



XVI Congreso de Metodología de las Ciencias Sociales y de la Salud

Universidad Autónoma de Madrid
8-10 de julio de 2019

LIBRO DE RESÚMENES



Índice

Simposios

Diferentes aproximaciones metodológicas al estudio e intervención sobre la violencia contra la mujer	1
Calidad metodológica y tamaño de efecto de programas de intervención	2
Modelos de variable latente discreta: alternativas para la medición en presencia de grupos heterogéneos . . .	4
Metodología observacional como 'mixed method' en sí misma (I)	6
Perfiles profesionales en metodología	7
Evaluación de la calidad métrica de instrumentos psicométricos: procedimientos y recomendaciones	8
Meta-análisis: avances estadísticos	9
Metodología observacional como <i>mixed method</i> en sí misma (II)	11
Estrategias de integración de evidencias de validez para el desarrollo de instrumentos de funcionamiento . .	13
Modelos longitudinales con variables latentes: aplicaciones en diferentes etapas del ciclo vital	15
Metodología mixta de investigación: desafíos, propuestas y beneficios de la integración de métodos cuantitativos y cualitativos	16
Aplicaciones metodológicas en la investigación sobre altas capacidades intelectuales	18
Metodología cualitativa en la investigación de altas capacidades	19
Meta-análisis: avances metodológicos	20
Retrato barroco. De la técnica de análisis más utilizada al problema de la pérdida de datos	22
Aplicaciones de la metodología observacional	23
Meta-análisis de generalización de la fiabilidad	25
Estimación y generalizabilidad en diseños observacionales	26
Soluciones metodológicas para las evaluaciones educativas de alto impacto	27
Construcción sistemática de ítems para instrumentos de medida	29
Modelos y técnicas para la estimación de sistemas complejos de conectividad cerebral con señal FMRI	31
Metodología observacional: aplicaciones	32
Meta-investigación: crisis metodológica y promoción de mejores prácticas científicas	34

Comunicaciones orales

Validación de escalas	36
Metodologías	38
Validación de Escalas: Aplicaciones	40
Validación de Escalas: procedimientos	41
Aplicaciones de la Metodología I	43
Aplicaciones de la Metodología II	45
Modelos psicométricos	47
Análisis de datos	49
Metodologías	51
SESIÓN DE PÓSTERES 1	52
SESIÓN DE PÓSTERES 2	59
SESIÓN PREMIO AEMCCO 2019	66

SIMPOSIO

Diferentes aproximaciones metodológicas al estudio e intervención sobre la violencia contra la mujer

Coordinan: Luis Manuel Lozano Fernández y José Luis Padilla García

Resumen. la violencia contra la mujer es un problema social y de salud que afecta a una de cada tres mujeres en el mundo siendo, por tanto, la violación de derechos humanos de mayor prevalencia. En España, entre el 1.9% y el 17.6% de las mujeres informaron haber sufrido violencia física por parte de su pareja o expareja en la Macroencuesta de Violencia contra la Mujer. La investigación, el desarrollo y evaluación de programas de prevención e intervención, plantea problemas metodológicos relacionados con la elaboración de instrumentos de medición, la evaluación de sus propiedades métricas y el análisis de datos desde diferentes aproximaciones. El objetivo del simposio es presentar los problemas metodológicos que plantean la investigación, evaluación e intervención sobre la violencia contra la mujer. Las comunicaciones que agrupa el simposio exponen los retos para identificar los efectos contextuales en las actitudes de justificación de la violencia contra la mujer utilizando modelos multinivel, la medición de la deseabilidad social en la evaluación de la agresión psicológica, la aplicación de modelos de regresión penalizada para evaluar el riesgo de reincidencia de agresores, el análisis de la equivalencia en estudios transculturales sobre violencia contra la mujer, y los desafíos de recurrir a métodos mixtos para la elaboración de instrumentos para evaluar las actitudes hacia la violencia contra la mujer de los agentes y miembros de las fuerzas y cuerpos de seguridad.

Aproximación multinivel al análisis de las actitudes de justificación hacia la violencia de género. Celia Serrano-Montilla¹, Luis Manuel Lozano¹ y José Luis Padilla¹

¹Universidad de Granada

En sociedades donde se mantienen actitudes que justifican la violencia de género la investigación señala una mayor prevalencia de este tipo de violencia, así como una menor reacción de las víctimas y de la sociedad. El objetivo del presente estudio es extender el conocimiento sobre los mecanismos que explican la justificación de la violencia de género. Para ello proponemos tres objetivos: (1) determinar el papel de variables individuales (edad, sexo, nivel educativo, estado civil y percepción del estatus social, sexismo, valores de autotranscendencia y conservación) y macro-sociales (desigualdad de género del país), (2) observar si las relaciones encontradas cambian en función del país y (3) determinar si dicho cambio es explicado por la desigualdad de género. Dado que las respuestas de los participantes del mismo país podrían correlacionar entre ellas más fuertemente debido a que viven bajo contextos similares, empleamos un enfoque multinivel. Así, los participantes (nivel 1) son agrupados en países de pertenencia (nivel 2). Se utilizaron datos de la Encuesta Mundial de Valores (2010 a 2014), un total de 81.516 participantes de 54 países. Se ejecutaron regresiones jerárquicas por pasos, así como, diferentes modelos multinivel (nulo, con pendientes e interceptos fijos y aleatorios). Se encontraron efectos fijos de las diferentes variables individuales y la desigualdad de género en la justificación de la violencia de género a través de los diferentes países. La relación

entre sexismo, valores de autotranscendencia y justificación cambia en función del país aunque dicho cambio no fue explicado por la desigualdad del género.

¿Estamos midiendo bien la deseabilidad social al estudiar la violencia contra la pareja? M. Carmen Navarro-González¹ y José Luis Padilla¹

¹Universidad de Granada

El hecho de poder medir y controlar la deseabilidad social (DS) en cuestionarios sobre comportamientos de violencia contra la pareja (VCP) podría ser de gran ayuda a la hora de detectar y prevenir casos de agresión psicológica. Sin embargo, existen evidencias en la literatura científica que nos hacen sospechar que las escalas más usadas para medir la DS no funcionan bien cuando se trata de estudiar constructos más comprometedores, como el de la VCP, ya que las personas podrían anticipar un riesgo mayor al reconocer este tipo de comportamientos. En este estudio proponemos evaluar la capacidad de la Escala de Deseabilidad Social de Marlowe- Crowne (MC-SDS) de detectar la DS en contextos de VCP. Para ello, se establecieron dos grupos de DS -alta y baja- según la puntuación obtenida en la MC-SDS por los participantes y se compararon las puntuaciones de ambos grupos en la escala de agresión psicológica de The Revised Conflict Tactics Scales (CTS-2) y en una adaptación del General Ecologic Behavior (GEB) para medir comportamiento ecológico, variable que se espera que sea menos comprometedora. Así, si la MC-SDS no discrimina bien ambos grupos en el caso de la VCP, el grupo de DS alta contestará de manera más socialmente deseable tanto en el GEB como en el CTS-2, mientras que el grupo de DS baja no contestará de manera socialmente deseable en el GEB, pero sí en el CTS-2, por tratarse de un contexto altamente comprometedor. Dicha comparación también se relaciona con los juicios sobre la deseabilidad social de los ítems de la CTS-2. Finalmente, se discuten posibles modificaciones en la forma tradicional de evaluar la DS.

Aplicando métodos de regresión penalizada para estudiar el riesgo de reincidencia en agresores de violencia de género. Manuel Martín-Fernández¹, Marisol Lila¹ y Enrique Gracia¹

¹Universidad de Valencia

En la regresión de Cox, cuando el número de variables predictoras es muy elevado o se espera un alto grado de comorbilidad entre ellas, el orden en el que se introduzcan estas variables puede afectar a los resultados. Esto puede alterar las variables que componen el modelo final, llegando incluso a soluciones contradictorias. Los modelos de regresión penalizada son una de las soluciones que más frecuentemente se utilizan en *machine learning* al tratar con estructuras de datos similares. En concreto, el procedimiento *adaptive least absolute shrinkage and selection operator (a-lasso)* es una de las estrategias más extendidas que permite seleccionar variables y penalizar a aquellas que covarían en exceso en un solo paso. En este estudio se presenta la aplicación de este procedimiento a una muestra de 393 hombres condenados por violencia de género, todos ellos participantes del programa de intervención Contexto, dirigido a rehabilitar agresores de este tipo de violencia. Se utilizó un total de 91 variables predictoras para modelar el riesgo de reincidencia a lo largo del tiempo del programa de intervención. Mediante el procedimiento *a-lasso* se obtuvo

un modelo consistente en seis variables: nacionalidad del participante, violencia sufrida en la infancia, la evaluación de riesgo de las terapeutas del programa, la ira como rasgo, el número de eventos estresantes en la vida de los participantes, y si han concluido o abandonado el tratamiento del programa de intervención. Se obtuvo por último una curva ROC para poner a prueba el ajuste del modelo, siendo el área bajo la curva 0,808.

Comparando las medias latentes en violencia de género entre España y Suecia. Manuel Martín-Fernández¹, Marisol Lila¹ y Enrique Gracia¹

¹Universidad de Valencia

La agencia de la Unión Europea para los derechos fundamentales llevó a cabo una encuesta en 2012 en la que se muestra la prevalencia de violencia física y sexual contra la mujer en los diferentes países de la UE. Esta prevalencia es calculada mediante el uso de dos escalas. Sin embargo, no se ha realizado hasta el momento ningún estudio sobre las propiedades psicométricas ni la invarianza factorial de ambas medidas. Esto hace imposible hacer comparaciones entre países, ya que no se puede descartar la posibilidad de que las diferencias en la prevalencia de violencia género se deban a un sesgo en las medidas o a diferentes interpretaciones de los ítems. En este estudio se comparó la violencia física y sexual contra las mujeres en España ($n = 1447$) y Suecia ($n = 1483$). Para ello se estudió la estructura factorial de las dos escalas de la encuesta y se encontró que un modelo de dos factores (i.e., violencia física y violencia sexual) ajustaba mejor en ambos países. Para evaluar la invarianza factorial se llevó a cabo una serie de análisis factoriales confirmatorios multi-grupo. Asimismo, se realizó un análisis DIF para datos de respuesta graduada. Los resultados apoyaron el nivel de invarianza escalar en ambas escalas entre España y Suecia, lo que permitió a su vez estimar las medias latentes en ambos factores. Se encontró que Suecia presenta mayores niveles de violencia física y sexual contra las mujeres, con tamaños del efecto pronunciados, lo que apoya la existencia de la "paradoja nórdica".

Adaptación de un instrumento para la evaluación de las actitudes hacia la intervención en casos de violencia contra la mujer mediante grupos focales. Celia Serrano-Montilla¹, Luis Manuel Lozano¹ y José Luis Padilla¹

¹Universidad de Granada

El estudio de las actitudes mantenidas por las Fuerzas y Cuerpos de Seguridad hacia la intervención en casos de violencia de género requiere la elaboración de instrumentos de evaluación adecuados al contexto cultural y normativo de los diferentes países. Para evaluar estas actitudes hay varias escalas, siendo la más empleada la de Chu y Sun (2014). Esta comprende cuatro dimensiones que hacen referencia a actitudes reactivas (mínima implicación policial y tolerancia hacia la violencia de género), y proactivas (importancia de la labor policial y pro-arresto). El objetivo del estudio fue mejorar el instrumento de medida de las actitudes policiales hacia intervención en violencia de género complementando la definición semántica original del constructo y generando nuevos ítems culturalmente relevantes en España. Los objetivos del estudio son: (1) obtener evidencias sobre la relevancia cultural de las dimensiones e indicadores de la definición semán-

tica original; y (2) ampliar el rango de actitudes posibles de los agentes hacia la intervención en violencia de género para incorporarlas al instrumento adaptado. Se empleó una metodología cualitativa de grupo focales. Un total de 8 grupos focales fueron llevados a cabo con agentes de los tres Cuerpos (Policía Local, Nacional y Guardia civil). El objetivo de la presentación es mostrar las aportaciones de la metodología cualitativa en la adaptación de instrumentos de evaluación de las actitudes hacia la intervención en casos de violencia contra la mujer. También señalaremos las fases en el desarrollo de instrumentos de evaluación en las que se pueden utilizar las evidencias cualitativas aportadas por los grupos focales.

SIMPOSIO

Calidad metodológica y tamaño de efecto de programas de intervención

Coordinan: Salvador Chacón Moscoso y Susana Sanduete Chaves

Resumen. En la práctica habitual, se observa una gran debilidad metodológica en la evaluación de programas de intervención de los diferentes ámbitos de aplicación, al no explicitarse todos los aspectos fundamentales de cómo se llevó a cabo la intervención o de si solventó el problema de la manera más eficaz. Esto dificulta la replicabilidad de las intervenciones exitosas y la acumulación integrada del conocimiento y, por tanto, el crecimiento y avance de la ciencia y de su aplicabilidad a la tecnología. En este simposio exponemos los avances de nuestro 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 a nivel nacional e internacional. Concretamente, se van a presentar seis comunicaciones: la primera comunicación muestra un planteamiento integrador acerca del modo de conectar diseño, medición y análisis desde el punto de vista de los *mixed methods*; la segunda comunicación propone un *checklist* para medir la calidad metodológica de estudios observacionales a partir de un estudio de validez de contenido y fiabilidad intercodificadores; la tercera comunicación estudia la estructura factorial del *Profile of Mood States* en pacientes con dolor crónico; la cuarta, presenta los resultados de un programa psicológico para incrementar la capacitación en padres de bebés prematuros; la quinta presenta la validación de un instrumento de medida de la violencia en parejas homosexuales en Chile; finalmente, la sexta comunicación estudia la eficacia de los programas de formación continua para trabajadores de la salud mental.

Una visión práctica sobre complementariedad de la calidad metodológica en diseños de evaluación de programas desde el punto de vista de los mixed methods. Salvador Chacón Moscoso^{1,2}, M. Teresa Anguera³, Susana Sanduete Chaves¹, José A. Lozano Lozano², Mariona Portell⁴ y José L. Losada³

¹Universidad de Sevilla, ²Universidad Autónoma de Chile, ³Universidad de Barcelona, ⁴Universidad Autónoma de Barcelona

Las decisiones sobre diseño, implementación y evaluación en los programas de intervención deben basarse en evidencias con calidad metodológica. Dependiendo del contexto de intervención, se

podrían aplicar diferentes tipos de metodologías. En la actualidad, aún existe la visión dicotómica que confronta los estudios experimentales/cuasi-experimentales frente a los no experimentales/basados en la metodología observacional. Esto dificulta la práctica de evaluadores y planificadores en la evaluación empírica de programas, un ámbito en el que la distinción entre los tipos de estudio cambia continuamente y es menos clara. Propondremos cómo conectar el diseño, la medición y el análisis de datos en programas de intervención basados en una perspectiva de métodos mixtos. Tomando como base el marco de validez clásico de los estudios experimentales/cuasi experimentales, y tras una revisión de la literatura sobre calidad metodológica en estudios primarios en revisiones sistemáticas de estudios experimentales/cuasi experimentales y de estudios basados en la metodología observacional, el presente estudio pretende: (1) especificar los elementos de diseño relevantes que se deben tener en cuenta para mejorar la validez y la generalización en la práctica de evaluación de programas; y (2) desde el punto de vista de los mixed methods, especificar el grado de correspondencia/complementariedad entre estos elementos en las dimensiones estructurales de un diseño de evaluación desde una perspectiva de complementariedad.

Escala de calidad metodológica para estudios basados en metodología observacional. Salvador Chacón-Moscoso^{1,2}, M. Teresa Anguera³, Susana Sanduete-Chaves¹, José A. Lozano-Lozano², José L. Losada³, Mariona Portell⁴

¹Universidad de Sevilla, ²Universidad Autónoma de Chile, ³Universidad de Barcelona, ⁴Universidad Autónoma de Barcelona

Antecedentes: no existen escalas referidas a los elementos mínimos que garanticen la calidad metodológica en estudios observacionales. En consecuencia, los profesionales no expertos en metodología observacional no disponen de una guía básica que oriente su práctica en este tipo de estudios. Se desarrolló una escala para medir los elementos mínimos de calidad metodológica que deben considerar los estudios observacionales aportando evidencias de su validez basada en el contenido de la prueba y fiabilidad. Método: 54 jueces, con al menos un año de experiencia en metodología observacional y estudios observacionales, evaluaron los ítems de la escala elaborada respecto a su relevancia, utilidad y viabilidad. Se seleccionaron aquellos ítems que obtuvieron al menos 0,5 en los índices de Osterlind en los tres aspectos evaluados. Dos codificadores los aplicaron a una selección aleatoria de artículos que utilizaron metodología observacional en fútbol y se estudió la fiabilidad intercodificadores mediante coeficientes kappa (κ) de Cohen. Resultados: la escala resultante constó de 16 elementos agrupados en 11 dimensiones, con coeficientes de fiabilidad adecuados. Conclusiones: se desarrolló un instrumento útil dirigido a profesionales no expertos en metodología observacional, para potenciar la calidad metodológica de los estudios observacionales en el ámbito de la intervención.

Estructura factorial del Profile of Mood States en pacientes con dolor crónico. Susana Sanduete Chaves¹, Francisco Javier Cano García¹, Salvador Chacón Moscoso^{1,2}, Celia López Jiménez³

¹ Universidad de Sevilla, ²Universidad Autónoma de Chile, ³Centro de Salud de Aznalcóllar, Servicio Andaluz de Salud

El dolor crónico es un problema de salud pública que requiere abordaje multidisciplinar. La *Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials* (IMMPACT), compuesta por académicos, agencias gubernamentales e industria farmacéutica, se gestó para mejorar la calidad de los ensayos clínicos a través de la mejora en su calidad metodológica. Desde el comienzo, IMMPACT propuso, como principal medida del área emocional en el dolor, el Profile of Mood States (POMS). Este instrumento es una lista de 65 adjetivos que miden siete estados de ánimo (tensión, depresión, enfado, vigor, fatiga, confusión y simpatía). Diseñado originalmente para pacientes psiquiátricos, ha sido validado también en población general, convirtiéndose incluso en un instrumento popular en el ámbito deportivo. Sin embargo, se conocen muy pocos estudios sobre dolor crónico donde se haya utilizado, y ninguno de ellos que haya testado su estructura factorial. Este es, pues, el objetivo del presente estudio. 588 pacientes con dolor crónico cumplieron el cuestionario. Se realizó el análisis factorial confirmatorio de la estructura original con los siete factores, y de la misma estructura tras eliminar el factor de simpatía y sus ítems, dado que el autor del instrumento terminó por excluirlos por razones psicométricas. No se encontró un ajuste apropiado en ninguno de los casos. Por ello, se propone una estructura alternativa del instrumento para la medición del estado emocional en pacientes con dolor crónico.

Resiliencia, estrés, depresión y satisfacción con un programa de apoyo a padres y madres de grandes prematuros: evolución desde el nacimiento hasta los 12 meses. Eva M. Padilla Muñoz¹, Merche Barbancho Morant², Susana Sanduete-Chaves¹, M. Dolores Lanzarote Fernández¹, Salvador Chacón-Moscoso^{1,3}

¹ Universidad de Sevilla, ²Consejería de Salud de la Junta de Andalucía, ³Universidad Autónoma de Chile

El nacimiento de un hijo gran prematuro (nacido antes de la semana 29 de gestación) y su posterior ingreso en la Unidad de Cuidados Intensivos Neonatales (UCIN) es un acontecimiento estresante que requiere ajustes emocionales que complican el ejercicio de una parentalidad positiva. Las intervenciones preventivas con las familias deben enfocarse tanto hacia la disminución de los factores de riesgo, como al incremento de los de protección. Nuestro objetivo es conocer el perfil emocional y comprobar la eficacia de un programa psicológico para incrementar la capacitación en un grupo de padres y madres de grandes prematuros. Conformaron la muestra 232 participantes (116 parejas heterosexuales). 64 de las parejas y sus bebés tuvieron un seguimiento estándar y 52 participaron en un programa de intervención de asesoramiento psicológico y apoyo social, proporcionando capacitación basada en información, que se extendía desde el nacimiento del bebé hasta los 12 meses de edad corregida. El programa presentó 12 sesiones, de 120 minutos y periodicidad mensual. Se desarrolló un análisis descriptivo de seguimiento de casos a partir del cual estudiamos las diferencias significativas entre las variables objeto de interés de los grupos de parejas. Los resultados encontrados pusieron de manifiesto una mejor capacitación en los padres y madres que han participado en el programa, pero queda pendiente la generalización de los cuidados centrados en el desarrollo de los bebés ingresados en UCIN, la atención a las necesidades emocionales de los padres y madres, así como su incorporación más activa en la atención al prematuro.

Diseño de un instrumento de medida sobre la violencia en parejas homosexuales en Chile: evidencia de validez basada en el contenido del test. José Antonio Lozano Lozano¹, Salvador Chacón Moscoso^{1,2}, Susana Sanduvete Chaves², Paula Andrea Bravo Armijo¹, Lylian Belén Pizarro Osorio¹ y Luis Alejandro Riveros Dinamarca¹

¹Universidad Autónoma de Chile, ²Universidad de Sevilla

Introducción: la problemática en parejas es una de las más frecuentes en las sociedades occidentales. Existen numerosos instrumentos para medir la violencia en parejas; sin embargo, no existe una escala que presente evidencias de validez y fiabilidad para las parejas homosexuales en Chile. **Objetivo:** diseñar un instrumento de medida sobre la violencia en parejas homosexuales en base a los resultados obtenidos en una revisión sistemática de la literatura especializada. **Método:** se realizó una revisión sistemática de la literatura utilizando 5 bases de datos (*Web of Science, Scopus, Medline, Pubmed y PsycINFO*). Las palabras claves utilizadas fueron: escala, violencia y homosexual en inglés y en español. Dos codificadores de manera independiente clasificaron las dimensiones y seleccionaron los ítems en función de la frecuencia de aparición, adecuación temática, relevancia y novedad de información, y se calcularon los índices de fiabilidad intercodificadores. **Resultados:** de los 334 artículos encontrados, finalmente se analizaron 34 de acuerdo con los criterios de inclusión. Las dimensiones resultantes de este proceso fueron violencia física (17 ítems), violencia psicológica (7 ítems), violencia sexual (6 ítems) violencia económica (5 ítems) y violencia normalizada (7 ítems). En base a estos resultados, se realizó un estudio de validez de contenido. Los ítems que obtuvieron un índice de Osterlind por encima de 0,5 en relevancia, utilidad y viabilidad conformaron la versión definitiva del instrumento.

