputación múltiple en modelos multinivel. ¿Es correcto utilizar la estrategia inclusiva?** Ricardo Olmos<sup>1</sup>, José Ángel Martínez-Huertas<sup>2</sup>, José Manuel Caperos<sup>3</sup>, Pablo García-Patos<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad Nacional de Educación a Distancia, <sup>3</sup>Universidad Pontificia Comillas

Durante las últimas décadas los métodos de referencia para trabajar ante el problema de la pérdida de datos (bajo mecanismos MAR y MCAR) han sido Máxima Verosimilitud (ML) e Imputación Múltiple (IM). De forma esquemática, IM consiste en generar múltiples copias imputadas de la base de datos original, de forma que se preserven tanto las relaciones que hay entre las variables como la incertidumbre debida a que los valores imputados difieren en las copias generadas. En cada una de esas copias se estima el modelo sustantivo (modelo en el que reside el interés del investigador) combinándose en un único modelo bajo las reglas de Rubin (1987). Una ventaja de la IM es que el modelo de imputación se ajusta antes del modelo sustantivo. Esto permite que el modelo de imputación sea más general que el modelo sustantivo (incorporando, por ejemplo, variables auxiliares que ayuden en el proceso de imputación). Una de las reglas más extendidas en IM ha sido la de la estrategia inclusiva (p.ej., Collins et al., 2001), estrategia que indica que un modelo de imputación debe ser al menos tan general como el modelo sustantivo, primándose la inclusión de múltiples variables o efectos que hagan un modelo de imputación muy general. En este estudio cuestionamos esta práctica en imputación múltiple dentro de los modelos multinivel. A partir de un trabajo exploratorio de simulación, mostramos que cuando se sobreparametriza el modelo de imputación (p.ej., incluyendo más efectos aleatorios de los que realmente son necesarios) la imputación es menos eficiente (los errores típicos aumentan) y se produce falta de consistencia (sesgo recuperando los parámetros del modelo sustantivo). Proponemos que, dentro del marco de los modelos multinivel, hacen falta indicadores de ajuste (como AIC, BIC, G2) que permitan orientar al investigador sobre cuál es el modelo de imputación adecuado.

**Tipologías de redes de apoyo social en personas mayores: aplicación del Análisis de Perfiles Latentes en la ola 8 del SHARE.** Zaira Torres<sup>1</sup>, Amparo Oliver<sup>1</sup>

<sup>1</sup>Universidad de Valencia

La clasificación en perfiles de redes sociales de apoyo permite identificar patrones con distintos lazos significativos y relaciones interpersonales. Estas tipologías se relacionan con variables de salud, bienestar y calidad de vida y su análisis es especialmente relevante en las personas mayores ya que es cuando más aumenta la

necesidad de apoyo social. Investigaciones previas muestran la existencia de cuatro, cinco o seis tipologías, por lo que el número concreto de clases está por determinar. Además, estos estudios emplean cluster análisis para la clasificación, usar en su lugar análisis de perfiles latentes puede mejorar la investigación ya que: permiten usar combinaciones mixtas de variables, tiene criterios objetivos para identificar el número de clases que mejor ajusta y evita resultados sesgados causados por las diferentes varianzas. Por ello, este trabajo pretende conocer el número de clases según su apoyo social en una muestra de 31341 personas con 65 años o más de 26 países Europeos e Israel, pertenecientes a la ola 8 del Survey of Health, Ageing and Retirement in Europe (SHARE). En concreto, se han llevado a cabo análisis de perfiles latentes en Mplus 8.6 donde se han probado seis modelos y (a) se han inspeccionado los datos, (b) evaluado los modelos, (c) su ajuste e interpretabilidad y (d) el patrón de perfiles del modelo retenido. La clasificación en cinco subclases se ha retenido por ser la que mejor ajuste mostró tanto en los índices (BIC: 487796.584, Entrophy: 0.917, LMR p value: .003) como teóricamente. Las clases identificadas fueron red “familiar”, “amigos”, “variada”, “otras personas” y “familia directa”. Esta investigación apoya clasificaciones previas y muestra que el análisis de perfiles latentes es una técnica útil para agrupar con ventajas frente al análisis de clúster que puede seguir usándose en futuros estudios en este campo.

**Estrategias cuantitativas para el estudio de la movilización en eventos de protesta. Aplicación al caso español (2000-2020).** Daniel López-Roche<sup>1</sup>

<sup>1</sup>Universidad Complutense de Madrid

A partir del enfoque del proceso político, en mi contribución propongo un modelo que analice los distintos factores asociados a la movilización en eventos de protesta, aplicándolo a España entre el 2000 y 2020. Para ello, planteo distintos aportes teóricos que puedan mejorar esta perspectiva y profundizo en la operativización de las variables y los análisis de datos y modelos estadísticos aplicables. Concretamente, sugiero interpretar la estructura de oportunidades políticas como una variable moderadora parcial de la relación entre agravios y eventos de protesta. En esto, delimito los elementos concretos de la estructura de oportunidades políticas e intento enfrentar las críticas que este concepto ha recibido, tratando de combinar la parsimonia y la exhaustividad teórica. De esta forma, busco generar una propuesta más replicable y contrastable. En cuanto al análisis de datos, elaboro una estrategia apropiada para datos muy frecuentes en sociología, estudios históricos y de eventos de protesta: diseños de caso único con pocas medidas repetidas y

muchos datos perdidos. Para ello, parto del uso de la regresión stepwise guiado por los supuestos teóricos y modelos de curva de crecimiento, aplicados a la regresión lineal múltiple y a la regresión de Poisson. Asimismo, trabajo con una novedosa base de datos que recopila los eventos de protesta en España y muchas de sus características entre el año 2000 y 2020 elaborada en el marco de los proyectos de investigación DISDEM y ECOPOL. Una de mis principales apuestas en este trabajo de cara a los estudios de acción colectiva y de la contienda política es tomar los eventos de protesta como objeto de investigación y fomentar su estudio desde perspectivas cuantitativas. Estos son muy versátiles para estudiar gran cantidad de fenómenos y están mejor definidos que otros conceptos más consolidados.