Eficacia de los programas de formación continua en profesionales de la salud mental: una revisión sistemática. Susana Sanduvete Chaves¹, Salvador Chacón Moscoso^{1,2}, José A. Lozano Lozano², Elena Garrido-Jiménez¹

¹ Universidad de Sevilla y ²Universidad Autónoma de Chile

Objetivos: estudiar la eficacia de los programas de formación para profesionales de la salud mental y detectar las características más relevantes de los estudios revisados. **Método:** la búsqueda se realizó en las principales bases de datos y se revisó también la literatura gris disponible. Las palabras clave fueron "work" AND "evaluation" AND "training program" AND "mental health", con las palabras en inglés y en español. No se limitó temporalmente la búsqueda, siendo realizada hasta junio de 2018. Los criterios de inclusión fueron: (a) el texto completo estaba redactado en español o inglés, (b) el texto completo estaba disponible en los recursos bibliográficos de la Universidad de Sevilla, (c) no estaba duplicado, (d) la formación estaba dirigida a profesionales en activo, (e) los profesionales eran trabajadores de salud mental, (f) se realizó una intervención y (g) el artículo no era una revisión sistemática o un meta-análisis. **Resultados:** se incluyeron 43 artículos en total. Un alto porcentaje (79.1%) encontró resultados positivos para estos programas de formación. Se revisaron las características más comunes en estas intervenciones como, por ejemplo, que era en los países anglohablantes donde más se realizaban

estos programas (51.9% del total). Se detectaron deficiencias relacionadas con cuestiones metodológicas, como que solo había evidencia empírica previa en un 53.5% de las ocasiones. **Conclusiones:** este trabajo determina que los programas de formación dirigidos a los profesionales de la salud mental suelen ser eficaces, aunque podrían hacerse algunas mejoras relacionadas con la calidad metodológica de los estudios examinados.

SIMPOSIO

Modelos de variable latente discreta: alternativas para la medición en presencia de grupos heterogéneos

Coordina: Jesús M. Alvarado Izquierdo

Resumen. Una de las propiedades fundamentales en medición es la invarianza de los instrumentos de evaluación, esta propiedad se aplica al estudio de variables latentes continuas, tanto cuando se asume continuidad en las variables observadas como cuando se aplican modelos para datos categóricos (ej., teoría de la respuesta al ítem -TRI). Ahora bien, cuando los grupos de evaluados son heterogéneos en cuanto a que sus respuestas (patrones de respuesta) no solo dependen del nivel en el rasgo sino también del grupo al que pertenecen se requiere o bien separar los grupos como paso previo a la aplicación de los modelos psicométricos tradicionales (AF o TRI) o bien aplicar directamente modelos de variable latente discreta que permitan obtener los parámetros tanto de los ítems como de los evaluados. En el presente simposio se analizan distintas estrategias para identificar y cuantificar la heterogeneidad, así como distintas alternativas tanto exploratorias como confirmatorias para el tratamiento de este tipo de datos, con aplicaciones concretas en ciencias de la salud y la medición de constructos relevantes en psicología y educación.

Segmentación mediante Análisis de Clases Latentes en la evaluación de la inteligencia emocional. Daniel Ondé¹, Santiago Sastre¹, Teresa Artola¹ y Jesús M. Alvarado¹

¹Universidad Complutense de Madrid

El interés por Análisis de Clases Latentes (LCA) ha crecido en las dos últimas décadas. A pesar de este interés, sigue siendo una técnica de análisis poco utilizada en comparación con otras técnicas tradicionales de segmentación de la muestra como el Análisis Clúster (CA). Los modelos tradicionales funcionan describiendo solamente las relaciones entre variables observables, mientras que el LCA parte de un modelo probabilístico en el que se asume la existencia de una variable latente categórica. El objetivo de este trabajo es mostrar las ventajas que tiene el LCA frente al CA para una correcta segmentación de la muestra, para lo que se emplearán los resultados de un estudio de simulación y una aplicación con datos reales en una prueba de inteligencia emocional (IE). El estudio de simulación se ha realizado con el paquete polCa del programa R, en donde se compara el comportamiento del LCA y del CA ante distintas distribuciones de muestra. Más concretamente, se simuló muestras formadas por 1, 2, 3 y 4 distribuciones de sujetos (clases), y distintos tamaños muestrales ($n = 200, 500$ y 1.000). Asimismo, se han analizado con LCA y CA

los datos de la prueba EMOCINE sobre IE. Los resultados reflejan un mejor comportamiento del LCA tanto en el estudio de simulación como en el caso aplicado. Además, mediante LCA se han identificado tres segmentos de interés respecto a la IE de las personas evaluadas, lo que refleja la necesidad de implementar las técnicas más adecuadas de análisis para una correcta clasificación.

El Análisis de *clusters* aplicado a ciencias de la salud. Marina Lucía Sánchez¹ y Jesús M. Alvarado¹

¹Universidad Complutense de Madrid

Los datos clínicos se caracterizan por ser complejos, de alta variabilidad y una mayoría de variables categóricas o discontinuas. Estas cualidades se asocian a problemas a la hora de intentar explorar la información que nos proporcionan, empujando a los investigadores en Ciencias de la Salud a buscar alternativas que cubran las necesidades de sus datos. El análisis de clusters (o clustering) es un método de análisis no supervisado, considerado de carácter exploratorio. Su objetivo es agrupar objetos a partir de un criterio general: que las muestras incluidas en cada grupo o cluster se asemejen más entre sí que con aquellas pertenecientes a otros clusters. Existen muchos algoritmos para realizar estas agrupaciones en función de las características de los datos (variables numéricas o categóricas), la forma de estudiar las diferencias entre muestras (diferentes cálculos de distancias) y las necesidades del estudio en sí mismo. Son estos elementos los que otorgan a este tipo de análisis de una gran versatilidad. Actualmente se utiliza en una gran variedad de campos, entre los que se incluye el aprendizaje automático, reconocimiento de patrones, minería de datos, análisis estadísticos generales y, más recientemente, está comenzando a sobresalir en ámbitos de investigación clínica. Entre otros, se encuentra en auge la aplicación al estudio de las comorbilidades clínicas: diagnósticos secundarios asociados a las dolencias principales de un paciente. Los análisis de clusters permiten observar grupos de sujetos en función de su combinación de diagnósticos y diversas características.

Aplicación de modelos de diagnóstico cognitivo en la determinación de diferencias en el dominio de habilidades cognitivas entre escuelas privadas y públicas. Pablo Cáceres Serrano¹ y Jesús M. Alvarado²

¹Pontificia Universidad Católica de Valparaíso, ²Universidad Complutense de Madrid

Las diferencias entre escuelas públicas y privadas se han estudiado desde muchas perspectivas y en particular considerando factores económicos y sociales, pues se asume que determinan en algún grado el bienestar y el desempeño de los estudiantes. En este trabajo, se intenta ofrecer una aproximación más detallada a las diferencias entre las escuelas pero evaluando si la segregación puede reflejarse en el dominio de habilidades a la base de una de las destrezas de mayor impacto en educación: La comprensión lectora. Para ello se recurre a la aplicación de los denominados modelos de diagnóstico cognitivo, con el fin de determinar la presencia de diferencias entre patrones de habilidades cognitivas presentes en las escuelas según su titularidad. No

obstante, y dado que los efectos como consecuencia de la segregación económica y social son esperables, se avanza en las diferencias más allá del componente verbal, para determinar si dicha segregación afecta también medidas cognitivas no verbales, como las destinadas a dar cuenta del factor *g*.

Evaluación de la propiedad de invarianza en aplicaciones a múltiples grupos. María Solar Sastre¹ y Jesús M. Alvarado²

¹Universidad Autónoma de Madrid, ²Universidad Complutense de Madrid

La invarianza de medida es una propiedad importante de las escalas, ya que, en caso de no darse, las conclusiones obtenidas pueden deberse a factores característicos de determinados grupos o situaciones que no son el objetivo de estudio. Esta propiedad es especialmente relevante en estudios y encuestas a gran escala, donde hay muchos grupos presentes y el objetivo suele ser la comparación entre ellos. La técnica que se ha utilizado tradicionalmente para comprobar la invarianza de medida es el Análisis Factorial Confirmatorio Multigrupo (AFCM); pero puede resultar inadecuada a medida que aumenta el número de grupos. Por ello, el objetivo de la investigación es comprobar el funcionamiento del AFCM con el aumento progresivo del número de grupos. Para ello, se realizó un estudio de simulación Montecarlo, con un modelo de un factor latente y seis indicadores, donde los factores manipulados fueron: el número de grupos (4, 8, 16 y 32); el tamaño muestral (250, 500 y 1000); el tipo de no-invarianza (pesos o intersecciones) y la magnitud de no-invarianza (de 0,1 a 0,4 en el caso de los pesos; de 0,1 a 0,5 en el caso de las intersecciones). Los resultados del estudio apuntan a que el funcionamiento del AFCM se deteriora a medida que aumentan el número de grupos.

La generalización de la fiabilidad y sus consecuencias en los meta-análisis sustantivos. Mercedes Ovejero-Bruna¹, Andrea C. Brabete¹ y Jesús M. Alvarado¹

¹Universidad Complutense de Madrid

La calidad de la medida es un asunto que va más allá de las aplicaciones individuales de un test a un grupo representativo. En los meta-análisis de Generalización de la Fiabilidad (RG) se estudia cómo la estimación combinada de la fiabilidad se ve afectada por diversos moderadores. Esta heterogeneidad de la fiabilidad puede afectar a otros meta-análisis de carácter más sustantivo, reduciendo su potencia. En el presente trabajo se estudia, mediante un caso aplicado, cómo varía la consistencia interna de las puntuaciones de las diferentes dimensiones de un test desde el punto de vista meta-analítico, y cómo dicha variabilidad afecta a un meta-análisis sustantivo basado en las correlaciones de Pearson. Los resultados de este estudio muestran que la consistencia interna tiene una alta variabilidad que se ve afectada en parte por la variabilidad de las puntuaciones, el tamaño muestral y la versión del test. La alta heterogeneidad observada en el estudio de RG se relaciona con la consistencia de los resultados encontrados en el meta-análisis de correlaciones de Pearson. Tras la exposición de resultados se ofrecen una serie de buenas prácticas a la hora de reportar la fiabilidad de las puntuaciones y futuras líneas de investigación.

SIMPOSIO

Metodología observacional como 'mixed method' en sí misma (I)

Coordinan: M. Teresa Anguera y Oleguer Camerino

Resumen. Después de varias décadas de un radical enfrentamiento entre partidarios de las metodologías cualitativa y cuantitativa, en el último cuarto de siglo se ha abierto un espacio muy relevante, el de los *mixed methods*, que en todo mundo se ha expandido de forma vertiginosa, facilitando la complementación, primero, y la integración, en los últimos años, de elementos cualitativos y cuantitativos. Sin embargo, todavía existen importantes discrepancias y diferentes corrientes internas, fruto seguramente de una cierta crisis de crecimiento. En este Congreso AEMCCO nos interesa enfatizar, en el marco del intenso desarrollo de los *mixed methods*, que la metodología observacional, considerando el perfil idóneo que la caracteriza, puede considerarse en sí misma como mixed method, teniendo en cuenta la transformación de datos que deriva del *connecting*, a lo largo de las etapas CUAL-CUAN-CUAL. En efecto, la recogida de datos es cualitativa, y todos los registros sistemáticos obtenidos tienen este carácter. Sin embargo, mediante la construcción de un instrumento de observación *ad hoc*, y un adecuado establecimiento de criterios de segmentación del flujo de conducta, tanto en observación directa como en indirecta, es posible sistematizar el registro, transformándolo en una matriz de códigos, que se puede tratar cuantitativamente, a pesar de disponer de datos cualitativos. Y los resultados cuantitativos de dicho análisis se interpretarán cualitativamente, completando el ciclo CUAL-CUAN-CUAL. Esta consideración abre un espacio relevante, que permite una intensificación del interés por el *quantitizing*, aspecto polémico a nivel internacional en *mixed methods*. En este simposio se presentan 5 comunicaciones, que abordan diferentes aspectos conceptuales, metodológicos y de aplicación.

LINCE PLUS: nueva plataforma de registro observacional. Alberto Soto¹, Oleguer Camerino^{1,2}, Marta Castañer^{1,2}, Xavier Iglecias¹ y M. Teresa Anguera³

¹Instituto Nacional de Educación Física de Cataluña (INEFC)-Universidad de Lleida-Barcelona, ²Instituto de Investigación Biomédica de Lleida (IRBLLEIDA), Universidad de Lleida, ³Universidad de Barcelona

Lince (Gabin, et al. 2012) es una plataforma de registro videográfico que por su versatilidad se ha ido consolidando, utilizándose en la actualidad en multitud de investigaciones de carácter observacional. La evolución natural de la plataforma, gracias a las ayudas institucionales de los proyectos de investigación I+D+I y del INEFC de Cataluña, se centra actualmente en la posibilidad de instalarse y/o utilizarse en cualquier sistema operativo o plataforma, incluyendo tabletas y móviles. También permite la ejecución de tareas de visualización en dos pantallas de vídeos sincronizados del mismo episodio en perspectivas y ángulos diferentes y el registro de datos observacionales de forma colaborativa y simultánea de varios observadores a la vez. Esta posibilidad de simultaneidad permitirá compartir el análisis del comportamiento mediante la comprobación de la concordancia, de los mismos episodios observados por diferentes observadores, y enriquecerlo con pers-

pectivas diferentes. La herramienta incorpora cálculos estadísticos instantáneos al haber incorporado el motor estadístico basado en Java y R (Ruiz y López, 2009) gracias a las adaptaciones del código libre de BeDataDriven (<http://www.bedatadriven.com/>). Esta nueva versión del programa Lince Plus permite el análisis de las conductas en una plataforma ágil y versátil de software libre, que en el futuro podrá incorporar técnicas de inteligencia artificial. Con estas incorporaciones hemos dado un paso importante en el nuevo LINCE PLUS para apoyar las necesidades investigadoras de nuestra comunidad científica.

A novel approach to integrate quantitative and qualitative data for fencing tactical efficiency analysis. Ana Teresa Campaniço¹, João José¹, M. Teresa Anguera² y Jorge Campaniço¹

¹University of Trás-os-Montes and Alto Douro, ²Universidad de Barcelona

This study illustrates how inertial quantitative data and the coach's qualitative evaluation of a fencer's tactical efficiency in training can be combined to provide autonomous, extrinsic feedback during individual practice. The device that produces this qualitative evaluation uses Observational Methodology, from a Mixed Methods perspective, and Machine Learning to detect the hidden patterns and relationships within both types of the data. This technological tool device is useful for behavior analysis in sport, in context, as well as other fields that study human behavior. 2000 fencing actions of two technical exercises were collected via an inertial sensor applied on the fencer's weapon hand. A dichotomous qualification of the gesture of each sample was also performed by the coach, based on two *ad hoc* observation instruments. 72 statistical features were extracted from the data to train an Artificial Neural Network. The resulting model achieved a 78% and 71,5% classification accuracy for each exercise respectively. As such, we demonstrate that not only is possible for an application to successfully classify highly variable data qualitatively, but it can also perform a quality control of how well the data describes the surrounding context.

Mixed systematic review en futsal masculino, considerando la observación sistemática como mixed method. María Preciado¹, Mauricio Orlando Torres², Daniel Lapresa³ y M. Teresa Anguera¹

¹Universidad de Barcelona, ²Universidad Nacional de Colombia, ³Universidad de La Rioja

No nos consta que previamente se hayan realizado revisiones sistemáticas de *futsal*/masculino desde la perspectiva *mixed method*. De acuerdo con el perfil que la caracteriza, destaca la integración realizada entre elementos cualitativos y cuantitativos, como exige la perspectiva *mixed method*. Los documentos primarios se ajustan a la metodología observacional y ésta en determinadas ocasiones puede ser considerada en sí misma *mixed method*. Está argumentado en Anguera y Hernández-Mendo (2016), en donde se alega que el planteamiento de la metodología observacional, basada en la complementariedad cualitativa-cuantitativa, consiste en una ordenada sucesión de etapas, que son: (1) obtención de datos cualitativos, (2) transformación progresiva en un proceso de cuantificación de registros -aunque el *quantitizing* que proponemos no es el utilizado por Tashakkori y Teddie (1998) o Sandelowski, Voils, y Knafp (2009)-, (3) tratamiento cuantitativo de la in-

formación desde técnicas analíticas adecuadas para datos categóricos (Anguera, Camerino, Castañer, y Sánchez-Algarra, 2014) y (4) interpretación de los resultados con retorno a lo cualitativo, perfectamente encajado en los *mixed methods*. Se ha efectuado una recopilación de 17 documentos primarios, que se han analizado desde diversos criterios procedimentales, tomando como referencia la guía adaptada de GREOM (Portell, Anguera, Chacón y Sanduverte, 2015) y valorando las vías de integración que se han utilizado por los respectivos autores.

CMASP: instrumento para clasificar la comunicación verbal y no verbal en psicoterapia desde la perspectiva mixed methods. Luca Del Giacco^{1,2}, Silvia Salcuni² y M. Teresa Anguera¹

¹Universidad de Barcelona y ²Universidad de Padova

Durante el intercambio comunicativo paciente-terapeuta, los componentes verbales y no verbales generan un proceso de influencia mutua. En particular, las conductas verbales, vocales y de interrupción regulan el cambio al impactar en la co-construcción de significados. La investigación psicoterapéutica tradicional considera estas dimensiones como independientes. Por tanto, los objetivos del estudio son: (a) desarrollar un instrumento (el Sistema de Análisis de Modalidades Comunicativas en Psicoterapia, CMASP) para analizar el discurso terapéutico superando los límites anteriores; (b) destacar su fiabilidad; (c) mostrar su funcionamiento mediante el análisis de las conductas vocales. Basado en la función performativa de la Teoría de los Actos de Habla, el CMASP es un sistema de observación indirecta ad hoc desde la perspectiva de los Mixed Methods que registra y analiza las conductas mencionadas combinando el enfoque cualitativo y cuantitativo. La muestra consta de 30 grabaciones de audio y transcripciones literales de 10 psicoterapias individuales (8.327 turnos de habla). El instrumento se deriva de la combinación de formato de campo y sistemas de categorías: 4 dimensiones, cada una caracterizada por un sistema de categorías exhaustivas y mutuamente excluyentes (33 categorías en total). El CMASP muestra alta fiabilidad intra e interobservadores a nivel global, dimensional y categorial, que lo hace flexible y capaz de identificar múltiples aspectos de la comunicación y evolución del discurso. La comunicación vocal transmite estados psicológicos y emocionales subyacentes. Mediante del análisis secuencial de retardos y análisis de coordenadas polares, el CMASP proporciona evidencias de cómo esta comunicación se desarrolla durante la interacción paciente-terapeuta.

El Estudio Longitudinal de Coyoacán: identificación de *T-patterns*. Carlos Santoyo¹, Gudberg K. Jonsson², Mariona Portell³, M. Teresa Anguera⁴, Andrea Allegro⁵ y Ligia Colmenares¹

¹Universidad Nacional Autónoma de México, ²University of Iceland, ³Universitat Autònoma de Barcelona, ⁴Universidad de Barcelona, ⁵Università Degli Studi di Padova

Este trabajo se dirige al desarrollo de una estrategia para la identificación de *T-patterns* en un estudio longitudinal con participantes preescolares. Se realizaron tres cortes de observación de seis meses cada uno, en un escenario preescolar en dos escenarios: aula y zona de juego. Con base en la metodología observacional se recogieron datos de tres niños en edad preescolar con la finalidad de identificar *T-patterns* a lo largo del tiempo, en distintos

escenarios. Se comparan datos molares con *T-patterns* en la perspectiva de encontrar vías para describir, explicar y predecir de mejor forma la persistencia en actividades vinculadas con las metas educativas y de socialización con aquellas no directamente relacionadas. Se discuten los resultados con base en la estabilidad y cambio de dichos patrones, las diferencias entre participantes y sus implicaciones educativas y evolutivas.

SIMPOSIO

Perfiles profesionales en metodología

Coordina: Miguel Sorrel

Resumen. Este simposio, compuesto por cinco presentaciones, pretende ofrecer un repaso al estado actual de diferentes perfiles profesionales en metodología. Se abordará la demanda a nivel internacional de psicómetras, las incursiones en el mundo de la ciencia de datos, el trabajo día a día en una editorial de test y en una empresa de recursos humanos, así como las aportaciones en investigación sanitaria clínica. Participan distintos profesionales y académicos con el objetivo de describir la situación actual y comentar las perspectivas de futuro.

Psicómetras por el mundo: un bien escaso. José Muñoz¹ y Avi Allalouf¹

¹Universidad de Oviedo, ²National Institute for Testing and Evaluation, Israel

El objetivo de esta presentación es revisar la necesidad que hay en la mayoría de los países desarrollados de profesionales especializados en la medición psicológica y educativa. Se inicia con el repaso de las vías existentes en España para llevar a cabo una especialización en psicometría, apuntando la necesidad de un Máster de ámbito nacional centrado en este campo. A continuación, se revisa la situación en varios países de los profesionales especializados en medición en ciencias sociales y de la salud. En concreto se comentan los casos de España, Israel, Rusia, Suecia, Noruega y Estados Unidos. En todos estos países el denominador común es una mayor demanda que oferta de psicómetras, variando en cada uno de ellos los sistemas de formación especializada. Se comentan las características de cada uno de ellos y se sacan algunas conclusiones generales comunes a todos ellos. Se analizan, finalmente, los sectores laborales que ofrecen un mayor atractivo y demanda para los profesionales, y se comentan algunas perspectivas de futuro que puedan orientar a las personas más jóvenes en proceso de especialización, o quienes estén pensando en elegir esta opción profesional.