### **Clasificación adaptativa en línea mediante la aplicación cdcats.** Miguel A. Sorrel<sup>1</sup>, Pablo Nájera<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

La idea de presentar únicamente unos pocos ítems seleccionados de forma óptima se remonta ya hasta cien años atrás. Y los primeros algoritmos que permitían su implementación adaptativa informatizada cuentan ya con cincuenta años. Sin embargo, a principios del presente siglo los altos costes asociados con la implementación de los primeros tests adaptativos informatizados (TAIs) llevaron a grandes nombres del panorama mundial a plantear serias dudas sobre si todo eran ventajas en cuanto a este formato de medición. Actualmente, han pasado otros 20 años desde esta fecha y el acceso a internet y a la tecnología necesaria para implementar TAIs ha evolucionado dramáticamente (en cuanto a precio, velocidad y capacidad de almacenamiento). El objetivo de este póster es presentar cdcats, una aplicación Shiny que toma código del paquete de {cdcatsR} para permitir la especificación y diseño de tests adaptativos informatizados basados en modelos de diagnóstico cognitivo. En el póster se detallan los métodos estadísticos que subyacen a cdcats y se describe un banco de ítems de Psicometría incorporado a la aplicación que permitirá a las personas asistentes determinar su nivel de conocimientos en Psicometría, a modo de ilustración. El póster permite exponer de manera didáctica cómo hoy en día es posible especificar un banco de ítems y recoger datos de manera online con test adaptativos informatizados sin coste económico asociado, haciendo uso solo de software libre. Se discute cómo este puede ser un buen primer paso para explorar la implementación de tests adaptativos y se comentan posibles vías de desarrollo posterior, de manera que pueda generalizarse el uso de esta metodología.

### **Propiedades Psicométricas de la Versión en Castellano del test de Necesidad de Afecto.** Rafael Gil<sup>1</sup>, Miguel A. Sorrel<sup>1</sup>, Javier Horcajo<sup>1</sup>, Miguel Sorrel<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

La necesidad de afecto (NA) fue propuesta para referirse a la motivación de aproximarse o alejarse de situaciones y estímulos con una alta emocionalidad. Maio y Esses, (2001) propusieron un test que ha permitido investigar sobre este constructo y su relación con otros fenómenos importantes en psicología social. En la presente investigación, tenemos como objetivo adaptar el test y validarlo en población general española. El test fue traducido al castellano por dos personas bilingües y después se realizó un proceso de back traslation para verificar su correcta traducción. Se llevó a cabo un primer estudio en el que se utilizaron medidas de validez convergente y validez discriminante (e.g., búsqueda de sensaciones) y referida a criterio (e.g., preferencia por películas emocionales). La muestra del estudio incluyó un total de 416 participantes, con un rango de edad de entre 18 y 87 años, pertenecientes a la población española general. La distribución según el género es de 180 hombres, 232 mujeres y 4 participantes con otra identidad de género. Respecto de la estructura interna, se realizó un análisis factorial con el objetivo de comprobar empíricamente la estructura conceptual propuesta para este constructo psicológico que diferencia entre dos factores (aproximación y evitación). Adicionalmente, se utilizó un análisis de variables únicas para reducir a una variable los ítems redundantes. Finalmente, se realizó un retest con una muestra de 100 participantes. El test de NA mostró unos buenos resultados en cuanto a la consistencia interna y la estabilidad temporal. Bajo la estructura de dos factores propuesta por los autores originales, los ítems se agrupan en dos factores de aproximación y evitación que correlacionan, congruentemente con la teoría previa, con las medidas de validez convergente, divergente y referida a criterio estudiadas. En definitiva, el test presenta unas cualidades psicométricas adecuadas para utilizarse con población española.

### **Modelos de árbol de procesamiento multinomiales con tiempos de reacción: comparación de procedimientos paramétricos y no paramétricos.** Anahí Gutkin<sup>1</sup>, Manuel Suero<sup>1</sup>, Juan Botella<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Los modelos de árbol de procesamiento multinomiales con tiempos de reacción (MPT-TR) son recientes y carecemos de un protocolo de modelado específico para los diferentes escenarios de distribución de los datos. Los procedimientos existentes difieren en la forma en que modelan los TR: algunos incluyen histogramas y otros asumen distribuciones, designados aquí como



procedimientos no-paramétricos y paramétricos, respectivamente. El objetivo de este trabajo es investigar, desde la simulación de un modelo de doble umbral (2HTM), qué procedimiento es el más apropiado considerando situaciones experimentales con probabilidades de procesos cognitivos extremos, manipulando el número de ensayos y las distribuciones latentes de TR. Los resultados preliminares indican que para elegir un procedimiento el experimentador no sólo debe tener una idea aproximada de la forma en que se distribuyen los TR (e.g.: suelen seguir distribuciones asimétricas positivas con una “fat tail”) sino que también deben hacer asunciones compatibles con los cambios de distribuciones de TR entre ramas (e.g.: las distribuciones de RT de estados de guessing del 2HTM tienen “fatter tails” que aquellos de estados de detección). Como el número de observaciones afectan a los procedimientos paramétricos y no paramétricos de forma diferencial, también se proveerá el número de observaciones necesarios para cada procedimiento y situación experimental.

**Aproximación metodológica al estudio de los procesos superiores de pensamiento en el contexto de los signos de tráfico.** Ana Hernando<sup>1</sup>, Antonio Lucas-Alba<sup>1</sup>, Andrés Sebastián Lombas<sup>1</sup>

<sup>1</sup>Universidad de Zaragoza

Este trabajo muestra una revisión de las principales aproximaciones metodológicas empleadas para abordar la comprensión de los signos de tráfico. La constante innovación en aspectos tecnológicos, la multiplicidad de normativas oficiales que regulan el diseño de la señalización vial o el incremento de la movilidad internacional son algunos de los factores que condicionan la investigación en este campo. Esta compleja realidad se aborda desde diversos ámbitos de conocimiento (psicología, ergonomía, ingeniería) y múltiples enfoques metodológicos. En numerosas ocasiones, siguiendo una fuerte vertiente aplicada, se focaliza en cómo optimizar la comprensión de los signos actuales, sin profundizar en el proceso de razonamiento subyacente. A través de modelos predictivos, se estudian las características sociodemográficas y las preferencias de los conductores en relación a la probabilidad de comprender u obedecer el mensaje exhibido en un signo (Ben-Bassat & Shinar, 2015; Ng & Chan, 2008; Chatterjee et al., 2002). Basándose en metodología cualitativa, Ben-Bassat y Shinar (2006) estudian cómo la aplicación de principios ergonómicos mejora la funcionalidad de los signos (Jamson & Mrozek, 2017; Ben-Bassat, 2019). Utilizando un paradigma experimental (exposición estímulos estáticos y dinámicos, uso de simulador de conducción, registro de movimientos oculares), diversos estudios han estudiado cuestiones técnicas de los signos

(color, formatos; Cristea & Delhomme, 2004) o la superioridad de un tipo de información (texto vs. pictograma; Roca, Insa & Tejero, 2018; Shinar y Vogelzang, 2013). No obstante, el reto de explicar cómo el conductor alcanza una conclusión sobre el significado de un signo ha recibido menos atención (Castro et al, 2008; Vargas et al., 2011). Con el fin de abordar esta cuestión, se considera prioritario asentar las bases de una línea de investigación experimental, estableciendo antecedentes metodológicos y estadísticos sobre los que guiar nuevos estudios.