Empleabilidad para metodólogos en el mundo de la ciencia de datos. Alejandro Martínez¹, Alejandro Hernández², Alejandro Vidal³ y Antonio López¹

¹Compai, ²Telefónica LUCA, ³BBVA Data&Analytics

¿Cuáles son las diferencias entre un perfil profesional puramente estadístico y uno más orientado a ciencia de datos? ¿Qué cualidades presentes en un metodólogo están buscando hoy en día las empresas? En esta comunicación hablaremos sobre la experien-

cia profesional de tres metodólogos en el campo de la ciencia de datos y expondremos dos casos de uso reales en los que se podrá comprender el papel de un metodólogo en estos equipos de trabajo. Además hablaremos de las habilidades adicionales que debe adquirir cualquier alumno del máster para comenzar a trabajar en el sector.

El papel del psicómetra dentro de una editorial de test. Belén Ruiz Fernández¹

¹TEA Ediciones

¿Qué hace un psicómetra? Probablemente todos los que estén leyendo este resumen tengan una idea muy clara de cuáles son las competencias y habilidades de un psicómetra, pero ¿qué hace un psicómetra en una editorial? Ante esta pregunta las respuestas empiezan a ser menos claras, con unos límites más difusos sobre sus funciones concretas. Como bien sabrán, el desarrollo de una prueba psicométrica tarda varios años. Desde que se plantea la idea de la medición de un constructo concreto hasta que se publica una prueba (compuesta por un manual, un cuestionario y un sistema de corrección), el psicómetra ha tenido que ir adaptando su rol a cada etapa de ese proceso siendo investigador, metodólogo, editor, coordinador de equipos de trabajo, escritor... Al igual que en otros ámbitos, la formación universitaria es muy específica y no se centra en el desarrollo de todas estas competencias transversales necesarias para la correcta consecución de todos los objetivos. Por ello, en esta presentación se pretende proporcionar una idea general del proceso completo del desarrollo de un test con el objetivo de tener una idea global y real de todas las implicaciones que un proyecto de estas características requiere y las competencias que la persona a cargo debe tener.

El papel del metodólogo en recursos humanos. Ignacio Jiménez¹ y Diego Lomas¹

¹Verisure Securitas Direct

En los últimos años los departamentos de recursos humanos empiezan a ser conscientes de la necesidad de tomar decisiones basadas en datos. Han aparecido distintos perfiles profesionales que, organizados en diferentes formatos según la compañía, vienen a satisfacer tales necesidades. Esta coyuntura ha aumentado sin duda alguna el valor del metodólogo en el mercado que, en conjunción con otros perfiles técnicos y de negocio, supone un importante activo para aquellas empresas que quieren implementar o mejorar sus procesos de toma de decisiones basadas en datos dentro del ámbito de recursos humanos. Sumado a ello, el perfil del metodólogo no se define en exclusiva por sus conocimientos técnicos, sino por una forma estructurada y lógica de enfrentar problemas abstractos que dota a este perfil de una polivalencia interesante y una capacidad de contribución a otras áreas analíticas de una empresa. En la exposición reflexionaremos sobre la aportación del metodólogo desde nuestra experiencia profesional.

Salidas profesionales del metodólogo en el contexto de la investigación sanitaria clínica. Alejandro de la Torre¹

¹CIBERSAM (Centro de Investigación Biomédica en Salud Mental)

Vivimos en un mundo de constante cambio, con innovaciones que se van incorporando a gran velocidad a los sistemas de producción, gestión y provisión de servicios. Esto implica cada vez más,

la incorporación de procesos de toma de decisión basados en la evidencia. El contexto sanitario no es un sector exento de esta fenomenología. Cada vez más se va integrando la decisión basada en la evidencia a la implementación de protocolos de diagnóstico e intervención en salud. El paso previo necesario es la aplicación de modelos estadísticos sólidos y contrastados que permitan manejar datos derivados de investigación. En esta presentación, se pretende poner en valor el papel del metodólogo en los contextos de investigación sanitaria clínica. Además, se pretende describir posibles ámbitos de aplicación del conocimiento metodológico en el ámbito de la investigación sanitaria. El metodólogo aporta un conocimiento altamente valorable y sofisticado en equipos de investigación en diversos ámbitos sanitarios. La transversalidad de este conocimiento posibilita la integración en equipos interdisciplinarios y dirigidos a diversas líneas de aplicación. En conclusión, la salida profesional del metodólogo en el contexto de la investigación clínica es bastante fructífera y gratificante, pero no exenta de desafíos y retos que deben ir superándose en los próximos años.

SIMPOSIO

Evaluación de la calidad métrica de instrumentos psicométricos: procedimientos y recomendaciones

Coordina: Jesús M. Alvarado

Resumen. El obtener medidas válidas que permitan inferir adecuadamente el nivel del rasgo de los sujetos a partir de los indicadores o ítems de los instrumentos psicométricos desarrollados por expertos, es una tarea compleja que a pesar de su larga trayectoria, desarrollos teóricos y de la intensa investigación desarrollada está aún pendiente de resolver cuestiones claves para determinar la calidad de los instrumentos. En el presente simposio nos centraremos en algunas de estas interrogantes ofreciendo soluciones o alternativas metodológicas a cuestiones como el filtrado y selección de ítems para el desarrollo de escalas de alta calidad métrica a partir de la aplicación de la TRI no paramétrica, la determinación de la calidad de los ítems mediante la aplicación del modelo de respuesta graduada en contextos de evaluación docente, el análisis de la fiabilidad en escalas multidimensionales, el control de los sesgos de respuesta y la necesidad de elaborar recomendaciones claras y fundamentadas sobre buenas prácticas para una medición de calidad.

Construcción de escalas de alta calidad métrica mediante la Teoría de Respuesta al ítem no paramétrica. Mirko Antino¹, Daniel Ondé¹ y Jesús M. Alvarado¹

¹ Universidad Complutense de Madrid

Las técnicas factoriales son la principal herramienta para explorar la dimensionalidad de los constructos. Estas técnicas en su formulación clásica asumen los supuestos de continuidad y normalidad multivariante. Sin embargo, cuando estos supuestos se incumplen las estimaciones de los parámetros son imprecisas y se pueden producir soluciones incorrectas al juzgar la dimensionalidad de los constructos. Una alternativa libre de supuestos distribucionales y pensada para datos no cuantitativos es el modelo de Mokken que forma de la denominada Teoría de Respuesta al

Item no paramétrica y que recientemente se ha observado un interés creciente por la técnica en el ámbito aplicado. En el presente trabajo de simulación, se muestran la capacidad del procedimiento para la construcción de escalas unidimensionales formadas por ítems de alta calidad métrica en escalas Likert. Los resultados muestran en un amplio conjunto de condiciones (tamaños muestrales, número de ítems, pesos factoriales y correlación entre factores) que MSA tiene un funcionamiento óptimo en ítems de calidad media (i.e., peso factorial 0.7) en el procedimiento estándar y que en ítems de calidad alta se requieren ajustes adicionales para obtener un funcionamiento adecuado.

Aplicación del modelo de Respuesta Graduada a cuestionarios de calidad docente. Lady Catheryne Lancheros Florian¹ y Vicente Ponsoda²

¹Universidad Complutense de Madrid, ²Universidad Autónoma de Madrid

La evaluación de la satisfacción de los estudiantes sobre la labor que realiza en docente es una cuestión que ha generado múltiples análisis e investigaciones con el objetivo de dar evidencias de su validez. Se han generado diversos tipos de evaluaciones (SET) por sus siglas en inglés, de acuerdo con el énfasis que le quiera dar cada institución educativa. En este trabajo se ajusta un modelo de la Teoría de Respuesta al Ítem (TRI) a los ítems del instrumento a través del cual, el estudiante valora la satisfacción con el profesor que le ha impartido un curso durante su formación universitaria mediante una escala de siete (7) ítems tipo Likert (UAM, 2015). Los datos provenían de las aplicaciones realizadas durante cinco años entre los periodos 2012/13 y 2016/17. Se aplicó el modelo de respuesta graduada (MRG) de Samejima a una muestra de 16.390 estudiantes universitarios. En primer lugar, se confirmó el supuesto de unidimensionalidad de los ítems. Se realizó la estimación de los parámetros del modelo (R:ltm[gm], Dimitris, 2006). La mayoría de los parámetros de localización tendieron a ubicarse en niveles medio alto de la escala. Los parámetros de discriminación adoptaron valores entre moderados y altos. La función de información del test evidenció que la escala es más precisa para discriminar individuos con niveles medios-bajos del rasgo. Los resultados evidenciaron los ítems que brindan más información sobre la satisfacción docente y plantea cuestiones sobre el uso de estos modelos dada las características de los datos.

La estimación de la fiabilidad en estructuras multidimensionales. Italo Trizano-Hermosilla¹ y Jesús M. Alvarado²

¹Universidad de la Frontera y ²Universidad Complutense de Madrid

Para estimar la fiabilidad, es decir la precisión, de las puntuaciones obtenidas con un instrumento de medida es necesario evaluar previamente la estructura latente al conjunto de ítems, resultando habitual encontrar estructuras con algún grado de multidimensionalidad. Uno de los modelos psicométricos adecuados para evaluar el grado en que un conjunto de ítems responde a una estructura unidimensional o multidimensional es el Bifactor, en donde se considera la presencia de un factor común a todos los ítems y algunos factores grupales o específicos de algunos subconjuntos de ítems. En este contexto, la estimación de la fiabilidad se puede obtener tanto para el factor general como para los factores específicos, siendo además posible obtener una estimación de la proporción de varianza explicada total de todos los factores considerados en el modelo. De esta forma se compara la estimación ofre-

cida por distintos coeficientes de fiabilidad, alfa, omega jerárquico y total, *greatest lower bound*, en condiciones de datos multidimensionales. Se discuten los sesgos encontrados al estimar la fiabilidad del factor general y la de todos los factores, así como las limitaciones que algunos estimadores ordinales pueden presentar en estas condiciones.

Procedimientos para el control de los sesgos de respuesta. Ana Hernández Dorado¹ y Jesús M. Alvarado¹

¹ Universidad Complutense de Madrid

La utilización de escalas balanceadas, es decir, la inclusión equitativa de ítems directos e inversos, es uno de los métodos clásicos para controlar el sesgo de aquiescencia (AQ). Sin embargo, ésta técnica no parece ser suficiente para que los test sean inmunes al sesgo. Por ese motivo se han desarrollado, desde diferentes perspectivas, métodos para el tratamiento de AQ. Se realizó un estudio de simulación Monte Carlo, con el objetivo de comparar el rendimiento de dos procedimientos en escalas unidimensionales, tanto balanceadas como no balanceadas: un análisis factorial confirmatorio (AFC), que incluía un factor de aquiescencia y un análisis factorial exploratorio (AFE) con rotación *target*. Ambos se contrastaron con un modelo "base" (análisis factorial confirmatorio de un factor) que ignoraba la presencia de AQ.

Recomendaciones y buenas prácticas para una medición psicológica de calidad. David Paniagua Sánchez¹, Marta Aparicio¹ y Jesús M. Alvarado¹

¹ Universidad Complutense de Madrid

En los últimos años, se está viendo un crecimiento en la importancia de la calidad de la investigación. Los principales problemas de la calidad en la investigación se derivan de una mala praxis: Falta de clarificación del marco teórico, problemas derivados del diseño de investigación (variables y muestra), así como una incorrecta utilización de las herramientas estadísticas o una mala interpretación y presentación de los resultados. Teniendo esto en cuenta, diversos autores han propuesto medidas de control de calidad que tuvieran un efecto positivo en la calidad de la investigación: medidas de preregistro y estudios de replicación, guías de estándares y una política de ciencia abierta. En este sentido, se pretende exponer de forma inicial estos conceptos y su relación con el marco de la medición.

SIMPOSIO

Meta-análisis: avances estadísticos

Coordina: José A López López

Resumen. El meta-análisis es una metodología de investigación que se ha consolidado como un método imprescindible para una acumulación eficiente de la investigación y, en consecuencia, para el correcto avance del conocimiento científico. Su característica esencial es que permite integrar los resultados de un conjunto de estudios primarios sobre una misma pregunta mediante la aplicación de métodos estadísticos. Las características peculiares de los datos meta-analíticos requieren investigación continua para

mejorar sus métodos estadísticos. El propósito de este simposio es presentar los resultados de cinco investigaciones dirigidas a tal fin. La primera comunicación, presentada por Fulgencio Marín Martínez, ofrece los resultados de una revisión sistemática de meta-análisis en el ámbito de la Educación para investigar las características metodológicas y estadísticas de los meta-análisis en ese ámbito. La segunda comunicación, presentada por José Antonio López López, ofrece un ejemplo ilustrativo de las bondades de los meta-análisis en red y, en especial, de su aplicación a intervenciones complejas. La tercera comunicación, presentada por Belén Fernández Castilla, habla de las posibilidades de los modelos multinivel en meta-análisis para tratar datos dependientes. La cuarta comunicación corre a cargo de Laura Badenes Ribera, y presenta evidencias sensibles sobre cómo pueden cambiar los resultados de un meta-análisis según el índice del tamaño del efecto que se utilice. Finalmente, Desirée M^a Blázquez Rincón presenta la quinta comunicación, en la que ofrece los resultados de un estudio de simulación dirigido a evaluar el sesgo y la eficiencia de diferentes estimadores de la varianza inter-estudios, con especial interés en el caso de que no se cumpla el supuesto de normalidad de la distribución de efectos paramétricos.

Una revisión sistemática de meta-análisis en el ámbito educativo. Fulgencio Marín Martínez¹, Eva Herrera Gutiérrez¹, Rubén López Nicolás¹, Julio Sánchez Meca¹, María Rubio Aparicio¹ y José Antonio López López¹

¹Universidad de Murcia

Se revisan 22 meta-análisis publicados en las cinco revistas de mayor índice de impacto en la categoría Education & Educational Research del JCR, con el objetivo de explorar las características del análisis estadístico aplicado en los meta-análisis. El número de estudios de los meta-análisis osciló entre 10 y 144. El 59,1% de los meta-análisis fueron de estudios experimentales sobre la eficacia de intervenciones educativas, el 27,3% fueron de estudios correlacionales y el 13,6% combinaron estudios experimentales y correlacionales en el mismo meta-análisis. Todos los índices del tamaño del efecto de un mismo meta-análisis se transformaron a índices d (68,2% de los meta-análisis) o a índices r (31,8%). El 86,4% de los meta-análisis aplicaron un modelo de efectos aleatorios, el 9,1% aplicaron tanto un modelo de efectos aleatorios como un modelo de efecto fijo, y tan solo un meta-análisis aplicó un modelo de efecto fijo. Se discuten las limitaciones de los meta-análisis en cuanto a la calidad del reporte de los análisis estadísticos, las técnicas de estimación por intervalo del efecto medio, la presentación de tablas y gráficos, el análisis de variables moderadoras, el abordaje del problema de la dependencia, el análisis del sesgo de publicación y el tratamiento estadístico de los *outliers*. Finalmente se aportan recomendaciones para la mejora de los futuros meta-análisis en el ámbito educativo.

El Meta-análisis en red como herramienta para identificar componentes efectivos en intervenciones complejas. José A. López López¹, Nicky J. Welton² y Deborah M. Caldwell²

¹Universidad de Murcia y (2) University of Bristol

El meta-análisis en red (MER) permite combinar la evidencia existente acerca de múltiples intervenciones para un mismo problema

de salud. En el ámbito de la salud mental, las intervenciones suelen ser complejas, ya que incluyen diferentes ingredientes activos o "componentes". Los modelos MER para componentes (MER-C), recientemente desarrollados dentro de un enfoque bayesiano, permiten estimar la contribución de cada componente y/o combinación de componentes a la efectividad total de una intervención compleja. En esta presentación se ilustra el uso de modelos MER y MER-C con una comparación entre diferentes tipos de terapia cognitivo-conductual (TCC) para el tratamiento de la depresión en adultos. Se incluyeron 91 ensayos clínicos aleatorizados y se encontró evidencia de que los pacientes que recibieron TCC mostraron un mayor descenso a corto plazo en las puntuaciones de sintomatología depresiva que los sujetos que recibieron tratamiento convencional, con una diferencia estandarizada de cambio medio de -1,11 (intervalo de credibilidad al 95% de -1,62 a -0,60) para TCC cara a cara, -1,06 (-2,05 a -0,08) para TCC híbrida y -0,59 (-1,20 a 0,02) para TCC multimedia, mientras que se observó un efecto perjudicial para los sujetos asignados a lista de espera (0,72; 0,09 a 1,35). No se encontró evidencia de ningún componente o combinación de componentes directamente implicados en esta efectividad diferencial; además, las estimaciones de los efectos de tratamiento sugieren que las modalidades híbridas y multimedia de la TCC podrían ser tan efectivas como la modalidad cara a cara, aunque estos resultados deben ser interpretados con precaución.

Revisión sistemática sobre el uso de modelos multinivel con más de un efecto aleatorio en meta-análisis. Belén Fernández Castilla¹ y Wim Van den Noortgate¹

¹Universidad de Leuven

La estructura de los datos en meta-análisis es jerárquica: los tamaños del efecto (nivel 1) están agrupados en estudios (nivel 2). Por lo tanto, el tradicional modelo de efectos aleatorios es equivalente a un modelo multinivel con un coeficiente aleatorio, que refleja la variabilidad de los tamaños del efecto entre estudios. Cuando la estructura de los datos es más compleja (p. ej., estudios que incluyen varios tamaños del efecto, o estudios que están agrupados en otros clústeres, como países), se pueden aplicar modelos multinivel con más efectos aleatorios que reflejen estas fuentes adicionales de variabilidad (p. ej., variabilidad de tamaños del efecto dentro de los estudios, o entre países). Esta revisión sistemática tiene como objetivos describir qué tipos de modelos multinivel (con más de un coeficiente aleatorio) se aplican en meta-análisis, y explicar cómo, en ocasiones, modelos multinivel más complejos deberían haberse aplicado para modelar adecuadamente la estructura jerárquica o cruzada de los datos. La búsqueda sistemática resultó en la selección de 178 meta-análisis, de los cuales 162 aplicaron modelos multinivel con dos efectos aleatorios, 8 usaron modelos con tres efectos aleatorios, 6 aplicaron modelos de efectos aleatorios cruzados y 2 utilizaron modelos multinivel con cuatro efectos aleatorios. Los resultados indican que modelos multinivel con dos efectos aleatorios son sistemáticamente aplicados para tratar la dependencia entre tamaños del efecto. Sin embargo, a veces la inclusión de tres o más efectos aleatorios, jerárquicos o cruzados, habría sido necesaria para modelar correctamente fuentes adicionales de dependencia. Algunos ejemplos sobre este último escenario son presentados.

La elección del índice del tamaño del efecto: su influencia en los resultados del meta-análisis. Laura Badenes Ribera¹, Cristian Molla Esparza¹ y Julio Sánchez Meca²

¹Universidad de Valencia, ²Universidad de Murcia

La *práctica basada en la evidencia* es una herramienta metodológica que enfatiza la importancia de la evidencia científica en la práctica profesional. En este contexto, el meta-análisis, una metodología de investigación, puede ayudar a los profesionales en la toma de decisiones bien informadas basadas en la mejor evidencia científica disponible dado que integra cuantitativamente los resultados de un conjunto de estudios empíricos sobre un mismo problema. El propósito de este estudio fue comparar el funcionamiento de diversos índices del tamaño del efecto de la familia de la diferencia de medias estandarizadas para: (a) estimar el efecto medio, (b) estimar la varianza inter-estudios y (c) contrastar la significación estadística del efecto medio. En concreto, se calcularon diferencias de medias estandarizadas en el postest y diferencias de medias estandarizadas entre los cambios del pretest al postest. Para ello se han elaborado diversas bases meta-analíticas a partir de estudios de meta-análisis publicados sobre eficacia de tratamientos, sobre las que se ha trabajado para valorar empíricamente en qué grado ofrecen resultados discrepantes los distintos índices del tamaño del efecto estudiados. Los resultados muestran diferencias muy relevantes cuando sobre una misma base de datos meta-analítica se aplican índices del tamaño del efecto diferentes. Se discuten las consecuencias que estos resultados pueden tener en la interpretación de resultados meta-analíticos.

Robustez de los diferentes estimadores de la varianza inter-estudios en meta-análisis de efectos aleatorios: una simulación Monte Carlo. Desirée M^a Blázquez Rincón¹, Julio Sánchez Meca¹, Juan Botella² y Manuel Suero²

¹Universidad de Murcia, ²Universidad Autónoma de Madrid

La varianza inter-estudios permite trabajar con sencillez con modelos de efectos aleatorios en meta-análisis, los cuales contemplan más fuentes de variabilidad que el modelo de efectos fijos, favoreciendo la generalización de los resultados meta-analíticos. La varianza inter-estudios constituye uno de los parámetros más importantes en los meta-análisis de efectos aleatorios, puesto que resulta necesaria para describir la distribución paramétrica y estimar el efecto medio. Actualmente contamos con diversos estimadores de la varianza inter-estudios, y varios métodos para calcular su intervalo de confianza. Estudios anteriores han dado cuenta del impacto en las conclusiones estadísticas que tiene la elección de un estimador u otro, así como de las diferencias de sesgo y eficiencia que presentan en función de variables como la magnitud de la varianza inter-estudios, el número de estudios incluidos, o el tamaño muestral de los estudios primarios. Sin embargo, otra de las variables que puede afectar a la estimación de la varianza inter-estudios es la desviación del supuesto de normalidad de la distribución paramétrica. Esto resulta especialmente relevante en los meta-análisis de generalización de la fiabilidad, dado el carácter asimétrico que suelen exhibir las distribuciones de los coeficientes de fiabilidad, una asimetría que no siempre se logra corregir con las transformaciones al uso. El presente estudio se sirve de la simulación Monte Carlo con el objetivo de: (1) comparar los resultados en cuanto a sesgo y eficiencia de los estimadores más habituales en contextos de no normalidad y (2) comprobar si se replican los resultados encontrados en estudios anteriores.