**Definición empírica de los puntos de corte para considerar efectos de tamaño de pequeño, mediano y grande para la eficacia de la punción seca.** Carlos Giménez-Pérez<sup>1</sup>, José Manuel Caperos<sup>1,2</sup>

<sup>1</sup>Fundación San Juan de Dios, <sup>2</sup>Universidad Pontificia Comillas

La punción seca es una de las técnicas de intervención más extendidas y estudiadas en el ámbito de la Fisioterapia musculoesquelética. En este campo, la interpretación de la magnitud de los efectos encontrados en los ensayos clínicos y metaanálisis suele hacerse en base a los valores propuestos por Cohen. Sin embargo, los tamaños del efecto en un área son dependientes del fenómeno de estudio, del tipo de metodología empleada o de las variables y métodos de medida utilizados. Para poder realizar una interpretación adecuada de los resultados es recomendable evaluar el tamaño de los efectos en cada campo de estudio concreto. Se ha obtenido una estimación del tamaño del efecto (mediante la diferencia de medias tipificada,  $g$ ) de 488 contrastes de hipótesis procedentes de 101 ensayos clínicos evaluando la eficacia de la punción seca. Se calcularon los puntos de corte para considerar un efecto como pequeño, mediano o grande como los cuartiles de la distribución de los tamaños del efecto recopilados, para la muestra total y en función de la variable de resultado analizada. En la muestra completa se encontró un efecto pequeño como  $g = 0,20$ , mediano  $g = 0,50$  y grande  $g = 1$ . Estos valores fueron superiores para los contrastes que evaluaban el dolor, con escala visual analógica ( $g = 0,25$ ;  $g = 0,57$ ;  $g = 1,17$ ) o con otras medidas ( $g = 0,16$ ;  $g = 0,51$ ;  $g = 1,18$ ), que en variables de resultado fisiológicas ( $g = 0,15$ ;  $g = 0,46$ ;  $g = 0,97$ ), funcionales ( $g = 0,18$ ;  $g = 0,37$ ;  $g = 0,79$ ) o psicológicas ( $g = 0,14$ ;  $g = 0,25$ ;  $g = 0,42$ ). Es importante que los investigadores conozcan los distintos factores que pueden estar influyendo en las estimaciones del tamaño del efecto de sus estudios.

**Estimadores del tamaño del efecto en ANOVA en condiciones de no normalidad.** F. Javier García-Castro<sup>1</sup>, Jaime Arnau<sup>2</sup>, Rafael Alarcón<sup>1</sup>, Roser Bono<sup>2</sup>, María J. Blanca<sup>1</sup>

<sup>1</sup>Universidad de Málaga, <sup>2</sup>Universidad de Barcelona

En el análisis de varianza (ANOVA), el tamaño del efecto es un indicador que estima la magnitud de la diferencia entre las medias de los distintos grupos, cuantificando el fenómeno de estudio. A través de simulación Monte Carlo, se ha estudiado el sesgo, la precisión y la exactitud de los principales estimadores del tamaño del efecto del ANOVA unifactorial ( $\eta^2$ ,  $\varepsilon^2$ , y  $\omega^2$ ). Sin embargo, esto se ha realizado principalmente con distribuciones normales. El objetivo del presente trabajo es el de analizar el sesgo, la precisión y la exactitud de los tres principales estimadores del tamaño del efecto del ANOVA a través de una simulación Monte Carlo ante la violación ligera, moderada, y severa de la normalidad. Se ha incluido un tamaño del efecto medio, con una diferencia de medias siguiendo un patrón lineal, y se ha manipulado el tamaño muestral por grupo y la desviación con respecto a la distribución normal (leve, moderada y severa). Los resultados mostraron tendencias similares a los estudios previos en todas las condiciones, siendo  $\eta^2$  el estimador con mayor sesgo en comparación con los otros dos tamaños del efecto. Las características de estos tamaños del efecto también variaron según el tamaño muestral y la distribución de la variable.

**Análisis psicométrico de la Escala de Dificultades en la Regulación Emocional (DERS-28) en jóvenes consumidores de sustancias.** Alba González-Roz<sup>1</sup>, Elena Gervilla<sup>2,3</sup>, Álvaro Postigo<sup>1</sup>

<sup>1</sup>Universidad de Oviedo, <sup>2</sup>Universitat de les Illes Balears, <sup>3</sup>Instituto de Investigación Sanitaria Illes Balears (IDISBa)

La desregulación emocional (DE), que integra tanto los déficits en la activación como en la ineffectividad de ciertas estrategias de regulación de las emociones, se ha vinculado con el consumo de sustancias. La Escala de Dificultades en la Regulación Emocional (DERS-28) es el instrumento de evaluación más utilizado para la evaluación de la DE. En España, no se disponen de estudios psicométricos que permitan garantizar la fiabilidad y validez del DERS-28 en personas consumidoras de sustancias. Este estudio tuvo como objetivo examinar las evidencias de validez de la DERS-28 en jóvenes adultos de la población española. La muestra la conformaron 1.713 [Medad= 19,55(DT = 1,69)] usuarios de sustancias que participaron en un estudio multicéntrico (Asturias, Aragón e Islas Baleares). La estructura factorial de la DERS se examinó empleando análisis factorial exploratorio ( $n = 580$ ) y confirmatorio ( $n = 1.133$ ). La fiabilidad se examinó mediante la Teoría Clásica y de la Respuesta al Ítem (TRI) (modelo de Samejima). La estructura factorial del DERS-28 se compone de 5 dimensiones (interferencia, no aceptación, falta de control, inatención y confusión). Las correlaciones entre los factores oscilaron entre ,255-.655.

El modelo bifactor mostró un buen ajuste, aportando evidencias sobre un factor general de DE y cinco factores específicos (RMSEA= ,065; SRMSR= ,067; NNFI= ,916; CFI= ,928). Tanto el factor general como los específicos mostraron una entidad suficiente, explicando una parte considerable de la varianza (wHG = ,833; wHS = ,352 - ,793), excepto falta de control (wHS = ,137). La fiabilidad de las dimensiones fue excelente ( $w = .820-.963$ ) y el análisis TRI indicó una elevada precisión ( $\theta = -2,5$  to  $+1,5$ ) del test. La DERS-28 es un instrumento fiable y válido para evaluar la DE en jóvenes usuarios de sustancias. Futuros trabajos deberán examinar su validez predictiva en relación con la gravedad de la adicción.

**Análisis Psicométrico de la Versión Española de la Brief Self-Control Scale: Evidencias de su Dimensionalidad, Invarianza y Correlatos con Personalidad y Bienestar.**

Estefania Guerrero<sup>1</sup>, Jorge Torres-Marín<sup>1</sup>, Juana Gómez-Benito<sup>1</sup>, Georgina Guilera<sup>1</sup>, Maite Barrios<sup>1</sup>

<sup>1</sup>Universitat de Barcelona

En 2004, June P. Tangney y colaboradores propusieron y empíricamente validaron la Brief Self-Control Scale (BSCS), una medida breve que permite diferenciar a los individuos en función de sus niveles en autocontrol disposicional. A pesar de la innegable popularidad de este instrumento ( $\geq 7000$  citas en Google Scholar), hoy en día continua vigente un debate sobre qué representación interna explica con mayor grado de precisión la estructura latente de este instrumento. Este trabajo presenta una evaluación psicométrica comprehensiva de la versión española de la BSCS, con especial énfasis en su estructura factorial. Utilizando una muestra de 1558 participantes, y guiados/as por la literatura previa sobre esta medida, testamos una serie de modelos que describen la BSCS como (A) unidimensional, (B) bifactorial (6 modelos que difieren en la naturaleza de los factores de la BSCS y/o la composición de sus ítems) o (C) tetrafactorial. Tras obtener evidencias de que el modelo de dos factores—autodisciplina y control de impulsos: Morean et al., 2014; revisado por Chiesi et al., 2020—presenta un ajuste superior para la versión española de la BSCS, exploramos otras propiedades psicométricas de la escala. Observamos que las puntuaciones de la versión española de la BSCS fueron fiables en términos de consistencia interna. También obtuvimos pruebas de la invarianza (parcial) de la medida con independencia del género y la edad. Por último, obtuvimos evidencias de una adecuada validez de criterio de nuestro instrumento al obtener correlatos coherentes con medidas de personalidad (Big Five) e indicadores de bienestar (satisfacción con la vida y felicidad subjetiva). En definitiva, la versión española de la BSCS parece arrojar garantías psicométricas suficientes como para

recomendar su administración en estudios con muestras hispanoparlantes que quieran estudiar diferencias individuales en autocontrol disposicional.

**Estableciendo criterios externos para la actualización de un meta-análisis.** Santiago Arranz-Orlandi<sup>1</sup>, Juan Botella<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

El cuerpo de investigaciones que nutren la práctica basada en la evidencia es tan amplio que el meta-análisis, como herramienta de síntesis cuantitativa, se ha constituido como referencia a la hora de tomar decisiones. No obstante, según avanza el tiempo, los meta-análisis correctamente realizados se van quedando desactualizados y cabe preguntarse cuándo merece la pena publicar una actualización. Es importante, de forma similar a las reglas secuenciales en los estudios primarios, establecer criterios a priori o externos que nos ayuden a decidir cuándo es conveniente actualizar un meta-análisis, sobre todo cuando sus resultados no son estadísticamente significativos. De lo contrario, se puede caer en la mala praxis de añadir casos (en este caso estudios) basándose en la significación, lo que desemboca en un progresivo aumento de la tasa de error tipo I. El objetivo del presente estudio es desarrollar una herramienta práctica para el metodólogo a la hora de decidir cuándo actualizar un meta-análisis basándose en dos criterios independientes y complementarios: la potencia y la precisión de la estimación combinada. Se ha desarrollado un script para el entorno estadístico R que permite estimar cuántos estudios serían necesarios para alcanzar: (a) un valor determinado de potencia, y/o (b) una determinada reducción de la amplitud del intervalo de confianza de la estimación combinada del tamaño del efecto. Estos cálculos se podrán realizar tanto para modelos de efecto fijo como de efectos aleatorios.

**Invarianza factorial de la Escala de Satisfacción con la Vida (SWLS) en adultos ingleses.** María J. Blanca<sup>1</sup>, Rebecca Bendayan<sup>2</sup>, F. Javier García-Castro<sup>1</sup>

<sup>1</sup>Universidad de Málaga, <sup>2</sup>King's College London

La satisfacción con la vida se ha definido como el componente subjetivo del bienestar. Existen diferentes instrumentos para su evaluación, siendo uno de los más utilizados la Escala de Satisfacción con la Vida (SWLS). Esta escala contiene cinco ítems con siete opciones de respuesta tipo Likert (1-fuertemente en desacuerdo; 7-fuertemente de acuerdo), obteniéndose una puntuación total que oscila entre 5 y 35. Puntuaciones altas son indicativas de un mayor nivel de satisfacción con la vida. En diferentes estudios llevados a cabo en distintas

poblaciones se ha obtenido una estructura de un único factor. Aunque se ha utilizado en múltiples investigaciones dentro del English Longitudinal Study of Ageing (ELSA), con una gran muestra representativa de población inglesa, no se ha estudiado su invarianza respecto al género. El objetivo del presente trabajo es el de analizar esta invarianza del SWLS en los datos recogidos en ELSA. La muestra estaba compuesta por 7790 adultos ingleses, y se ha llevado a cabo un análisis factorial confirmatorio de la estructura de un factor del SWLS y un análisis de invarianza entre hombres y mujeres. El modelo de un factor presenta un ajuste satisfactorio según los índices de ajuste del modelo para la muestra total y para las submuestras de hombres y mujeres, mostrando además una invarianza métrica entre ambos géneros. El análisis confirma la estructura de un factor del SWLS tanto para hombres como para mujeres. Este resultado, además es consistente a los encontrados en otros estudios con diferentes poblaciones.