SIMPOSIO

Metodología observacional como *mixed method* en sí misma (II)

Coordinan: M. Teresa Anguera y Mariona Portell

Resumen. Después de varias décadas de un radical enfrentamiento entre partidarios de las metodologías cualitativa y cuantitativa, en el último cuarto de siglo se ha abierto un espacio muy relevante, el de los *mixed methods*, que en todo mundo se ha expandido de forma vertiginosa, facilitando la complementación, primero, y la integración, en los últimos años, de elementos cualitativos y cuantitativos. Sin embargo, todavía existen importantes discrepancias y diferentes corrientes internas, fruto seguramente de una cierta crisis de crecimiento. En este congreso AEMCCO nos interesa enfatizar, en el marco del intenso desarrollo de los *mixed methods*, que la metodología observacional, considerando el perfil idóneo que la caracteriza, puede considerarse en sí misma como *mixed method*, teniendo en cuenta la transformación de datos que deriva del *connecting*, a lo largo de las etapas CUAL-CUAN-CUAL. En efecto, la recogida de datos es cualitativa, y todos los registros sistemáticos obtenidos tienen este carácter. Sin embargo, mediante la construcción de un instrumento de observación *ad hoc* y un adecuado establecimiento de criterios de segmentación del flujo de conducta, tanto en observación directa como en indirecta, es posible sistematizar el registro, transformándolo en una matriz de códigos, que se puede tratar cuantitativamente, a pesar de disponer de datos cualitativos. Y los resultados cuantitativos de dicho análisis se interpretarán cualitativamente, completando el ciclo CUAL-CUAN-CUAL. Esta consideración abre un espacio relevante, que permite una intensificación del interés por el *quantitizing*, aspecto polémico a nivel internacional en *mixed methods*. En este simposio se presentan 5 comunicaciones, que abordan diferentes aspectos conceptuales, metodológicos y de aplicación.

La observación indirecta en el marco de los *mixed methods*: propuestas de proceso. M. Teresa Anguera¹, Mariona Portell², Salvador Chacón Moscoso^{3,4} y Susana Sanduvete-Chaves³

¹Universidad de Barcelona, ²Universitat Autònoma de Barcelona, ³Universidad de Sevilla y ⁴Universidad Autónoma de Chile

El carácter documental de la observación indirecta incluye una pluralidad de tipologías que sigue ampliándose con la incorporación de nuevas tecnologías de registro, difusión y almacenamiento, y engloba básicamente textos documentales obtenidos por grabación magnetofónica de la conducta verbal de un participante en una situación natural, textos escritos procedentes de cualquier canal de comunicación, construcción de "objetos" relacionados con el objetivo, medidas de erosión, etc. En los últimos 25 años ha eclosionado el movimiento de los *mixed methods*, que implica la recogida, análisis e interpretación de datos cualitativos y cuantitativos en un mismo estudio, y orientados al estudio del mismo objetivo, y al que se ha dado incluso el rango de paradigma. Los *mixed methods* permiten posicionar la investigación en un continuo y admiten consecuentemente una cierta gradación, todavía no clarificada en su totalidad en la comunidad científica, pero sobre la cual existe un cierto consenso en entender que se apoya en la naturaleza del objetivo, tipo de datos, tipo de análisis, y tipo de inferencia. En este sentido, en la actualidad es posible cuantifi-

car el habla o el texto, desarrollando técnicas de obtención de datos que resultan sumamente útiles, así como aprovechando las inmensas posibilidades que nos brindan las aplicaciones informáticas. Para ello se requiere el seguimiento de un proceso que deberá canalizar la transformación del texto documental (o conducta oral transcrita, y demás tipos de datos) a matrices de códigos, que podrán ser tratadas cuantitativamente mediante técnicas estadísticas robustas, aunque no convencionales. Ello permitirá avanzar en estudios diacrónicos, de comparación de episodios, de casos múltiples, etc., que bascularán siempre en torno a un procedimiento riguroso y flexible a la vez.

Microanálisis de la comunicación: aplicaciones de la metodología observacional a la práctica profesional reflexiva. Paco Molinero¹

¹Universidad de Barcelona

La naturaleza interactiva y co-constructiva de los dispositivos de comunicación orientados a la facilitación de procesos de cambio encuentra en los métodos mixtos de la metodología observacional que empieza a ser aplicada en el desarrollo de un programa de investigación denominado "microanálisis" (Bavelas 2016). Las herramientas informáticas están llamadas a jugar un papel esencial en el desarrollo sistemático de este enfoque, el microanálisis, que el grupo de Bavelas ha llegado a definir como meta-método a muchas situaciones de la comunicación interpersonal. Las herramientas informáticas cumplen una función mediadora en el diálogo entre práctica profesional e investigación dando lugar a una práctica basada en la evidencia observacional, y una teoría desarrollada en diálogo continuado con los procesos de comunicación orientados al cambio. Este proceso de convergencia ha cobrado fuerza con los modelos colaborativos del cambio en los que la unidad de análisis es el lenguaje utilizado en la interacción (terapia focalizada en las soluciones y la entrevista motivacional). En este sentido, proponemos un rediseño de los procedimientos de supervisión y aprendizaje de las habilidades de intervención basadas en la comunicación en las que las herramientas informáticas ELAN y PRAAT juegan un papel muy importante para el registro y anotación / codificación de lo que ocurre en las sesiones atendiendo a las diferentes dimensiones de la comunicación (verbales y no verbales), y las dimensiones propias de los modelos teóricos. Para ello proponemos un procedimiento de trabajo en red con profesionales del cambio (psicoterapia, *coaching*, mediación..) que puedan incorporar estas herramientas en la mejora de su práctica profesional y de mejora de los modelos de intervención, esto es, profesionales-investigadores reflexivos que hagan realidad el lema *knowing by doing* y promuevan una práctica basada en la evidencia.

Equipos ganadores y perdedores en fútbol femenino: análisis de árbol de decisión. Rubén Maneiro¹, Claudio Casal², Toni Ardá³ y José Luis Losada⁴

¹Universidad Pontificia de Salamanca, ²Universidad Católica de Valencia, ³Universidad de A Coruña, ⁴Universidad de Barcelona

El análisis y estudio del fútbol femenino está experimentando un creciente apogeo de un tiempo a esta parte. Diversos trabajos han señalado la escasez de literatura científica asociada a esta modalidad de fútbol, mientras que los pocos estudios publicados han hecho referencia en su mayoría a aspectos fisiológicos, bio-

mecánicos o antropométricos. El objetivo principal de este estudio fue analizar cómo la variable *match status* o resultado parcial puede afectar al comportamiento táctico entre los equipos exitosos y no exitosos en el fútbol femenino de alto rendimiento. Para la obtención de la información, se ha llevado a cabo una observación sistemática sobre las posesiones de balón. Se han analizado 3.770 posesiones ofensivas durante el FIFA Womens World Cup 2015. Desde la perspectiva *mixed methods*, para identificar las interacciones se utilizó el análisis cuantitativo de árboles de decisión. Los resultados han permitido corroborar una serie de variables que modifican la conducta de los equipos exitosos VS los equipos no exitosos en función del resultado parcial del encuentro. Esto puede servir para proponer nuevos modelos de éxito en el fútbol femenino de más alto nivel, y por otro lado incrementar la producción científica sobre el fútbol femenino en este particular.

Un estudio de *mixed methods* en el modelo pedagógico de responsabilidad personal y social (MRPS). Oleguer Camerino^{1,2}, Alfonso Valero Valenzuela³, Queralt Prat¹, David Manzano Sánchez³ y Marta Castañer^{1,2}

¹INEFC, Universidad de Lleida, ²IRBLLEIDA, Universidad de Lleida, ³Universidad de Murcia

El presente estudio tiene como objetivo aplicar un programa basado en el "modelo de responsabilidad personal y social" (MRPS) de Hellison, tradicionalmente utilizado en educación física, a otras materias educativas. Pretendemos valorar la influencia del método en la responsabilidad, autonomía, motivación, autoconcepto y clima social del aula, comparando la conducta docente de tres profesores que aplican el MRPS con la de un experto en este modelo. Aplicamos la observación sistemática en 44 sesiones del primer semestre con un total de 120 estudiantes, entre 11 y 16 años ($M = 13,8$ años, $SD = 2,3$). Se contrastaron los resultados observacionales de la conducta docente, con *T-patterns* y coordenadas polares, de los 4 profesores con el comportamiento percibido de ellos y de los estudiantiles mediante el cuestionario TARE 2.0. El análisis *mixed methods* confirma cómo el MRPS mejoró los comportamientos docentes de los tres profesores en la formación comparado con el profesor experto y que este método influye positivamente en la participación del alumno en el desarrollo de la clase en contraste con el tratamiento tradicional, generando una percepción favorable en el alumno de sus necesidades psicológicas básicas, motivación hacia la E.F., deportividad y estilo de vida activo.

A partir de datos cualitativos, ¿se puede realizar un estudio diacrónico? M. Teresa Anguera¹, Gudberg K. Jonsson², Antonio Hernández Mendo³, Pedro Sánchez Algarra¹ y Mariona Portell⁴

¹Universidad de Barcelona, ²Iceland University, ³Universidad de Málaga, ⁴Universitat Autònoma de Barcelona

Hasta hace unos años, los datos cuantitativos se hallaban en el centro de una viva polémica entre quienes la ensalzaban y sus detractores. Se consideraba, por una parte, que permitían albergar una gran riqueza de información mediante registros de conducta, relatos, conversaciones, ..., pero, por otra, se los acusaba insistentemente de subjetividad y falta de consistencia. En la actualidad contamos ya con elementos que nos permiten adentrarnos en un profundo proceso de sistematización y transformación

de los datos cualitativos, a partir de diversas decisiones metodológicas (sobre dimensionalidad y criterios de segmentación en unidades) que conducirán a su transformación en matrices de códigos. Estas matrices de códigos pueden ser obtenidas y almacenadas diacrónicamente a lo largo de un proceso de recogida de datos, con la finalidad de escrutar su estructura desde diferentes perspectivas, que se corresponden con diferentes técnicas de análisis cuantitativo de datos cualitativos, y especialmente con detección de *T-Patterns*, análisis secuencial de retardos y análisis de coordenadas polares. Cada una de ellas tiene sus propios objetivos de análisis, que no excluyen que se puedan utilizar de forma complementaria entre sí. Las grandes posibilidades analíticas que comportan permiten fortalecer y empoderar enormemente el potencial de los datos cualitativos.

SIMPOSIO

Estrategias de integración de evidencias de validez para el desarrollo de instrumentos de funcionamiento

Coordina: Juana Gómez Benito

Resumen. La Organización Mundial de la Salud (OMS) conceptualiza la salud desde el modelo biopsicosocial y utiliza el término "funcionamiento" para representar el aspecto positivo y práctico de la salud: qué podemos o no podemos hacer en nuestro funcionamiento diario, más allá de tener una enfermedad u otra (World Health Organization, 2001). Dentro de este marco conceptual se elabora la "Clasificación Internacional del funcionamiento, de la discapacidad y de la salud" (CIF) que representa un marco integral para examinar el funcionamiento en personas con cualquier condición de salud. Dado que consta de más de 1.400 categorías, se han desarrollado "conjuntos básicos" (CBs) en su versión abreviada y extensa para condiciones de salud determinadas, con los indicadores que mejor describen dicha condición. El equipo de trabajo desarrolló un CB para la esquizofrenia y la primera aportación pretende situar el simposio en la secuencia de estudios que se han llevado a cabo hasta ahora para la elaboración de dicho CB siguiendo el proceso recomendado por la OMS. A partir de su elaboración, todos los CBs se han de validar para corroborar que los indicadores elegidos se adecuan a los objetivos específicos del CB desarrollado y este es el proceso en el que el equipo está inmerso en este momento y sobre el que pivotarán las demás aportaciones del simposio. La segunda aportación identifica los instrumentos de medida utilizados con mayor frecuencia en la investigación sobre funcionamiento en la esquizofrenia relacionándolos con la cobertura propuesta por el CB de esquizofrenia y llegando a la conclusión que no están diseñados específicamente para la esquizofrenia ni para evaluar funcionamiento, por lo que se constata la necesidad de desarrollar un instrumento que cumpla los requisitos adecuados. De ello se ocupa la tercera aportación que trata de relacionar los tres CBs elaborados en salud mental, encontrar los puntos comunes y proponer una primera versión de un instrumento para evaluar el funcionamiento en salud mental en el marco de la CIF. A continuación, con el fin de proporcionar mayor detalle entorno a las evidencias de validez del CB y a su posible generalización a nivel mundial, la cuarta aportación analiza los datos de diferentes estudios Delphi comparando la relevancia que los expertos de la salud de las seis regiones

geográficas propuestas por la OMS otorgan a las categorías de la CIF para describir el funcionamiento en la esquizofrenia. Complementando este estudio, la quinta aportación se propone valorar los factores que provocan el cambio de respuesta entre rondas favoreciendo el consenso entre los participantes. Finalmente en la sexta aportación se expone el estado actual de las evidencias de validez obtenidas hasta el momento, las que hay pendientes y futuros desarrollos como la utilización de grupos focales en línea. Todo este proceso aporta las bases necesarias para generar un sistema integral, multidisciplinar e internacional de evaluación de la funcionalidad en la esquizofrenia.

Conjunto básico de indicadores de funcionamiento para la esquizofrenia. M. Dolores Hidalgo¹, Juana Gómez-Benito^{2,3}, Nekane Balluerka⁴ y Arantxa Gorostiaga⁴

¹Universidad de Murcia, ²Universidad de Barcelona, ³Instituto de Neurociencias (UBNeuro), ⁴Universidad del País Vasco

El CB-CIF para la esquizofrenia se desarrolló siguiendo el proceso basado en la evidencia propuesto por la ICF Research Branch, centro colaborador de la OMS. Este proceso consistió en la realización de una primera fase preparatoria, que constó de cuatro estudios preliminares para seleccionar las categorías candidatas a formar parte del CB-CIF: un estudio empírico (Barrios, Gómez-Benito, Pino, Rojo y Guilera, 2018), una encuesta a expertos (Barrios, Guilera, Selb, y Gómez-Benito, 2017), un estudio cualitativo (Benítez, Pino, Padilla, y Cuevas-Parra, 2016) y una revisión sistemática de la literatura (Gorostiaga, Balluerka, Guilera, Aliri y Barrios, 2017). En la siguiente fase, se realizó un proceso de consenso y toma de decisiones en una conferencia internacional (Gómez-Benito et al., 2018) formada por profesionales de diferentes ámbitos de la salud expertos en el tratamiento de la esquizofrenia de las seis regiones del mundo de la OMS, que dio origen a la primera versión del CB-CIF para la esquizofrenia. El objetivo de esta presentación es precisamente describir todo el proceso metodológico de la elaboración de este CB-CIF desde una perspectiva multidisciplinar y multimétodo.

Funcionamiento en la esquizofrenia: revisión de instrumentos y evidencias de validez de contenido. Georgina Guilera^{1,2}, Maite Barrios^{1,2}, Barbara Braida¹ y M. Dolores Hidalgo³

¹Universidad de Barcelona, ²Instituto de Neurociencias (UBNeuro), ³Universidad de Murcia

Con este estudio se pretende: (a) identificar los instrumentos de medida utilizados con mayor frecuencia en la investigación sobre funcionamiento en la esquizofrenia; (b) analizar el contenido de estos instrumentos utilizando como marco de referencia el sistema de "Clasificación internacional del funcionamiento, de la discapacidad y de la salud" (CIF); y (c) comparar los indicadores extraídos de estos instrumentos con los cubiertos por los "conjuntos básicos" (CBs) para la esquizofrenia. Para ello, se realizó una revisión de los estudios sobre funcionamiento en la esquizofrenia utilizando las bases de datos Medline, CINAHL y PsycINFO. Los ítems de los 20 instrumentos más empleados se vincularon a categorías del sistema CIF. Los instrumentos más comúnmente utilizados fueron sobre todo de evaluación de síntomas clínicos (p. ej., la *Positive and Negative Syndrome Scale*), pruebas neuropsicológicas (p. ej., la *Wechsler Adult Intelligence Scale*), o de me-

dición del funcionamiento general (p. ej., la *Global Assessment of Functioning*). Al vincular los ítems a categorías CIF se obtuvieron 42 categorías: 25 de funciones corporales, 1 de estructuras corporales, 11 de actividades y participación y 5 de factores ambientales. Ningún instrumento por sí solo cubre el espectro de contenidos de los CIF-CBs. En su mayoría los instrumentos no están diseñados específicamente para la esquizofrenia o para evaluar el funcionamiento, sino que son pruebas genéricas estandarizadas para medir una amplia gama de variables en varias condiciones de salud. Los resultados constatan la necesidad de desarrollar un instrumento, aceptado internacionalmente, específico y completo de evaluación del funcionamiento en individuos con esquizofrenia.

Versión preliminar del "Cuestionario de valoración del funcionamiento en el trastorno mental severo": diseño y desarrollo. Juana Gómez Benito^{1,2}, Maite Barrios^{1,2}, Gomaa Said Mohamed Abdelhamid¹ y Oscar Pino^{1,3}

¹Universidad de Barcelona, (²) Instituto de Neurociencias (UBNeuro), ³Hospital General de Granollers Benito Menni

La "Clasificación internacional del funcionamiento, de la discapacidad y de la salud" (CIF) cuenta con más de 1.400 categorías, por lo que se requiere la formación de "conjuntos básicos" (CBs) a partir de la selección de las categorías relevantes para cada condición de salud. En la actualidad, en salud mental, se cuenta con los CBs de tres de los principales trastornos mentales: esquizofrenia, trastorno bipolar y depresión. El objetivo de este estudio es diseñar y desarrollar un instrumento en español e inglés para la valoración del funcionamiento de individuos que presenten un trastorno mental severo a partir de los CBs de la CIF. La primera fase consistió en identificar las categorías comunes en los tres CBs de la CIF. En una segunda fase se generaron ítems para cada una de las categorías comunes a los tres trastornos. Posteriormente se valoró la adecuación de los ítems por un panel de expertos. A partir de la comparación de los tres CBs se identificó un total de 34 categorías comunes a los tres trastornos: 10 para funciones corporales, 14 para actividades y participación y 10 para factores ambientales. La mayoría de los ítems generados a partir de estas categorías fueron confirmados por el panel de expertos. Este estudio ha permitido, por un lado, capturar los aspectos comunes del funcionamiento en la esquizofrenia, el trastorno bipolar y la depresión y, por otro, proponer una versión preliminar de un instrumento breve basado en la CIF para valorar el funcionamiento de los trastornos mentales severos.

Estudio cross-cultural del funcionamiento en la esquizofrenia. Laura Nuño^{1,2}, Emilio Rojo³, Maite Barrios^{2,4} y Juana Gómez Benito^{2,4}

¹Instituto Clínico de Neurociencias (ICN), ²Universidad de Barcelona, ³Hospital Benito Menni CASM, ⁴Instituto de Neurociencias (UBNeuro)

La "Clasificación internacional del funcionamiento, de la discapacidad y de la salud" (CIF) fue diseñada con la intención de ofrecer un sistema para describir el funcionamiento y la discapacidad con un lenguaje común y válido a nivel internacional. Los "conjuntos básicos" (CBs) para la esquizofrenia son una selección de categorías del sistema CIF específicas para el caso de la esquizofrenia. El objetivo de este trabajo es realizar un estudio cross-cultural donde se compara la relevancia que los expertos de la salud de las seis regiones geográficas propuestas por la OMS (i.e., África,

Américas, Asia Suroriental, Mediterráneo Oriental, Europa, y Pacífico Occidental) otorgan a las categorías de la CIF para describir el funcionamiento en la esquizofrenia. Para ello se analizan los datos de varios estudios Delphi donde se preguntó a un panel de expertos por los problemas que suelen encontrar en los pacientes con esquizofrenia. Sus respuestas se vincularon a categorías CIF y se seleccionaron aquellas que alcanzaron el consenso (?75%). Las categorías extraídas se compararon también con los CBs para la esquizofrenia. Un gran número de categorías fueron consideradas relevantes por los expertos en las seis regiones de la OMS. No obstante, algunas categorías CIF, especialmente aquellas pertenecientes al componente Actividades y participación y Factores ambientales, solo fueron consideradas relevantes en algunas de las regiones. En conclusión, aunque los expertos de las diferentes regiones de la OMS mayoritariamente aceptan las categorías CIF de los CBs, se observa que la relevancia de algunas categorías es mayor en algunas regiones frente a otras.