**ESCOLA: Una Nueva Herramienta de Evaluación del Compromiso Laboral.** Álvaro Postigo<sup>1</sup>, Francisco Prieto-Díez<sup>1</sup>, Marcelino Cuesta<sup>1</sup>, José Muñoz<sup>2</sup>

<sup>1</sup>Universidad de Oviedo, España, <sup>2</sup>Universidad Nebrija, España

Debido a su gran importancia en la predicción del rendimiento laboral, el concepto de compromiso laboral ha adquirido gran relevancia en los últimos años. El compromiso laboral se puede entender como la implicación y participación efectiva de las personas en el trabajo, la cual genera una valoración positiva del propio desempeño y de la actividad laboral realizada. El objetivo de esta investigación fue el desarrollo y análisis psicométrico de una nueva escala para la evaluación del compromiso laboral en población general española. Se empleó una muestra de 599 trabajadores en activo, siendo el 51% trabajadores por cuenta propia. El 53% fueron hombres y la media de edad fue de 44,41 años (DT = 8,78). La nueva escala desarrollada consta de 10 ítems y muestra una estructura esencialmente unidimensional. La fiabilidad de las puntuaciones fue excelente ( $\omega = ,92$ ), y se obtuvieron evidencias de validez en relación con el Clima Organizacional ( $r = ,540$ ), Personalidad Emprendedora ( $r = ,701$ ), Felicidad ( $r = ,674$ ), Reparación Emocional ( $r = ,470$ ), y Estabilidad Emocional ( $r = ,440$ ). La escala desarrollada para la evaluación del compromiso laboral muestra unas buenas propiedades psicométricas, constituyendo una herramienta adecuada para su utilización tanto en investigación como en contextos profesionales aplicados.

**Factores de personalidad relacionados con la omisión de respuestas en exámenes de opción múltiple en**

**estudiantes universitarios.** José Manuel Caperos<sup>1,2</sup>, Ricardo Olmos<sup>3</sup>, Rosa María Ortiz-Gutierrez<sup>4</sup>, Eduardo Estrada<sup>3</sup>

<sup>1</sup>Universidad Pontificia Comillas, <sup>2</sup>Fundación San Juan de Dios,

<sup>3</sup>Universidad Autónoma de Madrid, <sup>4</sup>Universidad Complutense de Madrid

El objetivo de este estudio es evaluar la relación entre los rasgos de personalidad y las omisiones de respuesta en los exámenes de elección múltiple. En una muestra de 201 estudiantes de primer año de las licenciaturas de fisioterapia y enfermería se recogieron datos de las escalas de razonamiento fluido (Raven), neuroticismo y responsabilidad (Neo-FFI), impulsividad (UPPS-P) y ansiedad (STAI estado y rasgo). También se recogieron datos de las omisiones, las respuestas correctas e incorrectas en los exámenes finales de cuatro asignaturas diferentes (Anatomía, Bioquímica, Fisiología y Estadística). Además, en una parte de la muestra (n=90) se forzó a los estudiantes a que tras las omisiones intentaran adivinar cuál era la respuesta correcta. Se encontró un factor latente en la omisión en los distritos exámenes que explicaba el 57% de la varianza. Las puntuaciones en dicho factor se relacionaron positivamente con la ansiedad estado experimentada después de completar la prueba de elección múltiple de Raven, pero no con la ansiedad rasgo ni con el neuroticismo. Además, la omisión se encontró inversamente relacionada con la responsabilidad, pero no con la impulsividad, ni la inteligencia. Cuando se fuerza a responder, el acierto en los ítems omitidos fue del 42,9% de ellos (SD = 27,9%), porcentaje significativamente mayor al esperado por azar (33%). La respuesta de los ítems omitidos hubiera incrementado la calificación media en el examen significativamente, aunque con poca relevancia práctica (0,2 puntos sobre 10, DT = 0,48). Los alumnos con puntuaciones latentes más altas en el factor omisión muestran un mayor incremento en la nota cuando se les fuerza a contestar. Nuestros resultados muestran que la personalidad está relacionada con el rendimiento en los exámenes de opción múltiple, pero con un bajo impacto en las calificaciones finales.

#### **Adaptación de la escala Situational Motivation Scale.**

Pablo Sayans-Jiménez<sup>1,2</sup>, Marta Godoy-Giménez<sup>1,2</sup>, Ángel García-Pérez<sup>1,2</sup>, Stefan Anca<sup>1</sup>, Regina Caparrós-Armand<sup>1</sup>, Fernando Cañadas<sup>1,2</sup>, Ángeles F. Estévez<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería)

La eficacia de las nuevas aplicaciones móviles para la salud depende, inevitablemente, de la motivación de los/as usuarios/as por su uso. En un contexto donde el uso de este tipo de app es creciente, la evaluación de las mismas implica conocer el grado de motivación por su

uso. La Situational Motivation Scale (SIMS) es una escala de 16 ítems que evalúa la motivación situacional para realizar una actividad. Este estudio se centró en realizar una traducción y adaptación cultural al castellano de la misma en el marco de un proyecto enfocado al desarrollo de MedMemoriApp (una aplicación para mejorar, en personas mayores de 60 años polimedizadas, el recuerdo tanto del momento de toma-medimento como de la enfermedad- medicamento). Siguiendo las directrices de la Comisión internacional de test, se realizaron dos traducciones ciegas inglés-castellano de los ítems seguidas de un juicio de expertos para conciliar ambas traducciones y adaptarlas culturalmente. Se modificaron las instrucciones y los ítems para que se ajustasen al objetivo del estudio y a la población. Un estudio piloto con una muestra de 11 participantes mayores de 60 años (población objetivo final) sirvió para recoger posibles incidencias y modificar aquellos aspectos que fuesen necesarios. Tras ello, se realizó una retrotraducción ciega castellano-inglés y un estudio comparativo final de las traducciones y retrotraducción por un tercer traductor profesional. La aplicación de esta escala pretende ofrecer información sobre la motivación por el uso de MedMoriApp. El proceso de entrenamiento con la app y la recogida inicial de datos, que se encuentran actualmente en marcha, ofrecerán resultados preliminares relacionados con las propiedades psicométricas de la escala y de sus ítems.