El consenso en el método Delphi: Análisis del cambio de respuesta. Maite Barrios^{1,2}, Núria Boixader¹, Georgina Guíllera^{1,2} y Laura Nuño^{2,3}

¹Universidad de Barcelona, ²Instituto de Neurociencias (UBNeuro) y ³Instituto Clínico de Neurociencias (ICN)

El método Delphi es una técnica grupal basada en la consulta a expertos que tiene como objetivo alcanzar el consenso entre sus participantes. El propósito del presente estudio es valorar los factores que provocan el cambio de respuesta entre rondas favoreciendo el consenso entre los participantes. En primer lugar, se valora el efecto de variables sociodemográficas y profesionales, y en segundo lugar, el efecto del acuerdo grupal en el cambio de respuesta individual. Se llevaron a cabo de manera independiente cinco estudios Delphi en tres rondas en los que participaron un total de 777 expertos en salud mental. Para cada participante se calculó la proporción de cambio de respuesta entre la segunda y tercera ronda teniendo en cuenta todas sus respuestas emitidas y se analizó el efecto de variables tales como la edad, sexo, profesión, años de experiencia profesional y región geográfica de procedencia. También se consideró el porcentaje de acuerdo obtenido en la segunda ronda para valorar el porcentaje de cambio en la tercera ronda. Ninguna variable sociodemográfica ni profesional tuvo un efecto significativo para explicar el cambio de respuesta. Sin embargo, al considerar el porcentaje de acuerdo obtenido en la segunda ronda y mostrado a modo de feedback en la tercera ronda, se observó que cuando este superaba el 75% se producía un cambio positivo (aumentaba el nivel de acuerdo entre los participantes), mientras que si era inferior al 75% el cambio de respuesta era negativo (se reducía el acuerdo entre los participantes). Este estudio evidencia el efecto del porcentaje de acuerdo del grupo a la hora de emitir una respuesta y alcanzar el consenso en el método Delphi.

Grupos focales en línea: una revisión sistemática. Paola Ruiz Celis¹, Annie Chai², Karina Campoverde³ y Juana Gómez Benito^{2,4}

¹Universidad Iberoamericana, Méjico, ²Universidad de Barcelona, ³Hospital General de Granollers Benito Menni, ⁴Instituto de Neurociencias (UBNeuro)

El uso de los grupos focales en línea (GFE) ha ido incrementando en la última década debido a que son una alternativa útil a los

grupos focales presenciales. Ante la falta de consenso sobre su uso y las características que deberían cumplir, se realizó una revisión sistemática de los estudios que utilizaron grupos focales en línea, sin restricciones temporales ni idiomáticas. Se utilizaron las bases de datos Web of Science, Ebsco y Scopus y la búsqueda identificó 2.234 artículos publicados. Los títulos y resúmenes fueron examinados para su elegibilidad según los criterios de inclusión establecidos y siguiendo la declaración PRISMA. Las características analizadas fueron: diseño de estudio, número de participantes, número de grupos; datos sociodemográficos de los participantes como edad, país y ciertos atributos particulares; la información del GFE a partir de la temporalidad (sincrónico-asincrónico), duración, el rol y características del moderador; el análisis de la información obtenida como el tipo de software utilizado, análisis, sistema de categorización y criterio de saturación; y el manejo ético y confidencial para la información recabada. A partir de los hallazgos encontrados en la revisión, se sugiere mayor claridad y especificación del uso, desarrollo y análisis de los GFE dentro de la investigación, a fin de promover adecuados criterios metodológicos entorno a las estrategias. Por tanto, se propone desarrollar una guía de reporte de información dentro de la investigación, análisis y reporte de resultados y el manejo ético y confidencial de la información obtenida por los participantes.

SIMPOSIO

Modelos longitudinales con variables latentes: aplicaciones en diferentes etapas del ciclo vital

Coordina: Alejandro de la Torre Luque

Resumen. Los estudios longitudinales se han aplicado en muy diversas disciplinas de las ciencias de la salud. Estos estudios pretenden analizar el efecto del tiempo (cambio intraindividual con el tiempo) en fenómenos humanos de diversa índole. Los modelos estadísticos con variables latentes proporcionan excelentes aproximaciones de análisis y predicción de cursos o trayectorias. Asimismo, se puede estudiar el efecto de predictores y/o covariables; o bien identificar trayectorias que subyacen al desarrollo global del proceso. En este simposio, se presentarán cinco aplicaciones de los modelos estadísticos con variables latentes para estudios longitudinales. En primer lugar, el Dr. Estrada introducirá los modelos de cambio latente en el desarrollo de la habilidad verbal infantil. En segundo lugar, el Dr. Ruiz presentará un estudio sobre el curso de la percepción de vulnerabilidad a la lesión en atletas. Los datos fueron modelados con curvas de crecimiento latente. Posteriormente, la Sra. Valeriano presentará una aplicación de estos mismos modelos al rendimiento cognitivo de personas mayores. En siguiente lugar, el Dr. De la Torre describirá la aplicación de uno de los modelos centrados en la persona (modelos mixtos de clase latente) en el estudio del curso de sintomatología depresiva en el envejecimiento, para identificar trayectorias (o cursos heterogéneos) de síntomas a medida que el individuo se hace mayor. Por último, el Sr. Javier de la Fuente introducirá aplicaciones complejas de los modelos de ecuaciones estructurales para modelar datos de alta dimensionalidad (datos genéticos). En conclusión, los modelos basados en variables latentes constituyen una alternativa analítica que cubre una amplia variedad de hipótesis de

investigación en estudios longitudinales. Asimismo, son bastante flexibles y aportan evidencia robusta.

Identificación de individuos con patrones atípicos de cambio latente. Eduardo Estrada^{1,2}, Emilio Ferrer², Bennett A. Shaywitz³, John M. Holahan³ y Sally E. Shaywitz³

¹Universidad Camilo José Cela, ²University of California, ³Yale University School of Medicine

En este trabajo presentamos un conjunto de técnicas estadísticas para identificar individuos dentro de una muestra que no cambian conforme al patrón típico del grupo. Cuando existen más de dos medidas repetidas, la aplicación de diversos modelos de cambio latente (*latent change score*, LCS) permite calcular la verosimilitud individual para cada caso bajo cada modelo y cuantificar cómo de atípica es cada trayectoria individual. Cuando existen solo dos ocasiones de medida, la aplicación de modelos LCS permite calcular los residuos del cambio latente a nivel individual para identificar cambios pre-post atípicos. Ilustramos el uso de estos métodos aplicándolos a una medida de habilidad verbal (WISC-R) tomada en una muestra amplia de individuos evaluados repetidamente entre los 6 y los 15 años. Basándonos en nuestros resultados, mostramos la eficacia de estas técnicas, ilustramos su utilidad en contextos aplicados, y discutimos posibles aplicaciones para el estudio del desarrollo psicológico.

Does sport injury risk perception change along the rehabilitation process? Víctor J. Rubio¹, Iván Sánchez-Iglesias², Miguel A. Ruiz¹, Constanza Pujals³, José Manuel Hernández¹

¹Universidad Autónoma de Madrid, ²Universidad Complutense de Madrid, ³Centro Universitario Ingá (Uningá)

Objectives: Perceived vulnerability (PV) refers to the extent to which an individual believes that s/he is subject to a health threat and it is supposed to contribute to increase motivation for active health-protective behaviors. The objective was assessing how PV of sport injury evolves at different moments related to the point in which an injury was sustained (before, at the moment of injury, when the athlete is discharged, six months later). **Design:** Within-subject design with four repeated measures of the PV Injury severity, self-control and perceived similarity were included as controlled variables. **Method:** One-hundred and twenty-four male non-professional soccer players (18-27 years, M = 23.4, SD = 2.1) participated in the study. A latent curve analysis (LCA) was used to study the pattern followed by PV over time. **Results:** The LCA model with both linear and quadratic terms yielded significant estimates ($p_0 = 6.62$), an upwards trend ($\eta_1 = 11.24$), and a Quadratic term ($\eta_2 = -11.42$). Only perceived similarity yielded significant effect on the model. The goodness of fit for the quadratic model was not adequate ($\chi^2 = 38.711$, $df = 3$, $p < 0.001$), and five different profiles were found in the sample. **Conclusions:** Experiencing an injury plays a role in the PV athletes reports, though with different profiles regarding how the PV evolves through time. However, in most of the cases such perception was progressively declining since the moment the athlete was injured which should be considered for promoting health-protective behaviors.

Estudio longitudinal del rendimiento cognitivo en el proceso de envejecimiento y su relación con la reserva cognitiva. Elizabeth Lucía Valeriano Lorenzo¹, Teodoro del Ser² y Miguel A. Fernández²

¹Universidad Complutense de Madrid, ²Centro de Investigación de Enfermedades Neurológicas - Fundación CIEN

Antecedentes: la discordancia frecuente entre patología cerebral y expresión clínica en los procesos neurodegenerativos se atribuye al contingente de recursos cognitivos acumulados tempranamente denominado *reserva cognitiva* (RC) (Stern Y., 2003). ¿Podría la RC temprana (RCt) afectar a la pendiente evolutiva del rendimiento cognitivo en los ancianos? Objetivos: (1) determinar la trayectoria de cambio en seis tareas cognitivas, durante 5 años de seguimiento y (2) analizar la influencia de la RCt en las trayectorias. Método: el Proyecto Vallecas es un estudio longitudinal con una muestra de 1.093 voluntarios con edades comprendidas entre 68-86 años, sin patología neurológica ni psiquiátrica relevante, para identificar factores predictores de deterioro cognitivo. Primero, se realizó un análisis factorial de RC de la cual se extrae la variable latente RCt. Segundo, se recogieron las puntuaciones de 6 parámetros cognitivos: estado cognitivo global (CG), memoria episódica inmediata (MEi) y diferida (MEd), memoria semántica, velocidad de procesamiento y viso-construcción, y empleando un modelo de curva latente se han estimado las trayectorias de evolución en cada una de ellas para el grupo de sujetos sanos (n=976) y para el grupo con deterioro cognitivo sobrevenido (n=117). Se realizó imputación múltiple de los valores perdidos. Resultados: las tareas CG, MEi y MEd muestran trayectorias de cambio significativo en ambos subgrupos. Para el grupo sano, la pendiente es creciente en tareas de MEi y MEd y decreciente en CG, y en el grupo con deterioro cognitivo el declive es más pronunciado en MEi y MEd. Además, la RCt es significativa y positiva con las pendientes de las tres tareas.

Trayectorias heterogéneas de síntomas de depresivos en población anciana: un estudio basado en modelos de crecimiento mixto latente. Alejandro de la Torre Luque¹, Javier de la Fuente¹ y José Luis Ayuso Mateos¹

¹Centro de Investigación Biomédica en Salud Mental (CIBERSAM), Universidad Autónoma de Madrid

La depresión es una condición psiquiátrica altamente prevalente en la edad anciana y asociada con importantes complicaciones de salud. Existen múltiples factores que pueden llevar a una escalada de sintomatología depresiva y a la instauración de condiciones de salud mental altamente incapacitantes. Los modelos mixtos de clase latente (MMCL) pueden contribuir en la identificación de trayectorias heterogéneas de desarrollo sintomatológico en esta etapa de la vida, dada su flexibilidad al no partir de la existencia de un curso unitario. Este estudio pretendía examinar la existencia de trayectorias diferenciales de síntomas de depresión en una muestra de 8317 adultos ingleses mayores de 65 años (53,10% mujeres; M = 73,87 años al inicio del estudio, DT = 6,53). Los participantes fueron evaluados bianualmente durante un periodo de 10 años. Se monitorizó sintomatología depresiva en este periodo. Se modeló el curso de la sintomatología usando MMCL, independientemente para hombres y mujeres. Como resultados, se identi-

ficaron tres trayectorias diferenciales de síntomas de depresivos para ambos sexos: una trayectoria mayoritaria con baja sintomatología, trayectoria subclínica de síntomas (en torno al 20% de participantes) y trayectoria de síntomas crónicos (en torno al 10% de participantes). Factores sociodemográficos como el nivel educativo y la historia de problemas psiquiátricos fueron asociados a trayectorias crónica y subclínica. Como conclusión, este estudio refleja la utilidad del uso de metodología centrada en la persona (MMCL) en la determinación de perfiles heterogéneos en el envejecimiento y en la búsqueda de alternativas personalizadas de abordaje preventivo y terapéutico.

Modelos de Ecuaciones Estructurales Genómicas: analizando la estructura genética multivariada de rasgos complejos. Javier de la Fuente¹

¹Centro de Investigación Biomédica en Salud Mental (CIBERSAM), Universidad Autónoma de Madrid

Los modelos de ecuaciones estructurales genómicas permiten modelar la arquitectura genética común a un conjunto de rasgos, utilizando estadísticos derivados de estudios de asociación del genoma completo (GWAS). Estos modelos permiten modelar la matriz de varianzas-covarianzas entre polimorfismos de nucleótido único (SNPs) asociados a fenotipos complejos. Así, es posible tanto la identificación de factores latentes responsables de la arquitectura genética multivariada de un conjunto de rasgos, como el modelado de la relación longitudinal entre los mismos. Por otro lado, esta aproximación permite testar la heterogeneidad de locus específicos, posibilitando así la identificación de variantes genéticas operando a distintos niveles de especificidad.

SIMPOSIO

Metodología mixta de investigación: desafíos, propuestas y beneficios de la integración de métodos cuantitativos y cualitativos

Coordinan: Isabel Benitez Baena y José Luis Padilla García

Resumen: La metodología mixta de investigación ha experimentado un renovado impulso en la última década. Parte de este impulso puede explicarse por la necesidad de abordar problemas de investigación cada vez más complejos en el ámbito de las ciencias sociales y de la salud. Entre los problemas de creciente complejidad se pueden apuntar: la diversidad de intereses y de agentes implicados en la investigación, y los nuevos modos y contextos para la obtención de datos (e. g., administración de cuestionarios por Internet). La metodología se enfrenta a renovadas y nuevas cuestiones como las bases epistemológicas de la metodología mixta, los fundamentos conceptuales, la actualización de los criterios de validez y replicabilidad de los estudios mixtos, así como los de la difusión de sus resultados en los medios tradicionales (conferencias, manuales y revistas científicas), y los emergentes (redes sociales). El simposio agrupa un conjunto de comunicaciones dirigidas a plantear, proponer y estimular el debate metodológico sobre los desafíos y beneficios de la metodología mixta. Los ponentes harán énfasis en las características que diferencian

sus estudios mixtos, y cómo han abordado la integración de metodologías cuantitativas y cualitativas en las diferentes fases de la investigación. Las comunicaciones abordan desde las bases epistemológicas de la metodología mixta, el estudio de problemas psicosociales como los dilemas morales integrando métodos cuantitativos y cualitativos, la elaboración de protocolos mixtos de evaluación de las necesidades formativas de los estudiantes, la integración de métodos cuantitativos y cualitativos en la metodología observacional, hasta los retos de la comunicación de resultados integrando las respuestas a preguntas cerradas y abiertas junto a "paradatos" de las encuestas por Internet.

Medición psicológica, métodos cualitativos y reflexividad.

Ana R. Delgado¹

¹Universidad de Salamanca

Un área de las ciencias del comportamiento en la que los métodos cuantitativos han obtenido un éxito incuestionable es el de la medición psicológica; sin embargo, son muchas las decisiones que deben tomarse subjetivamente, de manera convencional en el mejor de los casos, sobre cuestiones propiamente cualitativas tales como los contenidos de las pruebas. Por otra parte, es común que la investigación en ciencias sociales y de la salud parta de conceptos borrosos, de procedencia teórica diversa y solapada entre sí, por lo que se requiere contar con métodos que permitan obtener información de interés sin imponer a priori una estructura cuantitativa, no importa lo consensuada que esta sea. Consideramos que una adecuada combinación de métodos cualitativos y cuantitativos (p. ej., fenomenología descriptiva + modelo de escalas de calificación en el estudio de la personalidad; análisis temático + modelo de Rasch en el estudio de problemas sociales; o análisis de contenido + modelo de crédito parcial en áreas educativas) puede ayudar a resolver los problemas planteados, siempre y cuando se tenga presente que el uso de los métodos cualitativos exige atender a los supuestos ontológicos y epistemológicos: la reflexividad es parte integral del proceso de investigación.

Una aproximación mixta al estudio de dilemas morales (Neurodile). Pilar Aguilar¹, Alejandro Galvao¹ e Isabel Benítez¹

¹Universidad Loyola Andalucía

La Filosofía ha mantenido durante siglos un gran debate en torno a la toma de decisiones morales. Desde una perspectiva filosófica, dos han sido las opciones más recurrentes a la hora de guiar las decisiones morales: la perspectiva deontológica centrada en el cumplimiento de las normas morales y la consecuencialista, que prioriza la consideración de las consecuencias de las decisiones. Recientemente, la psicología ha adoptado este mostrando un interés creciente por los procesos psicológicos que subyacen a estas decisiones morales, siendo, principalmente, analizadas desde un paradigma experimental. Estos estudios previos han mostrado las diferencias que existen en las decisiones que toman las personas, pero aún no se ha profundizado en el origen de dichas diferencias. El objetivo de este estudio es investigar los procesos de respuesta asociados a la toma de decisiones desarrolladas ante dilemas morales e indagar los motivos que llevan a las personas a tomar decisiones deontológicas o consecuencialistas.

Para ello, se ha diseñado un estudio mixto con dos fases. En la primera fase, los participantes responden a 45 dilemas morales mientras se registran datos psicofisiológicos. En la segunda fase, los participantes participan en una entrevista cognitiva donde se recoge información sobre los elementos utilizados para tomar decisiones ante los dilemas morales. Se describirán algunos resultados obtenidos en cada una de las fases y se presentará la propuesta de integración. Se discutirán las limitaciones del estudio y los retos ocurridos en el desarrollo del protocolo de entrevista y otras fases del proyecto.

Evaluación de variables asociadas al rendimiento académico de estudiantes universitarios: un modelo de convergencia.

Juan F. Luesia¹, Isabel Benítez¹ y Milagrosa Sánchez-Martín¹

¹Universidad Loyola Andalucía

El rendimiento académico universitario puede estar determinado por distintas variables que han sido ampliamente estudiadas desde distintos enfoques como, por ejemplo, el de la evaluación de la inteligencia basada en competencias cognitivas. Sin embargo, es difícil encontrar un modelo comprensivo de competencias capaz de predecir cómo será el rendimiento académico de los estudiantes una vez comiencen sus estudios universitarios. Hasta ahora, el análisis de estas variables se ha basado únicamente en datos cuantitativos. Sin embargo, integrar estos datos con otros de naturaleza cualitativa podría enriquecer la interpretación de los resultados y optimizar el proceso de evaluación. El objetivo de este estudio es proponer un diseño mixto para desarrollar un protocolo de evaluación donde se recoja información relevante que permita identificar las necesidades de los estudiantes universitarios. Para ello, se planteará la integración de datos procedentes de una revisión sistemática y datos recogidos mediante entrevistas y grupos focales. Los resultados reflejarán un modelo de convergencia que incluya las competencias más relevantes para los estudiantes universitarios. Se discutirán las limitaciones del modelo y se enfatizarán los beneficios de la integración frente a la interpretación aislada de cada una de las fuentes de información.

El *quantitizing* como planteamiento innovador de la integración cualitativo-cuantitativo en observación sistemática a partir del orden o secuencia. M. Teresa Anguera¹

¹Universidad de Barcelona

En los últimos años se ha desatado una interesante polémica acerca las vías de integración entre elementos cualitativos y cuantitativos bajo la perspectiva *mixed methods*. Se han evidenciado diferentes posiciones, así como discusiones acerca del grado de integración que se puede lograr. Desde la observación sistemática, que es una metodología que puede considerarse como *mixed method* en sí misma, partimos de las vías totalmente consolidadas propuestas por autores pioneros (Cresswell y Plano Clark), como *merging*, *connectingy embedding*. Entre ellas, el *connecting* permite la alternancia de etapas QUAL-QUAN-QUAL, que legitima el planteamiento genérico de los *mixed methods*, en tanto que se logra una integración total entre elementos cualitativos y cuantitativos. Si consultamos la abundante documentación empírica existente en la actualidad, nos percatamos de que en su gran mayoría

se aplican solamente técnicas analíticas convencionales (estadística descriptiva, correlaciones ordinales, aplicación de la teoría de la generalizabilidad, análisis de correspondencias, análisis de conglomerados, análisis de varianza, ...). Prestigiosos autores de *mixed methods* se han referido conceptualmente, y también técnicamente, a la cuantificación y a todo lo que implica la transformación de los datos. Pero una gran diferencia de nuestra propuesta a partir de la metodología observacional, a favor de ésta, es que la cuantificación es más robusta, ya que no se apoya en el recuento de ocurrencias de conducta, es decir, en la frecuencia, sino en los parámetros primarios de orden o secuencia. A partir de ellos es posible desarrollar análisis cuantitativos robustos, como análisis secuencial de retardos, análisis de coordenadas polares, y detección de *T-Patterns*, que permite desvelar la estructura de los datos registrados desde diferentes perspectivas, y en los últimos años hemos comprobado su creciente aplicabilidad, uso y resultados obtenidos, tanto individualmente como de forma complementaria.

Integrando evidencias cualitativas sobre los procesos de respuesta y para datos para extender el modelo cognitivo del proceso de pregunta-y-respuesta en encuestas por Internet. Dörte Naber¹ y José Luis Padilla García¹

¹Universidad de Granada

El creciente recurso a la administración de encuestas por Internet plantea un reto al modelo cognitivo tradicional del proceso de pregunta-y-respuesta (Tourangeau, 1984, 2018). Al mismo tiempo, el desarrollo actual de las aplicaciones para implementar cuestionarios web permite obtener evidencias cualitativas (webprobing), y cuantitativas en diversas modalidades de "paradatos": tiempo de respuesta, patrones de teclado, movimientos del cursor, etc. El objetivo del estudio es mostrar cómo se pueden integrar evidencias cualitativas y cuantitativas para evaluar el funcionamiento de preguntas en cuestionarios web. Presentaremos resultados preliminares del análisis de evidencias cualitativas, resultados de análisis psicométricos y "paradatos" de 1.000 participantes (500 en Alemania y 500 en España), a preguntas individuales y preguntas multi-ítem sobre bienestar de la Ronda 8ª de la Encuesta Social Europea, dentro de un diseño de métodos mixtos. En esta presentación, nos centraremos en la integración de resultados y cómo dicha integración aporta evidencias sobre la "ecología" del proceso cognitivo de pregunta-y-respuesta en cuestionarios web. También discutiremos cómo la integración de resultados de diferentes métodos permite incrementar la validez de los estudios transculturales.

SIMPOSIO

Aplicaciones metodológicas en la investigación sobre altas capacidades intelectuales

Coordina: África Borges del Rosal

Resumen: Las "altas capacidades intelectuales" es un tema de creciente interés, debido a las necesidades educativas de este alumnado. Los problemas más importantes se centran en la presencia de mitos y conceptos erróneos, por lo que la determinación

del concepto resulta fundamental. Por ello, resulta fundamental clarificar las características que definen a este alumnado, lo que redundará en una adecuada intervención educativa, tanto en las instituciones educativas como teniendo en cuenta su entorno. En el presente simposio se toman en cuenta estos aspectos.

Elaboración de un cuestionario sobre el conocimiento en altas capacidades intelectuales. Josué Pérez¹, Leire Aperribai², Leire Garamendi² y África Borges¹

¹Universidad de La Laguna, ²Escuela Universitaria de Magisterio Begoñako Andra Mari, BAM

Introducción: el conocimiento adecuado sobre los alumnos superdotados por parte de la población resulta complicado en la medida en que existe un intenso debate a nivel científico entre numerosas definiciones y propuestas teóricas diferentes que pueden favorecer la persistencia de mitos y estereotipos sobre este fenómeno, generando confusión. Los mitos sobre "altas capacidades intelectuales" son habituales e influyen en la percepción, el diagnóstico y la intervención educativa sobre estos alumnos. Objetivo: el presente estudio pretende elaborar un cuestionario para evaluar el conocimiento sobre los alumnos con altas capacidades mediante los diferentes mitos y estereotipos asociados a la superdotación. Método: el cuestionario de "Conocimientos en altas capacidades intelectuales" se creó a partir de la revisión de un cuestionario de mitos sobre la superdotación previo. Se llevó a cabo un proceso de adecuación del instrumento por medio de la revisión de la literatura y otros instrumentos similares. Se trata de un cuestionario breve y rápido de aplicar que consta de 38 ítems con respuesta tipo Likert de 4 puntos sobre el grado de acuerdo con cada afirmación presentada. Se aplicó a un total de 313 personas de distintos colectivos pertenecientes a varias comunidades autónomas. Resultados: los análisis de fiabilidad mediante alfa de Cronbach devuelven valores altos. Las correlaciones ítem-test muestran ser significativas y adecuadas, salvo en el caso de 4 ítems, que serán eliminados para la validación del instrumento en un siguiente estudio. Conclusiones: el cuestionario "Conocimientos en altas capacidades intelectuales" se muestra como adecuado y fiable para su aplicación y validación como un nuevo instrumento que evalúa los conocimientos previos o representaciones que se tienen sobre el alumnado superdotado, pudiendo identificarse, de esta manera, errores, mitos o estereotipos existentes sobre éstos.

Diferenciación del talento en la memoria de trabajo: variables relevantes en la distinción entre el alumnado más dotado y sus compañeros normativos. E. Verche¹, E. Rodríguez Naveiras², P. Hernández Lastiri³, R. Montero³ y A. Borges³

¹Universidad Europea de Madrid, ²Universidad Europea de Canarias, ³Universidad de La Laguna

Los estudios referentes al alumnado de altas capacidades señalan diferencias cognitivas con sus compañeros de menor capacidad intelectual. Esto es así cuando se trata de funciones ejecutivas, una de las cuales es la memoria de trabajo. Un reciente meta-análisis puso de manifiesto una mejor memoria de trabajo entre alumnado con distintas denominaciones de talento y el grupo normativo, especialmente en memoria de trabajo verbal. Sin embar-

go, en el análisis de variables moderadoras, considerando como tal el tipo de talento y la forma en que fue medido, no aparecen diferencias significativas. Esto resulta especialmente relevante, porque, por una parte, sería preciso analizar los procedimientos evaluativos que determinan la presencia de alta capacidad intelectual, determinando el más relevante y, por otro, que las distintas etiquetas de talento no son importantes en la diferenciación de una mejor memoria de trabajo.

La comprensión lectora ¿está mediada por el formato utilizado? Daniel Valbuena¹ y África Borges¹

¹Universidad de La Laguna

Dentro del proceso lector, la comprensión lectora es la habilidad llave para adquisición de conocimiento y entendimiento del propio texto. Con el avance de las nuevas tecnologías en el ámbito educativo es relevante estudiar el funcionamiento diferencial del medio utilizado en la lectura: el papel tradicional, o las tabletas electrónicas. Los estudios que analizan los efectos de las nuevas tecnologías para el proceso lector miden la eficacia autopercebida, la fluidez lectora y la comprensión lectora en población normal, obteniendo resultados de mejoría con el uso de nuevas tecnologías. Por otra parte, los estudios realizados con alumnado de altas capacidades señalan los beneficios para este colectivo tienen las nuevas tecnologías en el proceso de enseñanza aprendizaje. Para poner a prueba el funcionamiento diferencial del proceso de lectura, con una muestra de 19 estudiantes de primaria con diagnóstico de alta capacidad, y un diseño de medidas repetidas, contrabalanceado en cada aplicación de la prueba, donde en una de ellas se presentaba un texto adaptado a la edad de cada participante en papel y en la segunda aplicación, con tres semanas de distancia, en el otro formato. Se presentan los resultados obtenidos, discutiendo la relevancia que tiene la elección de formato para el proceso de aprendizaje.

Comportamientos personales y familiares percibidos entre escolares con altas capacidades y sus madres. Gabriela López Aymes¹, Ilse Giovanna Tlalmis Gama¹ y Gabriel Reyes Valenzuela¹

¹Universidad Autónoma del Estado de Morelos

En este trabajo se analizan las percepciones de escolares con altas capacidades sobre su comportamiento y el de sus madres, así como el de las madres sobre sus hijos/as que participaron en un programa de enriquecimiento extracurricular. En el estudio participaron 20 menores y sus madres que provienen de familias de clase media del municipio de Cuernavaca, Morelos (México). La información se recabó mediante una entrevista semiestructurada. Los resultados muestran que los niños se perciben responsables respecto a sus deberes escolares, sensibles hacia los demás; sin embargo no les resulta fácil hacer amigos y son algo impacientes. Se sienten queridos y escuchados por sus madres, pero reciben castigos cuando no se portan bien. Por otro lado, las madres perciben un buen ajuste en su rol educativo y muestran prácticas de crianza positivas que les permiten enfrentar los desafíos familiares; así mismo consideran que sus hijos emplean estrategias de socialización y comunicación asertiva, sin embargo de manera

general perciben que sus hijos discuten fácilmente con los demás y les molesta no conseguir lo que quieren. Se analizan las repercusiones para la orientación educativa.

Programa de mentoría para alumnado universitario con altas capacidades, ATENEA-ULL. Evaluación de conferencias plenarias. África Borges del Rosal¹, Manuela Rodríguez-Dorta¹, Triana Aguirre Delgado¹, José Miguel Pérez Álvarez¹ y Pablo Rodríguez Flores¹

¹Universidad de La Laguna

Los programas dedicados al alumnado de altas capacidades están, por lo general, orientados a alumnado preuniversitario. Por este motivo, el Vicerrectorado de Estudiantes de la Universidad de La Laguna (ULL) ha impulsado un programa pionero cuyo objetivo es dar respuesta a las necesidades educativas específicas del alumnado con altas capacidades intelectuales en la etapa universitaria. El Programa de Mentoría Atenea-ULL consta de dos tipos de actividades impartidas por docentes universitarios. Por un lado, las actividades específicas que pueden ir dirigidas a alumnado de un grado en concreto, de varios afines o estar abiertas a todo el alumnado. Con ellas se brinda la posibilidad de conocer aspectos de la investigación y de la realidad profesional de su área del conocimiento y/o de otras, que por lo general no se abordan en las clases ordinarias, así como acercarlos al profesorado experto y que entre ellos puedan compartir opiniones, ideas e intereses, creando así una red social de conocimiento intergeneracional, además de brindarles la oportunidad de iniciarse en la labor investigadora. Por otro, incluye conferencias plenarias que ofrecen una visión general sobre diversas temáticas de interés. Ambas actividades son evaluadas por el alumnado asistente, a través de un formulario de evaluación online creado para cada actividad que se va desarrollando en el transcurso del programa. Se presentan aquí los resultados obtenidos en las evaluaciones de las conferencias plenarias. Se pone de manifiesto un alto grado de satisfacción por parte del alumnado, quien señala las distintas conferencias impartidas como interesantes, entretenidas, dinámicas, etc.

SIMPOSIO

Metodología cualitativa en la investigación de altas capacidades

Coordina: María Cadenas Borges

Resumen. La metodología cualitativa ha ido adquiriendo cada vez más importancia para el estudio de las ciencias del comportamiento. No obstante, cuando los análisis se hacen siendo extraídas las categorías resultantes por el propio investigador, está sometido a subjetividad. Por ello, ganan cada vez más relevancia los análisis de datos textuales mediante estadística, que permiten extraer la información esencial incluida en un texto, con la ventaja añadida de que es posible incluir muestras de participantes mayores que en los estudios cualitativos tradicionales, sin un gran coste de tiempo y esfuerzo. Se presentan aquí varios estudios con muestras de alumnado de altas capacidades o sus familias o

profesorado, usando dos softwares que analizan datos textuales: ALCESTE e IRAMUTEQ. Se señalan las similitudes y diferencias de ambos procedimientos.

Concepciones de los progenitores de alumnado de altas capacidades sobre lo que ello significa. María Cadenas¹, África Borges², Rubens Montero², Elena Rodríguez Naveiras¹ y Lidia García González¹

¹Universidad Europea de Canarias, ²Universidad de La Laguna

Las diferentes concepciones que existen acerca del alumnado con altas capacidades intelectuales y la existencia de mitos asentados en buena parte de la población sin formación en el tema, incluidos los progenitores, hace que el estudio de las características definitorias y la disipación de las falsas creencias sobre las altas capacidades sea una tarea constante para la investigación en este campo. Debido a la importancia de una detección temprana y una premura en la intervención, una gran parte de este alumnado es menor, siendo relevante la intervención conjunta menor/progenitor. Puesto que es importante conocer las creencias de los progenitores acerca de las altas capacidades y las preferencias y expectativas en cuanto a programas de intervención, este estudio analiza desde una perspectiva cualitativa las creencias que tienen los progenitores con uno o más hijos e hijas con altas capacidades intelectuales. La muestra está compuesta por progenitores de alumnado de altas capacidades, 28 hombres y 210 mujeres, con una media de edad de 42 años, residentes en España, Argentina y México, que respondieron a la pregunta abierta ¿Qué entiende Usted por alta capacidad intelectual? El análisis de los datos se ha hecho mediante el software ALCESTE 2018 creado en Francia en 1986 por la sociedad IMAGE, que describe, clasifica y sintetiza un texto. Se obtienen tres clases, que explican el 81% del corpus: Forma Diferente de Ser y aprender, Destacar en Diversas Áreas, Cociente Intelectual Superior, por lo que se concluye que la concepción de los progenitores, aunque puede diferir como se muestra en el agrupamiento obtenido, asimila alta capacidad a un funcionamiento cognitivo superior.

¿Qué son altas capacidades? Concepciones del alumnado universitario de altas capacidades y de sus profesores. Triana Aguirre¹, África Borges¹, María Sillió¹, Alvaro Martín¹, Manuela Rodríguez-Dorta¹ y Rubens Montero¹

¹Universidad de La Laguna

No hay consenso en lo que se refiere a qué se entiende por altas capacidades. Por ello, es importante conocer la idea que el alumnado universitario de altas capacidades tiene sobre el particular, al igual que su profesorado. En este estudio, se solicitó a los participantes de un programa pionero para alumnado de altas capacidades (profesorado y alumnado) implementado por la Universidad de La Laguna, que respondieran a una pregunta abierta sobre la idea que tenían sobre la conceptualización de altas capacidades, respondiendo a ello ambos colectivos. Las respuestas fueron analizadas con el programa ALCESTE, dando en cada grupo dos clases, que tienen en común la consideración de habilidades la capacidad intelectual superior. Se comentan las diferencias existentes entre las clases obtenidas en cada colectivo.

Papel de los progenitores en la detección del alumnado de altas capacidades. Elena Rodríguez-Naveiras¹, Julián Bethancort², M. Dolores Valadez², África Borges³ y Rubens Montero³

¹Universidad de La Laguna, ²Universidad de Guadalajara (México), ³Universidad de La Laguna

Los progenitores del alumnado de altas capacidades juegan un papel fundamental en la detección inicial, que da lugar a los pasos posteriores, esto es, el diagnóstico definitivo que confirma sus altas capacidades, dando lugar a que puedan recibir la adecuada respuesta educativa. En el presente estudio se analiza mediante ALCESTE la respuesta que dieron los progenitores a una pregunta abierta sobre qué les hizo suponer que sus hijos/as tenían altas capacidades, en el entorno de un proceso de selección de alumnado con CI de 130 o superior, para su posterior inclusión en una escuela agrupada por capacidad. El ALCESTE obtenido arroja tres clases. Se comentan los resultados, señalando la importancia de una adecuada idea que los progenitores posean sobre lo que significa tener capacidad superior para la adecuada detección de este alumnado.

Las relaciones del alumnado en el contexto del aula. Raquel Matías¹, África Borges¹ y Grecia Ortiz²

¹Universidad de La Laguna, ²Universidad de Guadalajara (México)

Uno de los temas que sigue suponiendo un importante debate en altas capacidades es la relación que este alumnado tiene con compañeros y profesorado en el contexto escolar, llegando incluso a considerarse esta población como tendente a poder presentar acoso. Dada la relevancia del tema, se ha llevado a cabo un estudio, con 37 estudiantes de primaria diagnosticados de altas capacidades, a los que se les preguntó sus relaciones con pares y profesorado a través de una encuesta semiestructurada. Los resultados fueron analizados con un software que estudia datos contextuales, IRAMUTEQ. Los resultados muestran adecuadas relaciones con su entorno, lo que lleva a plantear las razones del debate de las dificultades relacionales de este alumnado.

SIMPOSIO

Meta-análisis: avances metodológicos

Coordina: Julio Sánchez Meca

Resumen. Aunque el meta-análisis es una metodología de investigación consolidada, su aplicación requiere de mejoras para garantizar la validez de las conclusiones de los meta-análisis. En este simposio se presentan cuatro investigaciones en las que se abordan, o dan respuesta, a diversos aspectos y problemas que tienen que ver con esta metodología. En la primera comunicación Josep Maria Losilla presenta la herramienta ROBS para ser aplicada en el contexto de meta-análisis que integran estudios no experimentales (cohortes, casos y controles y transversales). La segunda comunicación, presentada por Carme Viladrich, compara y discute la calidad de las herramientas para valorar el riesgo de sesgo en los estudios primarios que utilizan escalas de calidad de

vida, y su repercusión en los meta-análisis de esos estudios primarios. En la tercera comunicación Dolores Frías presenta una panorámica del meta-análisis en red y de su utilidad para realizar comparaciones entre tratamientos alternativos que no se han comparado de forma directa en los estudios primarios. Finalmente, la cuarta comunicación corre a cargo de Susana Sanduvete, en la que presenta los resultados de un meta-análisis de alta calidad dirigido a investigar el efecto que el contexto familiar ejerce en el desarrollo del lenguaje de niños con implante coclear, enfatizando los aspectos y dificultades metodológicos encontrados en dicho meta-análisis.

ROBS (“Risk of bias of Observational Studies”). Josep Maria Losilla¹, Isabel Oliveras¹, Jaume Vives¹, Emelina López², Juan A. Marín García³ y María F. Rodrigo²

¹Universitat Autònoma de Barcelona, ²Universitat de València, ³Universitat Politècnica de València

La evaluación del riesgo de sesgo (RdS) de los estudios primarios que se incluyen en revisiones sistemáticas (RS) y meta-análisis (MA) es un aspecto central en el proceso de síntesis de los resultados de investigación. En esta comunicación presentamos la herramienta ROBS (“Risk of bias of Observational Studies”), que hemos desarrollado partiendo del conocimiento del equipo de investigación en el desarrollo de la herramienta Q-Coh, así como de los resultados de una exhaustiva meta-revisión original de las RS y MA de estudios no experimentales publicadas en la última década en el ámbito de la psicología sanitaria. ROBS permite evaluar el RdS de estudios no experimentales con diseños de cohortes, de casos y controles y transversales, es relativamente rápido de administrar y ofrece unos buenos índices de fiabilidad entre jueces y de validez concurrente y discriminante. La orientación de ROBS difiere de las herramientas propuestas hasta la fecha en tres aspectos relevantes: (a) se centra en la evaluación de la relación entre exposiciones y respuestas en lugar de en el diseño del estudio; (b) aún en una única herramienta la evaluación del RdS y de la calidad del reporte, para ayudar a juzgar si la publicación incluye toda la información necesaria para valorar el RdS; y (c) requiere justificar la evaluación de los potenciales sesgos, para facilitar el análisis posterior de los resultados del estudio en el contexto de RS y/o MA, especialmente en relación con la valoración de la heterogeneidad y de la credibilidad de los resultados de la síntesis.

El rol del uso previsto de los cuestionarios de calidad de vida en la evaluación de sus propiedades de medida: una meta-revisión. Carme Viladrich¹, Sonia Lorente¹, Josep María Losilla¹ y Jaume Vives¹

¹Universitat Autònoma de Barcelona

Las revisiones sistemáticas (RS) y meta-análisis (MA) sobre instrumentos de medida en el ámbito sanitario deben evaluar tanto el riesgo de sesgo (RoB) de los estudios primarios como las propiedades de medida de los instrumentos. Aunque existen múltiples propuestas para llevar a cabo dichas evaluaciones, a menudo éstas no toman en consideración el uso previsto de los cuestionarios analizados (i.e., el argumento interpretativo, que puede

ser, por ejemplo, descriptivo, evaluativo o predictivo) a la hora de seleccionar las evidencias y propiedades específicas que se deben valorar. La presente meta-revisión analiza las RS y MA sobre instrumentos de evaluación de la calidad de vida (QoL) publicadas en la última década, con los siguientes objetivos: (a) identificar los instrumentos de evaluación del RoB y los estándares e instrumentos de evaluación de las propiedades de medida utilizados; (b) comparar las características de los estándares e instrumentos identificados; (c) describir la relación que los autores de las RS y MA establecen entre la evaluación de las propiedades de medida y el/los uso/s previsto/s de los cuestionarios sobre QoL. Las guías Medical Outcomes Trust y COSMIN son las más utilizadas en las RS y MA analizadas para evaluar las propiedades psicométricas de los cuestionarios. Nuestros resultados también indican que los autores de las RS y MA mayoritariamente no tienen en cuenta el RoB de los estudios primarios, ni tampoco el uso previsto de los cuestionarios sobre QoL cuando evalúan sus propiedades de medida o cuando realizan recomendaciones con base en dicha evaluación.

Meta-análisis en red. Dolores Frías Navarro¹, Marcos Pascual Soler², Franco Lacedonia¹ y Olaya Molina Palomero²

¹Universidad de Valencia, ²ESIC Business & Marketing School of Valencia

La acumulación de evidencia sobre múltiples tratamientos para un mismo problema ha generado actualmente una red de intervenciones y comparaciones que se puede representar y analizar a través de un estudio de meta-análisis en red. El meta-análisis en red (también conocido como meta-análisis de múltiples tratamientos o meta-análisis de comparaciones mixtas) es una técnica estadística que se ha desarrollado como una extensión del meta-análisis tradicional con la ventaja de que permite integrar comparaciones indirectas entre dos tratamientos que no han llegado a realizarse en ningún estudio a través de un comparador común (cuando no se dispone de información o la información directa es de baja calidad). La técnica combina diferentes estudios sobre un mismo tema o pregunta de investigación para sintetizar simultáneamente el efecto de más de dos tratamientos (red de tratamientos) que se comparan entre sí respecto a un comparador común, llevando a cabo comparaciones directas (a partir de estudios que directamente comparan intervenciones), indirectas y mixtas (combinan la estimación directa e indirecta de los efectos en un único efecto del tratamiento). Además, permite calcular los efectos relativos de todos los tratamientos incluidos en la red, facilitando una jerarquía de las intervenciones basada en la probabilidad que tiene cada tratamiento de ser el más o el menos eficaz. El meta-análisis en red se ha convertido en una fuente de evidencia fundamental entre los investigadores y los desarrolladores de guías y agencias de salud (desarrollo de fármacos, por ejemplo) y es una herramienta que la psicología va incorporando en sus estudios. El objetivo de nuestro trabajo es presentar la técnica del meta-análisis en red y los fundamentos de su análisis, su evolución, valorar su utilidad en el campo de la psicología y destacar sus ventajas y limitaciones metodológicas.

El Impacto del contexto familiar en el desarrollo del lenguaje en niños con implante coclear: una revisión sistemática y meta-análisis. Susana Sanduvete Chaves⁴, Salvador Chacon Moscoso^{4,6}, Daniel Holzinger^{1,2,3}, Magdalena Dall^{1,2}, David Saldaña⁵ y Johannes Fellinger^{1,2,7}

¹Konventhospital Barmherzige Brüder, ²Johannes Kepler University Linz, ³University of Graz, ⁴Universidad de Sevilla, ⁵Laboratorio de Diversidad (Universidad de Sevilla), ⁶Universidad Autónoma de Chile, ⁷Medical University of Vienna

Objetivo: realizar una revisión sistemática y meta-análisis para evaluar la influencia del entorno familiar en el desarrollo del lenguaje en niños con implantes cocleares. **Diseño:** las búsquedas en las bases de datos describían: entorno familiar, características del niño y desarrollo del lenguaje. Los estudios tenían que: evaluar las asociaciones entre variables familiares y variables que afectaban directamente al niño con los resultados lingüísticos, estudiar niños implantados antes de los 5 años, medir el lenguaje con instrumentos estandarizados, publicarse entre 1995 y febrero de 2018 y haber sido revisados por pares. La calidad metodológica se evaluó adaptando una checklist validada. **Resultados:** se estudiaron 27 publicaciones de diferente calidad metodológica, encontrándose un efecto alto para la calidad y la cantidad del input lingüístico parental en los primeros años, $r = 0,563$, $p < 0,001$, sin riesgo de sesgo de publicación. Se encontraron beneficios bajos-moderados casi significativos para la implicación de la familia, $r = 0,380$, $p = 0,052$, 95% CI = [-0,004, 0,667]. El resultado para el estatus socioeconómico fue estadísticamente significativo con una correlación baja-media, $r = 0,233$, $p < 0,001$, 95% CI = [0,104, 0,354]. **Conclusiones:** se constata que el input lingüístico parental durante los primeros años tras la implantación predice los resultados lingüísticos, siendo necesarios los programas de intervención temprana que proporcionen apoyo a los padres incrementando la exposición de los niños a contextos conversacionales de calidad.