**Validación de las mediciones de la versión española de la Relationship Power Inventory.** María Alonso-Ferres<sup>1</sup>, Celia Serrano-Montilla<sup>2</sup>, Inmaculada Valor-Segura<sup>1</sup>, Francisca Expósito<sup>1</sup>

<sup>1</sup>Universidad de Granada, <sup>2</sup>Universidad de Córdoba

Las dinámicas de poder en las relaciones de pareja son fundamentales a la hora de negociar conflictos. Es por ello que el presente estudio pretende abarcar el proceso de validación de las mediciones de la "Spanish Adaptation of Relationship Power Inventory" (SARPI; Farrell et al., 2015), basada en una de las teorías más actuales y apoyadas del constructo, a partir de diferentes fuentes de validez con el fin de desarrollar un argumento de validez que apoye las interpretaciones propuestas para las mediciones y sus usos previstos para el ámbito de la investigación. Concretamente, se obtuvieron evidencias basadas en el contenido del test, en la estructura interna y en la relación con otras variables. En el Estudio 1, un total de seis expertos/as realizaron una tarea para valorar la calidad técnica, relevancia y representatividad de los ítems al contenido de la SARPI. En el Estudio 2, se administró la escala a dos muestras diferentes de población adulta. En concreto, en la Muestra 1 (N = 400), de prueba, se exploró la dimensionalidad. En la Muestra 2 (N = 755), muestra de validación, se confirmó la estructura interna de la escala

y se obtuvo evidencia de validez externa e invarianza de la medida. El análisis factorial exploratorio y confirmatorio mostró un buen ajuste para la estructura de cuatro factores (poder personal en el resultado, poder personal durante el proceso, poder de la pareja en el resultado y poder de la pareja durante el proceso). Estas dimensiones fueron invariantes entre hombre y mujeres y se obtuvieron evidencias de validez basadas en la relación con otras variables (dependencia de la pareja, estrategias de resolución de conflictos y bienestar psicológico). La interpretación propuesta para la mediciones de la SARPI son apoyadas para los usos propuestos. Este trabajo ha sido publicado en la revista *Current Psychology*.

**Análisis psicométrico de la IGDS9-SF y de la GDT: Un estudio con el análisis de escalas de Mokken.** Laura Maldonado-Murciano<sup>1</sup>, Halley M. Pontes<sup>2</sup>, Maite Barrios<sup>1</sup>, Juana Gómez-Benito<sup>1</sup>, Georgina Guilera<sup>1</sup>

<sup>1</sup>Universidad de Barcelona, <sup>2</sup>Birkbeck University of London

La Internet Gaming Disorder Scale-Short-Form (IGDS9-SF; Pontes & Griffiths, 2015) y la Gaming Disorder Test (GDT; Pontes et al., 2021) son dos test estandarizados que evalúan el trastorno por uso de videojuegos bajo el marco teórico de la APA y de la OMS, respectivamente. Las correspondientes versiones en español han demostrado adecuadas propiedades psicométricas bajo la teoría clásica de los test y modelos paramétricos de la teoría de respuesta al ítem (Maldonado-Murciano et al., 2020, 2021). El objetivo principal de este estudio es evaluar ambos instrumentos a través del Análisis de Escalas de Mokken (MSA) aplicando el modelo de homogeneidad monótona y el modelo de doble monotonicidad. Participaron en este estudio un total de 605 jugadores de videojuegos (42.31% mujeres, Media(edad) = 23.98, SD(edad) = 9.21). Los resultados indicaron unidimensionalidad, independencia local, monotonicidad latente y ausencia de interacción de las curvas características de los ítems en ambas escalas. En conjunto, la IGDS9-SF y la GDT encajan en los dos modelos del MSA, reafirmando sus excelentes propiedades psicométricas. Este estudio concluye que ambas escalas son psicométricamente aptas para evaluar el uso problemático de los videojuegos.

**Aproximaciones al estudio de la multidimensionalidad.** Karen Martínez<sup>1</sup>, Paula Elosua<sup>1</sup>

<sup>1</sup>Universidad del País Vasco

El estudio de la estructura factorial de un instrumento de evaluación psicológica es una fase ineludible en el proceso de validación de las puntuaciones. Son varias las aproximaciones que permiten evaluar la multidimensionalidad de una escala; a los modelos

factoriales exploratorios y confirmatorios clásicos se unen los modelos bifactoriales y los modelos ESEM. El objetivo de este trabajo es comparar los distintos acercamientos desde una perspectiva aplicada que permita al investigador conocer las ventajas de cada uno de ellos. Para ello analizamos la estructura interna del cuestionario de remisión (EDI3 RF) construido a partir del Eating Disorder Inventory (EDI3). Se trata de una medida breve de autoinforme compuesta por 25 ítems agrupados en 3 escalas (Obsesión por la delgadez, bulimia e insatisfacción corporal). El estudio se llevó a cabo sobre una muestra compuesta por 2.304 mujeres provenientes de población comunitaria que aleatorizamos en varias submuestras para evitar la capitalización del azar. El trabajo muestra las ventajas e inconvenientes relacionados con cada uno de los modelos evaluados, y ofrece pautas prácticas que puedan ayudar al investigador en la selección e interpretación de las distintas soluciones.

**Adaptación española de la subescala de Conocimiento Emocional del Schultz Test of Emotion Processing (STEP).** Ana Isabel Vergara<sup>1</sup>, Natalia Alonso-Alberca<sup>1</sup>

<sup>1</sup>Universidad del País Vasco UPV/EHU

Existe abundante evidencia de la relación entre el conocimiento emocional (CE) y el ajuste psicológico, social y escolar, así como de su vinculación con la conducta adaptativa en la etapa de Educación Primaria. Sin embargo, son escasos los instrumentos adaptados al castellano dirigidos a medir el CE. El Schultz Test of Emotion Processing (STEP; Schultz y cols., 2010) es un instrumento destinado a evaluar el procesamiento de la información emocional en niños/as entre los 7 y 11 años, compuesta por tres subescalas. En este estudio se presenta la adaptación de la primera subescala, dirigida a evaluar el CE, que consta de 25 secuencias de video breves (5-10 sg.) en las que tienen lugar situaciones cotidianas de contenido emocional, y en las que el niño/a debe indicar la emoción que siente el/la protagonista. El hecho de utilizar estímulos dinámicos otorga validez ecológica a la prueba, solventando una de las principales limitaciones existentes en este ámbito de estudio. La adaptación al castellano se llevó a cabo mediante una traducción directa-inversa de los diálogos de las diferentes secuencias, y utilizando actores/actrices para el doblaje de los videos originales. La muestra estuvo compuesta por 255 niños y niñas (47,8% niñas), con edades entre los 8-11 años. El análisis factorial confirmatorio de la versión adaptada mostró una estructura unidimensional (CFI = 0,91, TLI = 0,90) con adecuada consistencia interna (alfa de Cronbach = 0,76). Con objeto de aportar evidencias de validez de relación con otras variables, se realizó un modelo de ecuaciones estructurales analizando la influencia del sexo, la edad y la habilidad verbal en el CE, así como la influencia de éste

en variables conductuales. Los resultados obtenidos muestran la influencia de la habilidad verbal en el CE, así como que un mayor CE se asocia a menores problemas internalizantes, externalizantes y escolares.

## Sesión Premio AEMCCO Jóvenes Investigadores 2022

**Análisis de la potencia estadística de los estudios sobre la eficacia de la punción seca.** Carlos Giménez-Pérez<sup>1</sup>, José Manuel Caperos<sup>1,2</sup>

<sup>1</sup>Fundación San Juan de Dios, <sup>2</sup>Universidad Pontificia Comillas

Uno de los problemas más importantes señalados en los últimos años en la investigación en las ciencias del comportamiento y de la salud es la baja potencia estadística de muchos de los contrastes de hipótesis realizados. Este problema se relaciona con una mayor tasa de errores tipo II, una menor probabilidad de que un resultado significativo refleje un efecto real y una sobreestimación del tamaño del efecto en los metaanálisis que incluyan dichos resultados. En este trabajo se ha calculado la potencia estadística de los ensayos clínicos sobre la eficacia de la punción seca, una técnica de intervención de la Fisioterapia. Se analizaron 488 contrastes procedentes de 101 estudios. Para los cálculos de potencia la hipótesis alternativa se situó en: 1) la estimación de tamaño del efecto de un metaanálisis que incluyera el contraste considerado; 2) los valores operativizados para definir un efecto pequeño, mediano y grande (calculados a partir de los cuartiles de la distribución de los tamaños del efecto recopilados). Para estas condiciones se calculó la potencia de cada contraste y la media de potencia de cada estudio. La mediana de potencia estadística de los estudios fue de 0,42 (Amplitud intercuartil, AIQ = 0,43), solo un 14,9% presentan una potencia mayor de 0,8. Considerando los contrastes, la potencia mediana fue de 0,31 (AIQ = 0,670) y solo el 25,0% fue mayor de 0,8. Teniendo en cuenta los valores para efectos pequeños ( $g = 0,2$ ), medios ( $g = 0,5$ ) y grandes ( $g = 1$ ), la potencia mediana de los estudios fue de 0,10, 0,38 y 0,92, respectivamente. La baja potencia estadística encontrada en los ensayos clínicos sobre la eficacia de la punción seca amenaza la validez de la conclusión estadística de estos estudios y apunta a la necesidad de mejorar las prácticas de investigación a este respecto.

**Bifactor: Un paquete para el análisis factorial exploratorio de estructuras jerárquicas.** Marcos Jiménez<sup>1</sup>, Francisco Abad<sup>1</sup>, Eduardo García-Garzón<sup>2</sup>, Luis Eduardo-Garrido<sup>3</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>Universidad Camilo José Cela, <sup>3</sup>Pontificia Universidad Católica Madre y Maestra

Actualmente, se reconoce que los errores de especificación en un análisis factorial confirmatorio pueden conducir a serios problemas de estimación de los parámetros e interpretación de los resultados. Por esta razón, los modelos de ecuaciones estructurales exploratorios están experimentando un auge de popularidad. Sin embargo, los programas estadísticos actuales contienen un limitado número y tipos de criterios de rotación y usan algoritmos de optimización que convergen lentamente al estimar grandes estructuras factoriales. En consecuencia, nuestro grupo de investigación ha desarrollado un nuevo paquete estadístico para el lenguaje R, denominado bifactor, con la intención de superar estas limitaciones. Dicho paquete no solo ofrece nuevos algoritmos y posibilidades de rotación para investigar estructuras jerárquicas complejas como las comúnmente encontradas en inteligencia, personalidad y psicopatología, sino que también garantiza una alta velocidad de estimación. Esta eficiencia computacional procede de varias fuentes: programación de los algoritmos en lenguaje C++, implementación de rutinas de optimización newtonianos con mayor velocidad de convergencia y paralelización de tareas. Además, el paquete bifactor también permite seleccionar entre distintos métodos de extracción, computar índices de ajuste y estimar errores típicos, lo que lo convierte en una herramienta autosuficiente para estimar, evaluar e interpretar modelos factoriales exploratorios. En esta charla se demostrará la utilidad práctica del paquete bifactor para estimar estructuras jerárquicas complejas reanalizando casos empíricos.

**Broad Autism Phenotype- International Test (BAP-IT): Propiedades psicométricas mediante modelos de redes empíricas.** Marta Godoy-Giménez<sup>1,2</sup>, Ángel Pérez-García<sup>1</sup>, Antonio González-Rodríguez<sup>1,2</sup>, Fernando Cañadas<sup>1,2</sup>, Angeles F. Estévez<sup>1,2</sup>, Pablo Sayans-Jiménez<sup>1,2</sup>

<sup>1</sup>Universidad de Almería, <sup>2</sup>Centro de Investigación en Salud de la Universidad de Almería

El fenotipo ampliado del autismo (BAP) comprende comportamientos relacionados con el autismo que se expresan a niveles subclínicos y se distribuyen entre familiares de personas diagnosticadas con autismo y la población general. A pesar de su extenso estudio, los test actualmente disponibles para evaluarlo muestran severos problemas psicométricos (operacionalizaciones desfasadas; evidencias desfavorables relacionadas con su contenido, estructura interna y estabilidad en sus adaptaciones). Este trabajo presenta el BAP-IT, construido simultáneamente en castellano e inglés, y sus propiedades psicométricas comparadas en dos muestras, española e inglesa ( $n = 970, 533$ ). La estabilidad de estas propiedades se replicará en otra muestra española y se

explorarán evidencias de validez adicionales en relación con otras variables. En el estudio 1 se analizan, por muestra, las comunalidades y la dependencia local de los ítems (análisis de varianza única; EBIC<sub>g</sub> y correlaciones de Spearman). Los resultados exponen dos comunidades (alineadas con ambos dominios del TEA según el DSM-5) estables y un par de ítems (versión española) con  $wTO = .20$  (comparten parte del enunciado). No se considera colapsar ambos ítems debido a su diferencia en contenido específico. Seguidamente, se estudiaron las diferencias entre las redes de ambas muestras usando moderación mediante modelos gráficos mixtos (mgm; EBIC,  $\gamma = 0.25$ ), descartando diferencias en las conexiones entre los ítems en función de la versión. En el estudio 2 se replicaron los análisis del estudio 1 en una nueva muestra española ( $n = 460$ ). Se confirma la estabilidad de las dos comunidades, se descarta la presencia de dependencia local y se estiman las puntuaciones de red para las dos dimensiones. Seguidamente, se estima una nueva red empleando dichas puntuaciones y las observadas de las subescalas del BAPQ, SRS-II y RCS. Se muestran evidencias favorables para el BAP-IT y permiten compararlas con las del BAPQ.