SIMPOSIO

Retrato barroco. De la técnica de análisis más utilizada al problema de la pérdida de datos

Coordinan: M^a Paula Fernández García y Roser Bono Cabré

Resumen. Ya sea en condiciones accidentales o planificadas, sea cual sea la conducta que se observe, y para centrar el foco, valgan aquellas observadas en el ámbito de las ciencias sociales, de la educación y de la salud, siempre se ha considerado implícita o explícitamente su carácter dimensional. El desarrollo del software hace posible analizar los datos utilizando técnicas estadísticas sofisticadas que permiten mapear, analizar, factorizar y modelar la complejidad de la conducta investigada. La primera comunicación presentada revela el incremento en la frecuencia de uso de los modelos de ecuaciones estructurales y modelos multinivel. Esto sólo se entiende por el interés en el estudio de conjuntos de múltiples variables relacionadas, organizadas en diferentes niveles de agregación y que fluctúan de modo aleatorio en base a variables que condicionan la conducta. Posiblemente

el problema que más atención está recibiendo de los investigadores centrados en el desarrollo de las técnicas estadísticas sea cómo conservar la potencia y robustez de las técnicas cuando se debe hacer frente a la pérdida de datos que irremediamente se produce en el curso de una investigación. Es tan improbable que realizando una investigación longitudinal no se produzca pérdida de datos que algunos investigadores consideran que "la matriz de datos completa" es un caso especial de la "matriz de datos incompleta". Tres comunicaciones se centran en este aspecto. La primera estudia el poder que tienen los criterios de ajuste para discriminar si el mecanismo de pérdida subyacente es MCAR, MAR o MNAR. La segunda investiga la superioridad del modelo de patrones mezclados sobre el modelo mixto cuando la pérdida es MNAR. La tercera investigación, de la mano del concepto de validez estructurada propone directrices para elaborar un protocolo que permita perder el menor número de datos posible y reducir el impacto de las pérdidas que irremediamente se produzcan. Como George E. P. Box expresó: "All models are wrong; some models are useful". No es difícil encontrar paralelismo entre las características del barroco y las que definen el análisis de datos actual. Valga el título por tanto.

Tendencias actuales en metodología de investigación y análisis de datos en psicología. María J. Blanca¹, Rafael Alarcón¹ y Roser Bono²

¹Universidad de Málaga, ²Universidad de Barcelona

El objetivo de esta investigación fue analizar las prácticas actuales con respecto a la metodología de investigación, técnicas de análisis de datos y programas estadísticos utilizados en la investigación psicológica. Para ello, se realizó una revisión de la investigación empírica publicada en revistas europeas y americanas pertenecientes a las principales categorías del Journal Citation Reports situadas en primer o segundo cuartil. Para la revisión se diseñó una clasificación exhaustiva de las diferentes metodologías y técnicas estadísticas. Se examinaron 288 artículos publicados en 2017 y 2018, identificándose 663 técnicas de análisis de datos. Los resultados mostraron que las metodologías más frecuentes fueron la experimental y correlacional. De forma coherente, las pruebas paramétricas para comparación de medias y las técnicas para analizar la relación entre variables fueron más prevalentes. Entre estas últimas, la regresión jerárquica y los modelos de mediación fueron los más usados. Comparando estos resultados con los encontrados en investigaciones previas, se observa un importante decremento de la estadística no paramétrica y de los modelos de regresión que utilizan procedimientos de selección automática de predictores. Asimismo, también se observa un aumento en el uso de ecuaciones estructurales para identificar relaciones entre variables y de los modelos que utilizan estructura jerárquica de los datos o modelos multinivel. Estos resultados indican que la investigación psicológica utiliza procedimientos más avanzados de análisis de datos que los encontrados en estudios previos. En relación con los programas estadísticos, el IBM SPSS fue el más usado, seguido de MPLUS, PROCESS MACRO y AMOS. Agradecimientos: este trabajo ha sido subvencionado por el Ministerio de Ciencia, Innovación y Universidades, proyecto PSI2016-78737-P (AEI/FEDER, UE).

Evaluación de la potencia estadística de diferentes criterios de selección para discriminar entre diferentes mecanismos de pérdida en tres escenarios distintos. Pablo Livacic Rojas¹, Guillermo Vallejo², Paula Fernández² y Ellián Tuero Herrero²

¹Universidad de Santiago de Chile, ²Universidad de Oviedo

El objetivo del presente trabajo es evaluar la potencia estadística de cinco criterios de información para discriminar entre diferentes mecanismos de pérdida de datos en diseños de medidas repetidas a los que subyacen diferentes estructuras de covarianza representativas de tres escenarios distintos. Se realizó una investigación Monte Carlo generando 1.000 simulaciones de cada una de las condiciones que resultan de cruzar los niveles de las variables: tamaño total de muestra, emparejamiento de las matrices de covarianza y el tamaño muestral, número de ocasiones de medida, igualdad y desigualdad de las matrices de dispersión, distribución subyacente, mecanismo y porcentaje de pérdida. Los resultados muestran que los criterios de selección AIC, AICCR y HQICR tienen los niveles de potencia promedio más alto en todas las condiciones analizadas, y los criterios de selección BICR, CAICR muestran los niveles de potencia promedio más bajos para todas las condiciones. Sin embargo, los tres primeros seleccionan mejor en los escenarios 2 y 3 cuando Σ es NE y los grupos son pequeños ($n=50$), la relación entre las matrices y los grupos es negativa y la distribución es no normal. A su vez, BICR, CAICR realizan mejor ejecución cuando Σ es ARH y $n=50$. Inicialmente, los criterios de selección AIC, AICCR y HQICR serían más recomendables bajo los escenarios 2 y 3 (matrices Σ NE, grupos pequeños, relación Σ/n negativa y distribución normal), mientras que, BICR, CAICR serían más recomendables bajo los escenarios 2 y 3 (cuando Σ es ARH, relación Σ/n positiva y distribución no normal). Agradecimientos: este trabajo ha sido subvencionado por el Fondo Nacional de Desarrollo Científico y Tecnológico de Chile, proyecto FONDECYT. Ref.: 1120271.

Mecanismo de pérdida MNAR monótono y aleatorio. Evaluación de la robustez de los procedimientos MLM y MPM en un diseño de medidas parcialmente repetidas. Paula Fernández¹, Guillermo Vallejo¹, Pablo E. Livacic Rojas² y Ellián Tuero Herrero¹

¹Universidad de Oviedo, ²Universidad de Santiago de Chile

La pérdida de datos en una investigación provoca de modo inevitable pérdida de potencia de prueba, sin embargo, no es la peor consecuencia que se puede producir. Cuando la pérdida de datos obedece a un mecanismo de pérdida no aleatorio, pérdida MNAR, también está comprometida la inferencia sustantiva y es muy difícil poner a prueba las hipótesis de interés. El MLM tienen un comportamiento excelente cuando los datos están completos y cuando la pérdida es MAR. El MPM es una de las alternativas posibles de análisis en condiciones de pérdida MNAR. Realizamos una investigación Monte Carlo para examinar la robustez experimentada por las fuentes de variación de un diseño de medidas parcialmente repetidas cuando los datos son analizados por ambos procedimientos en condiciones de pérdida MNAR monótona y aleatoria. Realizamos 5000 simulaciones de 1440 condiciones resultantes de manipular distintos niveles de las variables: total tamaño de muestra, heterogeneidad de Σ , razón entre el tamaño de los grupos y el de las varianzas, grado de diferencia entre el tamaño de los grupos, forma de la distribución de

los datos, control de la heterogeneidad y cualidad del patrón MNAR. Los resultados ponen en evidencia que el MPM tiene el mejor comportamiento, pero no es robusto en todas las condiciones ni para todas las fuentes de variación del mismo modo, además, mayor tamaño de muestra no implica siempre un mejor comportamiento. El hecho de que la pérdida sea monótona o aleatoria no provoca ninguna variación sustancial en el comportamiento de ninguno de los métodos analizados. Agradecimientos: este trabajo ha sido subvencionado por el Ministerio de Economía y Competitividad de España, proyecto PSI2015-67630-PSIC.

Directrices para reducir el impacto de la pérdida de datos. El concepto de Validez Estructurada aplicado a la planificación, realización, seguimiento y análisis de una investigación. Paula Fernández¹, Guillermo Vallejo¹, Ellián Tuero Herrero¹, Pablo E. Livacic Rojas² y Feliciano Ordóñez Fernández^{3,4}

¹Universidad de Oviedo, ²Universidad de Santiago de Chile, ³Facultad Padre Ossó de Oviedo, ⁴Universidad Internacional de La Rioja

La mejor solución es táctica, no técnica. Todos los investigadores dedicados al desarrollo y al estudio de técnicas estadísticas para abordar el análisis de los datos en condiciones de pérdida convergen en que la única solución para poner a prueba las hipótesis de modo razonable es cuidar la investigación para perder la menor cantidad de datos posible, y para que las pérdidas que de modo inevitable se producen tengan el menor impacto posible en la prueba de la hipótesis. Proponemos un protocolo de investigación amparado en coordenadas filosóficas y metodológicas que implica diferentes directrices en cada una de las fases de la misma, planificación, realización, seguimiento y análisis. El protocolo debe realizarse en un orden determinado inalterable. Si se realiza de modo correcto, la validez de constructo y la validez interna estarán garantizadas, la pérdida de datos se producirá en un porcentaje pequeño y será MCAR o MAR en mayor medida porque el margen para una pérdida MNAR será muy reducido, el análisis de los datos devolverá resultados garantizando la validez de conclusión estadística, y el alcance de la validez externa será conocido. Llevar a cabo el protocolo con éxito implica trabajo en equipo satisfactoriamente coordinado y la complicidad de los sujetos que participan en la investigación. Posiblemente, este último aspecto sea el punto más vulnerable del protocolo propuesto, y difícil de mantener en algunas investigaciones, sin embargo, es una pieza imprescindible en este planteamiento. Agradecimientos: este trabajo ha sido subvencionado por el Ministerio de Economía y Competitividad de España, proyecto PSI2015-67630-PSIC.

SIMPOSIO

Aplicaciones de la metodología observacional

Coordina: Elena Rodríguez Naveiras

Resumen. A la hora de estudiar el comportamiento en el contexto natural en el que se da, la metodología observacional ofrece un procedimiento de registro sistemático objetivo. Esta metodología rigurosa y a la vez flexible, da la posibilidad de utilizar instrumentos de observación que permiten obtener información clara y fiable de lo que sucede en diferentes contextos de aplicación. 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 el presente simposio se presentan diversos trabajos en los que se hace uso de instrumentos de observación para la evaluación de la conducta del docente en educación básica en México; la evaluación de interacciones de los alumnos en un programa extraescolar socioafectivo; la presentación de un primer instrumento para la evaluación de la conducta del mediador y, por último, el análisis de indicadores comportamentales para determinar la consecución de los objetivos previstos en las actividades que se llevan a cabo en un programa extraescolar.

Protocolo de observación de la mediación-técnicas. Triana Aguirre Delgado¹, África Borges del Rosal¹ y María Aránzazu Calzadilla Medina¹

¹Universidad de La Laguna

En el ámbito de las Ciencias Sociales y de la intervención social en general, se asiste en los últimos años a un significativo incremento del estudio y de la práctica de la mediación como instrumento para solucionar conflictos. Entre las carencias más relevantes en la investigación sobre mediación se encuentran dos campos fundamentales: las intervenciones de los mediadores y la relación entre dichas intervenciones y los resultados obtenidos en el proceso. Es por ello por lo que se sitúa el punto de mira en las intervenciones de los mediadores, en las estrategias y técnicas utilizadas para llevar adelante el proceso de mediación, pero sin perder la perspectiva de la eficacia de estas. 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 y de la propia mediación pasa ineludiblemente por la debida formación y cualificación de las personas mediadoras que han de llevarlo a cabo. Una de las formas más adecuadas para estudiar el comportamiento en contextos naturales, es la metodología observacional, ya que permite extraer información muy rica y flexible. No obstante, en la actualidad no existen instrumentos que evalúen el comportamiento de la persona mediadora durante el proceso de mediación, y por ende que evalúen a las personas en formación en mediación. Por ello, es necesario ofrecer directrices de intervención en la mediación y un acercamiento a la definición aquellas habilidades y conductas que deben ser entrenadas y aprendidas por las personas que ejercen la mediación. En esta comunicación se presenta la estructura del primer instrumento de observación de PROME-TE.

Protocolo de observación de funciones docentes en educación básica México, versión 1 (PROFUNDO-EBM, v.1). Doris Castellanos Simons¹, Elia Torres¹, África Borges del Rosal² y Manuela Rodríguez Dorta²

¹Universidad Autónoma del Estado de Morelos, ²Universidad de La Laguna

Uno de los retos de la sociedad actual es ofrecer una educación de calidad. Esto hace necesario evaluar todos los procesos y agentes implicados en el sistema educativo y, concretamente, en el contexto de enseñanza-aprendizaje que tiene lugar en el aula. Es aquí donde la figura del docente cobra especial relevancia, siendo diversos los procedimientos empleados para evaluar su

comportamiento como profesional educativo. Desde la comparación con estándares internacionales, hasta la revisión de su portafolios, pasando por entrevistas y autoinformes a alumnado, compañeros y el propio docente. No obstante, una de las formas más adecuadas para estudiar el comportamiento en contextos naturales, la constituye la metodología observacional, ya que permite extraer información muy rica y flexible. Se presenta aquí la adaptación al contexto mexicano de Educación Básica de un protocolo de observación previo creado para analizar el comportamiento docente de maestros de Educación Primaria y Especial en el marco de la educación en España. Este instrumento está basado en un modelo teórico que recoge las funciones que un docente lleva a cabo antes, durante y después del proceso educativo que se desarrolla en el aula. De estas funciones docentes, el protocolo operativiza aquellas que son directamente observables durante el transcurso de las clases.

Evaluación formativa del cumplimiento de indicadores a través de vídeos. Rubens Montero¹, Belén Ortega¹ y Adalberto González¹

¹Universidad de La Laguna

Uno de los procesos más relevantes en la evaluación de un programa es idear instrumentos que permitan la evaluación tanto sumativa como formativa del mismo. Por esta razón, para determinar en esta última, la consecución de los objetivos planteados en el desarrollo de las actividades propuestas por el PIPAC (Programa Integral para Altas Capacidades). Se ha ideado un cuestionario de valoración para cada sesión. Las sesiones son diseñadas por un educador principal que la implementa en compañía de dos educadores secundarios, junto al diseño de la sesión se planifican indicadores que permitan determinar de forma operativa si los objetivos previstos se han alcanzado. El desarrollo de las sesiones es grabado en video sistemáticamente. Al finalizar cada sesión los educadores participantes completan este cuestionario desde cualquier dispositivo móvil, quedando así un registro de su evaluación al terminar la sesión. En el presente trabajo se compara la información contenida en estos registros con la evaluación posterior por observadores externos y entrenados de las sesiones grabadas en video, contrastando así la valoración de los educadores.

Determinación de relaciones diádicas en un programa psico-educativo. Pablo Hernández Lastiri¹, María Cadenas² y África Borges¹

¹Universidad de La Laguna, ²Universidad Europea de Canarias

Un factor relevante en el desarrollo de las personas son las relaciones tempranas con los pares. Tanto la interacción como el ajuste social acostumbran a ser objetivos de contextos intra y extraescolares destinados a la atención de alumnado con altas capacidades. Así, es necesario su estudio en contextos formales e informales, más aún cuando pretenden el desarrollo socioafectivo o social de esta población. En este trabajo se estudian las interacciones sociales con los compañeros de dos participantes del Programa Integral Para Altas Capacidades (PIPAC) de la Universidad de La Laguna, como una ampliación de un estudio previo. Para apresar las conductas relevantes se emplea metodología observacional, utilizando códigos del instrumento Protocolo

de Observación de la Interacción en el Aula (PINTA, v.5), siendo necesario discriminar los compañeros concretos que interactúan en lugar de codificarlos como un todo. Se observa cierta constancia en el intercambio de conductas con compañeros específicos cuando las frecuencias de las conductas del emisor y receptor son equiparables entre sí. El interés de este estudio reside en la evaluación de conductas de interacción, diferenciando entre los compañeros, de un programa socioafectivo que promueve las relaciones sociales adecuadas para propiciar el ajuste de sus participantes.

SIMPOSIO

Meta-análisis de generalización de la fiabilidad

Coordina: Fulgencio Marín Martínez

Resumen. Los meta-análisis de generalización de la fiabilidad son un tipo especial de meta-análisis en los que se integran estadísticamente los coeficientes de fiabilidad reportados por los estudios primarios que han aplicado un determinado test, con objeto de comprobar si la fiabilidad de las puntuaciones del test no cambia o si, por el contrario, varía en función de factores tales como la variabilidad de las puntuaciones del test, la población de referencia de la muestra, la longitud del test o la versión/adaptación del mismo. Las peculiares características de los meta-análisis de GF han llevado a la realización de investigación metodológica y estadística para optimizar su aplicación. En este simposio se presentan cuatro investigaciones realizadas en la Unidad de Meta-análisis de la Universidad de Murcia dirigidas a mejorar la metodología de los meta-análisis de GF. La primera comunicación, presentada por Julio Sánchez Meca, ofrece los resultados de la aplicación de la guía REGEMA a los meta-análisis de GF realizados sobre tests psicológicos. La segunda comunicación corre a cargo de Desirée Blázquez Rincón y presenta los resultados de un estudio de simulación para investigar el sesgo de reporte en los meta-análisis de GF. La tercera comunicación, presentada por Carmen López Ibáñez, presenta los resultados de aplicar diferentes métodos de integración estadística a un amplio número de meta-análisis de GF. Finalmente, la cuarta comunicación, presentada por Laura Badenes Ribera, ofrece los resultados de integrar meta-analíticamente coeficientes alfa frente a coeficientes omega, con objeto de investigar diferencias en los resultados.

Prácticas en el Reporte de Meta-análisis de Generalización de la Fiabilidad: Evaluación con la Guía REGEMA. Julio Sánchez Meca¹, Fulgencio Marín Martínez¹, Rosa M^a Núñez Núñez², María Rubio Aparicio³, José Antonio López López¹, Desirée M^a Blázquez Rincón¹, Carmen López Ibáñez¹, Rubén López Nicolás¹, José Antonio López Pina¹ y Juan José López García¹

¹Universidad de Murcia, ²Universidad Miguel Hernández de Elche, ³Universidad de Alicante

Los meta-análisis de generalización de la fiabilidad (GF) son un tipo particular de meta-análisis cuyo objetivo es investigar las fuentes que provocan variabilidad en los coeficientes de fiabilidad obtenidos en diferentes aplicaciones de un mismo test. Existen en la actualidad diversas guías elaboradas para orientar

el correcto reporte de los meta-análisis. Entre éstas destacan las diversas versiones de la guía PRISMA, MOOSE, AMSTAR, o MARS. Sin embargo, ninguna de estas guías está pensada para su aplicación a meta-análisis de GF. Por sus características peculiares, los meta-análisis de GF requieren una guía específicamente ideada para éstos. Es por ello que, recientemente, nuestro equipo de la Unidad de Meta-análisis de la Universidad de Murcia elaboró la guía REGEMA, dirigida a este tipo de meta-análisis. El propósito de esta investigación es aplicar la guía REGEMA a todos los meta-análisis de GF realizados hasta la fecha sobre tests psicológicos, con objeto de comprobar el grado en que estos meta-análisis cumplen con las buenas prácticas de reporte. Para este fin se realizó una revisión sistemática que permitió localizar más de 150 meta-análisis de GF. Cada uno de ellos se sometió a la aplicación de la guía REGEMA, la cual se aplicó doblemente a cada meta-análisis por dos codificadores debidamente entrenados. Los resultados mostraron un acuerdo inter-jueces alto para la mayoría de los ítems de la guía, y moderado para algunos de ellos. También se ha encontrado una elevada variabilidad en el grado de cumplimiento de los diferentes ítems de la guía. *Investigación financiada por el Ministerio de Economía, Industria y Competitividad y por los Fondos FEDER (proyecto nº PSI2016-77676-P).

Sesgo de reporte en los meta-análisis de generalización de la fiabilidad: una simulación Monte Carlo. Desirée M^a Blázquez Rincón¹, Carmen López Ibáñez¹, Julio Sánchez Meca¹, Juan Bottella² y Manuel Suero²

¹Universidad de Murcia, ²Universidad Autónoma de Madrid

Vacha-Haase propuso el procedimiento conocido como 'generalización de la fiabilidad' (GF), destinado a integrar coeficientes de fiabilidad independientes obtenidos con un mismo instrumento de medida. Desde entonces, se ha hecho evidente el alto porcentaje de estudios primarios incluidos en este tipo de meta-análisis (MA) que, o bien omiten, o bien inducen la fiabilidad del instrumento valiéndose de estudios anteriores. Para probar que los estudios que omiten o inducen la fiabilidad lo hacen por razones meramente aleatorias y no con el objetivo de esconder un coeficiente de fiabilidad bajo (o sesgo de reporte), es necesario que el meta-analista cuente con herramientas estadísticas *ad hoc*. Sin embargo, en el campo de la GF sólo existe una contribución sólida: el número de seguridad de Howell y Shields (2008). Dadas las limitaciones que presenta este método (necesidad de modelar el sesgo y carencia de reglas para determinar la magnitud del resultado), creemos necesario desarrollar otros procedimientos. El presente estudio tiene por objetivos: (1) ajustar los procedimientos de detección de sesgo de publicación propuestos en el campo del MA al terreno de la GF, y (2) comparar en términos de sesgo y eficiencia los resultados de dichos procedimientos en bases de datos simuladas. Esperamos que los resultados de este estudio: (a) ayuden a los meta-analistas a valorar, e incluso corregir, la posible influencia del sesgo de reporte en sus estudios de GF, y (b) contribuyan a sensibilizar a la comunidad científica sobre este fenómeno, tan poco estudiado en el ámbito de la GF. *Investigación financiada por el Ministerio de Economía, Industria y Competitividad y por los Fondos FEDER (proyecto nº PSI2016-77676-P).