**Diseños longitudinales acelerados para el estudio del desarrollo en sistemas bivariados.** Pablo Fernández-Cáncer<sup>1</sup>, Eduardo Estrada<sup>1</sup>

<sup>1</sup>Universidad Autónoma de Madrid

Objetivo. Recoger datos psicológicos en investigación longitudinal puede ser una tarea complicada, especialmente cuando el cambio que interesa estudiar se produce en escalas temporales de años o décadas. Cuando los individuos son evaluados repetidamente durante muchos años, aparecen distintos problemas metodológicos, como efectos de aprendizaje, abandono y sesgos de selección resultantes de la pérdida de participantes. Los diseños longitudinales acelerados (accelerated longitudinal designs, ALD), mediante la pérdida de datos planeada, permiten estudiar procesos que abarcan muchos años en un marco temporal mucho más reducido. Los ALDs se han utilizado principalmente en el campo del desarrollo cognitivo, donde el foco suele estar en identificar las secuencias temporales entre variables (e.g., ¿Son los cambios en la estructura cortical los que preceden a los cambios comportamentales, o viceversa? ¿Son estas influencias bidireccionales?). No obstante, ningún estudio ha evaluado la capacidad de los ALDs para recuperar dichas secuencias temporales entre dos procesos latentes. En este estudio, sugerimos una serie de diseños acelerados con distintos patrones de pérdida de datos y los ponemos a prueba utilizando un modelo de puntuación de cambio latente bivariado

(bivariate latent change score, BLCS) en tiempo continuo. En concreto, nos centraremos en la recuperación de los parámetros de emparejamiento o coupling, que informan sobre las interrelaciones y secuencias temporales entre procesos. Método. Hemos llevado a cabo un estudio de Monte Carlo para evaluar la capacidad de los modelos BLCS en tiempo continuo en combinación con los ALDs para recuperar trayectorias de procesos de desarrollo. Para ello, se han generado datos de dos procesos interrelacionados bajo distintas condiciones de tamaño muestral y tipo de ALD. Resultados y discusión. Discutimos los hallazgos más relevantes, ofrecemos recomendaciones sobre cómo implementar ALDs, y animamos a los investigadores a adoptar modelos BLCS en tiempo continuo para analizar datos de ALDs.

**Invarianza factorial aproximada: Evaluando la precisión y sensibilidad de la optimización Alignment en condiciones de invarianza parcial con múltiples grupos.**

Ricardo Rey-Sáez<sup>1,2</sup>, Manuel Martín-Fernández<sup>2</sup>

<sup>1</sup>Universidad Complutense de Madrid, <sup>2</sup>Universidad Autónoma de Madrid

La invarianza factorial (IF) es un requisito necesario para garantizar la equivalencia de la medida entre distintos grupos, permitiendo a los investigadores realizar comparaciones entre las medias latentes sin incurrir en un grave problema de validez. La investigación en IF desde los modelos de ecuaciones estructurales se ha centrado en mayor medida en la IF con variables continuas, restando protagonismo a las variables categóricas, donde es necesario tener en cuenta consideraciones extra para establecer los distintos niveles de IF. A medida que aumenta el número de grupos es fácil que no podamos asumir IF, siendo posible estudiar la equivalencia de la medida de forma aproximada. La optimización alignment (Asparouhov y Muthén, 2014) permite estudiar qué grupos son aproximadamente invariantes entre sí al mismo tiempo que posibilita realizar comparaciones entre las medias latentes de los distintos grupos sin asumir IF, representando una potencial alternativa al uso de los índices de modificación para establecer invarianza parcial. El objetivo del presente estudio es comparar el funcionamiento del AFC-MG y la optimización alignment en datos dicotómicos bajo distintos escenarios de simulación, manipulando asimetría, tamaño muestral, número de grupos, invarianza y diferencia en las medias latentes. Los resultados muestran el buen desempeño de Alignment a la hora de detectar parámetros y grupos no-invariantes en condiciones de invarianza parcial con ítems dicotómicos simétricos y pronunciadamente asimétricos,

siendo una prometedora alternativa al AFC-MG cuando no podemos sostener la IF de forma exacta.

### **¿Qué nos dicen los valores perdidos respecto del rendimiento académico futuro de los estudiantes?**

Susana Sanz<sup>1</sup>, Esther Ulitzsch<sup>2</sup>, Carmen García<sup>1</sup>, Ricardo Olmos<sup>1</sup>, Steffi Pohl<sup>3</sup>

<sup>1</sup>Universidad Autónoma de Madrid, <sup>2</sup>IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, <sup>3</sup>Freie Universität Berlin

Medir de forma precisa los conocimientos de las personas en ítems de opción múltiple es una de las principales preocupaciones en el estudio de la evaluación académica. Sin embargo, esta medición no solo refleja el conocimiento adquirido de los examinados, pues su nivel de habilidad puede verse afectado por su comportamiento a la hora de contestar (test-taking behaviour), provocando que la evaluación sea menos justa. En este contexto, el estudio de los valores perdidos es especialmente relevante, no solo para mejorar las estimaciones de los parámetros sustantivos con respecto a los ítems y los examinados, sino también porque pueden proporcionarnos información interesante respecto del comportamiento de los estudiantes ante los ítems de opción múltiple. Se ha enfocado esta cuestión intentando encontrar qué variables pueden afectar a la omisión o a no alcanzar ciertos ítems en un test. Sin embargo, también es relevante analizar si los valores perdidos pueden ser predictores del rendimiento académico futuro de los estudiantes. Para este objetivo se tienen en cuenta variables como el nivel educativo que alcanzan, sus aspiraciones laborales o su auto-regulación. Se utilizarán datos del National Educational Panel Study (NEPS), pertenecientes a estudiantes alemanes, y que proporcionan información longitudinal sobre los procesos educativos y el desarrollo de competencias. Se aplicarán modelos sobre omisiones, como el de Holman y Glas (2005), donde se parte del Modelo de Crédito Parcial o del Modelo de Crédito Parcial Generalizado para estimar conjuntamente el nivel de rasgo de los estudiantes y un parámetro de tendencia a omitir, o el de Rose et al. (2010) donde la omisión se considera una variable latente que habría que incluir en el modelo de medida. Se espera encontrar que tanto las omisiones como los ítems no alcanzados ayudan a explicar su rendimiento académico posterior.