Meta-análisis de generalización de la fiabilidad: una comparación entre diferentes estrategias de análisis estadístico. Carmen López Ibáñez¹, Desirée M^a Blázquez Rincón¹ y Julio Sánchez Meca¹

¹Universidad de Murcia

El enfoque meta-analítico de generalización de la fiabilidad (GF) trata de analizar la variación de los coeficientes de fiabilidad de un test en sus diferentes aplicaciones, tratando de generalizar dicha fiabilidad al propio test. Uno de los objetivos principales de los estudios GF es la obtención de un promedio representativo de los coeficientes de fiabilidad. Para el cálculo del promedio y su intervalo de confianza, se han propuesto diferentes métodos analíticos en función de: (a) transformar o no transformar los coeficientes de fiabilidad, y (b) el modelo estadístico asumido (mínimos cuadrados ordinarios, efectos fijos, coeficientes variables y diferentes versiones del modelo de efectos aleatorios). De esta combinación surgen 13 métodos para el cálculo del coeficiente de fiabilidad promedio y 18 métodos para el cálculo de la amplitud confidencial. El objetivo de este trabajo es comprobar si existen diferencias relevantes entre los diferentes métodos analíticos tanto para el cálculo del coeficiente promedio como para el cálculo de la amplitud confidencial. Con este propósito se llevó a cabo una revisión sistemática metodológica, para lo cual se obtuvieron más de 120 bases de datos reales procedentes de meta-análisis de GF. La aplicación de los diferentes métodos para promediar coeficientes alfa y sus intervalos de confianza dio lugar a resultados variables, especialmente en cuanto a los intervalos de confianza. Se discute el alcance de estos resultados para la realización de meta-análisis de GF. *Investigación financiada por el Ministerio de Economía, Industria y Competitividad y por los Fondos FEDER (proyecto nº PSI2016-77676-P).

Coefficiente alfa versus omega: un Estudio comparativo de su funcionamiento en meta-análisis de generalización de la fiabilidad. Laura Badenes Ribera¹, María Rubio Aparicio², Cristian Molla Esparza¹ y Julio Sánchez Meca³

¹Universidad de Valencia, ²Universidad de Alicante, ³Universidad de Murcia

La fiabilidad, una de las propiedades psicométricas más importantes al aplicar un test psicológico a una muestra de participantes, no es una propiedad inherente del test, sino de las puntuaciones en una aplicación determinada del mismo, por lo que la fiabilidad puede variar dependiendo de las características de la muestra a la que se administra. El meta-análisis de generalización de fiabilidad (GF) es un método para integrar estadísticamente las estimaciones de fiabilidad obtenidas en diferentes aplicaciones de un test. Por lo tanto, el meta-análisis GF permite caracterizar la fiabilidad media de las puntuaciones obtenidas mediante un test en múltiples estudios y situaciones, y estimar el grado de variabilidad en los coeficientes de fiabilidad en diferentes tipos de muestras, contextos y características del estudio. De esta manera, es posible determinar en qué circunstancias un test produce las puntuaciones más fiables, para qué tipo de personas y en qué contextos. El propósito de este estudio fue comparar el funcionamiento de dos coeficientes de fiabilidad, alfa de Cronbach y omega para: (a) estimar la fiabilidad media como consistencia interna del test y (b) estimar la varianza inter-estudios. Para ello se elabó

boró una base meta-analítica, con una muestra de 26 estudios publicados que aplicaron la Escala de Auto-compasión y reportaron dichos coeficientes, sobre la que se ha trabajado para valorar empíricamente en qué grado ofrecen resultados discrepantes estos dos coeficientes de fiabilidad de consistencia interna. Se discute el alcance de las diferencias observadas entre los dos coeficientes. *Investigación financiada por el Ministerio de Economía, Industria y Competitividad y por los Fondos FEDER (proyecto nº PSI2016-77676-P).

SIMPOSIO

Estimación y generalizabilidad en diseños observacionales

Coordina: Ángel Blanco Villaseñor

Resumen. Aunque existe en la literatura científica el término "estimación muestral", nos vamos a referir con ello a los procesos de precisión de estimación en el tamaño muestral de las diferentes facetas que integran un estudio observacional con el fin de poder llevar a cabo estudios de observación directa del comportamiento, guiados por las estructuras de intercambiabilidad de facetas, que nos ofrecen los diferentes planes de medida multi-método de la teoría de la generalizabilidad (TG). En distintas situaciones en las que realizamos mediciones para evaluar el comportamiento observado podemos encontrarnos con ciertos problemas, sobre todo en los valores que toman los errores de medida en cualquier modelo de validez de generalización, donde no podemos determinar realmente si parte de la variabilidad observada se debe a la variabilidad real debida a diferentes situaciones o simplemente al propio artefacto estadístico. Es decir, los factores contextuales o situacionales que son responsables de esa situación específica pueden ofrecernos un resultado diferente en la cantidad de error de medida en diferentes estudios. De tal forma que los coeficientes que presentan un error de medida determinado pueden estar confundiendo la variabilidad debida a las propias fluctuaciones aleatorias en una situación similar con las debidas a otras situaciones diferentes. El objetivo de este simposio será también evaluar el error de medida de los diferentes tamaños de muestra desde la perspectiva de la TG. La TG permite realzar la naturaleza multidimensional del error, la importancia del diseño observacional utilizado para registrar los datos de un determinado estudio y focaliza su atención en la generalizabilidad de las inferencias que realizamos basándonos en nuestras mediciones. La TG nos proporciona procedimientos rigurosos y altamente flexibles para estimar la generalizabilidad de las inferencias o decisiones que proporcionan las diferentes fuentes de error provenientes de las estructuras multi-método.

Generalización y estimación muestral de los resultados de estudios observacionales. Ángel Blanco Villaseñor¹

¹Universidad de Barcelona

Ante las diversas amenazas de riesgo existentes que afectan a la precisión de un registro, podemos actuar diseñando un plan de evaluación de la calidad basado en las distintas facetas o fuentes de variación propias de un estudio determinado, con el fin de

conocer en qué medida en las distintas posibilidades de organización de la información funcionan los componentes de variación correspondientes a cada una de las facetas así como la magnitud del componente del error. El propósito de este trabajo consiste en aplicar en el ámbito observacional diferentes estructuras de diseño basados en la teoría de la generalizabilidad (TG). En los estudios empíricos que presentamos y en otras situaciones de medida frecuentemente no se trata de diferenciar individuos, sino más bien de diferenciar contextos, tratamientos, escalas de medida, observaciones, sesiones, ocasiones de registro, etc. En palabras de Cardinet, Tourneur y Allal (1976), esto significa asumir el principio de SIMETRÍA, es decir que sucesivos objetos de medida pueden ser evaluados dentro de un mismo diseño.

Evaluación de las funciones ejecutivas en alumnado de Educación Secundaria Obligatoria: medida y optimización mediante la teoría de la generalizabilidad. Elena Escolano Pérez¹ y Ángel Blanco-Villaseñor²

¹Universidad de Zaragoza, ²Universidad de Barcelona

Las funciones ejecutivas son procesos cognitivos de alto nivel cuya característica común es la coordinación del procesamiento de la información y el control de la acción para alcanzar objetivos previamente determinados. Presentan un prolongado curso de desarrollo a lo largo del ciclo vital, alcanzado un pico de crecimiento importante durante la adolescencia. Actualmente las funciones ejecutivas constituyen objeto de estudio de gran interés dada su relevancia e implicación en numerosos ámbitos de la vida, como por ejemplo, el éxito académico. En relación con todo ello, el objetivo de este estudio es analizar si los resultados referidos a la evaluación de las funciones ejecutivas a través de la administración del cuestionario AEFI en una muestra de 263 adolescentes estudiantes de Educación Secundaria Obligatoria pueden ser generalizables a una muestra de mayor tamaño y características similares a la estudiada. Los resultados arrojan un adecuado índice de generalizabilidad tanto para la muestra total (0,92) como para cada uno de los grupos de estudiantes participantes en el estudio (alumnado de 3º ESO = 0,87; alumnado de 4º ESO = 0,90). La estimación de la precisión ha sido desarrollada por la teoría de la generalizabilidad, asumiendo que hay otras fuentes de variación además de las diferencias individuales e integrando cada una de estas fuentes de variación en una estructura global, que permite aplicaciones particulares de la teoría estadística del muestreo. La teoría de la generalizabilidad reconoce explícitamente las múltiples fuentes de error de medida (observadores, sesiones, días, momentos, etc.).

Teoría de la generalizabilidad. Análisis de la variabilidad técnico-táctica en el fútbol de alto rendimiento: estimación de la precisión y generalización. Rubén Maneiro Dios¹ y Ángel Blanco-Villaseñor²

¹Universidad Pontificia de Salamanca, ²Universidad de Barcelona

El análisis de la variabilidad en el fútbol ya ha sido puesto a prueba en trabajos precedentes de naturaleza física y fisiológica con resultados interesantes de cara a la optimización del rendimiento físico y condicional. En cambio, el análisis de la variabilidad todavía no ha sido puesto a prueba en estudios de naturaleza técnico-táctica. Por ello, el propósito de este estudio es analizar la va-

riabilidad técnica y táctica de los jugadores durante su desempeño deportivo en el fútbol de alto rendimiento, con respecto a seis variables situacionales: demarcación, fase del campeonato, tipo de contacto, zona de inicio, zona de finalización y el rival. Se han establecido 4 niveles de éxito como indicadores de rendimiento de la fase ofensiva de la Selección Española de Fútbol en la UEFA Euro 2012. Resulta halagador para el científico social, sabedor de las múltiples fuentes de variación que inciden en un problema determinado, disponer de una vía que le asegura conocer el 'peso' de cada una de ellas respecto las demás, así como las 'ratios' relativas a la incidencia que cada una puede desempeñar en relación con las restantes, o, lo que es mucho más importante, los recursos de que se debería disponer –niveles de las respectivas facetas– y el esfuerzo que se debería desempeñar para alcanzar unos objetivos óptimos de acuerdo con el plan previsto. Indudablemente, este planteamiento persigue de forma complementaria, pero con calado no menos importante, otra finalidad que tiene un carácter económico. Pretende, en efecto, situarnos, con total objetividad, ante la posibilidad de adoptar una serie de decisiones relativas al costo-efectividad del planteamiento investigador, e, incluso, al costo-eficacia en la mayoría de aplicaciones.

Comportamiento de la teoría de la generalizabilidad para los estudios de validez, fiabilidad y estimación en actividades deportivas. Carmen Rosa Sánchez-López¹ y Ángel Blanco-Villaseñor²

¹Universidad de la Laguna y ²Universidad de Barcelona

Una de las características propias de la metodología observacional (MO) es la necesidad de realizar un control expreso sobre la adecuada calidad de los datos obtenidos para la investigación. Actualmente uno de los procedimientos más utilizados para ello tiene su base en la teoría de la generalizabilidad (TG) (Cronbach, Gleser, Nanda y Rajaratnem, 1972; Cronbach, Rajaratnam y Gleser, 1963). El proceso de identificación de la fiabilidad, validez y precisión de una medida supone detectar que el error cometido es mínimo y que dicho error no le afecta. La TG unifica las definiciones de fiabilidad, validez y precisión, considerando este error como una faceta más o fuente de variabilidad para el estudio. En los estudios observacionales hay variaciones importantes en la metodología y en las técnicas utilizadas. Sin embargo, hay una característica que es común a todas ellas: sólo observamos y registramos una muestra del comportamiento de los individuos. Bajo estos supuestos, en este estudio se intentará precisar el número mínimo de sesiones de observación necesarias para obtener índices adecuados tanto en fiabilidad como en validez y precisión dependiendo de la actividad deportiva a observar.

SIMPOSIO

Soluciones metodológicas para las evaluaciones educativas de alto impacto

Coordina: Rubén Fernández Alonso

Resumen. Las administraciones educativas organizan y gestionan un gran número de pruebas y evaluaciones educativas. Probablemente las más conocidas son las evaluaciones de diagnóstico del sistema educativo y las evaluaciones internacionales. Se

trata de evaluaciones con alto impacto social como atestiguan las noticias de prensa que reseñan la aplicación de las mismas en los centros y la publicación de sus resultados. Sin embargo, las evaluaciones precitadas son sólo una pequeña parte de los programas de evaluación de las administraciones educativas. Anualmente las administraciones convocan, regulan y ejecutan decenas de pruebas que tienen tres finalidades: la obtención de un título oficial en diferentes etapas o enseñanzas (ESO, Bachillerato, Formación Profesional, enseñanza de idiomas...); el acceso a estudios no universitarios (Formación Profesional, enseñanzas musicales, artísticas, deportivas...); y la concesión de premios extraordinarios al rendimiento al alumnado que finaliza estudios en las enseñanzas de régimen general y de régimen especial. Este segundo tipo de pruebas, probablemente con menos huella mediática, tienen un importante impacto académico y personal en los aspirantes que concurren a las mismas toda vez ya que les permiten obtener un premio, un título válido en la Unión Europea o la continuación de estudios post-obligatorios. Este simposio presenta una muestra sustantiva de las soluciones analíticas implementadas por el grupo de Psicometría de la Universidad de Oviedo para la explotación de las bases de datos de estas evaluaciones, estudios internacionales de comparación de sistemas educativos (Elena Govorova y otros y Javier García-Crespo), evaluaciones nacionales o regionales para el seguimiento de cohortes (Álvaro Postigo y otros) y estudios especiales para el desarrollo de rúbricas de cara a la concesión de premios al rendimiento académico (Marcos Álvarez-Díaz y Rubén Fernández-Alonso).

Variables contextuales que hacen que un estudiante sea resiliente o low-performer en lectura. Países de la UE participantes en PIRLS. Francisco Javier García Crespo¹

¹Universidad Complutense de Madrid

Introducción: un estudiante resiliente es aquel que a pesar de encontrarse en una situación desventajada socioeconómicamente su rendimiento es muy superior al esperado. Por el contrario, un estudiante es *low-performer* si parte de una situación socioeconómica en desventaja y además su rendimiento es muy inferior al esperado. Se trata de evaluar que variables contextuales y que valores toman estas para que dos alumnos que se encuentran en la misma situación socioeconómica uno pase a ser resiliente y sin embargo el otro sea *low-performer*. Método: para llevar a cabo esta investigación se ha utilizado la base de datos de la evaluación internacional a gran escala PIRLS. Se han seleccionado los datos de los países de la Unión Europea que participaron en PIRLS. En un primer paso se clasifica a las y los estudiantes en resilientes, *low-performer* o de rendimiento esperado (serán de interés para este estudio las dos primeras categorías). Pasamos a seleccionar las variables de contexto extraídas del cuestionario del alumnado y de sus familias susceptibles de tener impacto positivo o negativo en las categorías a analizar. Formularemos dos modelos de regresión logística binaria, en el primero de ellos la variable dependiente será la resiliencia, mientras que en el segundo será la calidad de *low-performer*. Finalmente, analizaremos los coeficientes de las regresiones para evaluar el impacto que tienen las variables seleccionadas sobre las variables dependientes, nos fijaremos fundamentalmente en los odds ratio obtenido en los análisis y si estos son estadísticamente significativos o no. Resultados: se observa de manera generalizada que aquellas variables

que tienen impacto positivo en la resiliencia lo tienen en sentido contrario para el alumno *low-performer*, aunque el comportamiento es heterogéneo para los distintos países del estudio. Destacan la confianza en la lectura o el gusto por la misma. Conclusiones: políticas educativas dirigidas a potenciar el gusto por la lectura para conseguir una mayor confianza compensarían la desigualdad socioeconómica de partida.

Factores asociados al rendimiento en PISA 2015: análisis internacional. Elena Govorova¹, Isabel Benítez², Milagrosa Sánchez² y Noelia Muñoz²

¹Universidad de Oviedo, ²Universidad Loyola Andalucía

Introducción: en el presente trabajo se realiza el análisis de factores asociados al rendimiento en el contexto internacional mediante el uso de modelos jerárquicos lineales. Método: se utilizan los datos de 35 países de la OCDE participantes en PISA 2015 con una muestra de 237.720 alumnos. Se construye un modelo de dos niveles en el que los factores del nivel 1 (alumno) incluyen las variables demográficas y socioeconómicas, así como las variables relacionadas con el ambiente de aprendizaje y la implicación de los alumnos. Los factores del nivel 2 (escuela) incluyen el nivel socioeconómico de los centros, así como la información relativa a la disponibilidad de recursos, la autonomía en la toma de decisiones y los métodos de enseñanza y aprendizaje utilizados por los docentes. Resultados: los resultados reflejan la combinación de factores que tienen un impacto significativo en los resultados en cada uno de los países. Así, el nivel socioeconómico del estudiante, el disfrute de ciencias y la cooperatividad son las variables que muestran el efecto positivo y significativo en el rendimiento en todos los países analizados. A nivel de centro se destacan las estrategias basadas en la enseñanza dirigida por el profesor que tienen impacto positivo en el rendimiento en el 70% de los países. Conclusiones: los efectos de factores asociados al rendimiento varían entre países en cuanto a su dirección y magnitud por lo que, a la hora de diseñar las políticas educativas, es importante tener en cuenta el contexto individual de cada país.

El esfuerzo percibido modula el rendimiento académico: un enfoque longitudinal con modelos de clases latentes. Álvaro Postigo¹, Marcelino Cuesta¹, Eduardo García Cueto¹ y José Muñoz¹

¹Universidad de Oviedo

Introducción: el rendimiento académico de los estudiantes depende de variables personales y contextuales. Dentro de las personales juegan un papel relevante tanto las de carácter aptitudinal como las orécticas (personalidad, actitudes, manejo de emociones). El objetivo del presente trabajo es analizar en qué medida una variable oréctica como el esfuerzo académico percibido por los estudiantes influye en el rendimiento académico evaluado longitudinalmente. Método: se utilizó una muestra de 5371 estudiantes evaluados en dos momentos temporales separados por cuatro años: cuarto curso de Educación Primaria y segundo de Educación Secundaria. Se utilizó un análisis de clases latentes, una técnica que permite identificar grupos subyacentes a partir de unos indicadores observables. También se utilizó un análisis de transiciones latentes, una de las extensiones de esta técnica para

estudios longitudinales, que permite estudiar las transiciones entre los grupos latentes en los diferentes momentos temporales. Además, se realizó un ANOVA de medidas repetidas para analizar la influencia del esfuerzo percibido en el rendimiento académico. Resultados: se identificaron tres grupos latentes de esfuerzo percibido en ambos momentos temporales. El rendimiento académico disminuye entre la primera y segunda medición, si bien no lo hace linealmente, viniendo modulado por la evolución en el nivel de esfuerzo percibido. Conclusiones: la percepción que los estudiantes tienen de su compromiso académico es una variable que afecta a los resultados escolares. La intervención sobre dicha variable podría ser muy eficaz para la mejora del rendimiento académico.

Diseño y validación de una rúbrica para la evaluación de una producción abierta en un proceso de premios extraordinarios oficiales. Marcos Álvarez-Díaz¹ y Rubén Fernández-Alonso¹

¹Universidad de Oviedo

Introducción: los procesos de premios extraordinarios de las distintas etapas y enseñanzas forman parte del sistema educativo y si bien en ocasiones se dirimen con una prueba objetiva en otras la selección debe pasar por la evaluación de una producción abierta. La evaluación de dichas producciones abiertas en contextos competitivos experimenta con frecuencia discrepancias entre los expertos encargados de su valoración constituyendo esta circunstancia un reto para la evaluación en los sistemas educativos. Estas discrepancias pueden mitigarse utilizando rúbricas precisas que minimicen las diferencias entre los jueces. **Objetivo:** el objetivo de la presente investigación fue diseñar y validar una rúbrica analítica de evaluación que permitiese objetivar lo más posible la valoración de una producción abierta como es la interpretación musical en una situación competitiva (Premios Oficiales Extraordinarios de Enseñanzas Artísticas de Música). **Método:** un equipo de tres catedráticos superiores de música elaboró una rúbrica analítica compuesta de cinco criterios de corrección y tres niveles de calificación, con sus respectivos indicadores. Para la validación de la rúbrica dos psicómetras estudiaron su dimensionalidad, las fuentes de error, la fiabilidad entre jueces y la consistencia interna de los evaluadores. **Resultados:** se confirma la unidimensionalidad esencial de la rúbrica, no se encontraron efectos diferenciales entre los jueces, ni diferencias significativas en la consistencia interna de cada juez. **Conclusiones:** el uso de una rúbrica como herramienta de evaluación de la interpretación musical en un contexto competitivo repercute positivamente en una mayor fiabilidad y objetividad de los resultados, tanto a nivel de la consistencia interna como del acuerdo entre jueces.

SIMPOSIO

Construcción sistemática de ítems para instrumentos de medida

Coordina: Rafael Moreno Rodríguez

Resumen. Este simposium participa del convencimiento de que la construcción de ítems de instrumentos de medición psicológica, lejos de ser una tarea dependiente fundamentalmente de la inspiración y disposición natural de ciertas personas, puede y de-

be responder a criterios claros y eficientes que permitan obtener los niveles deseados de las propiedades psicométricas. Para esa construcción dos son las vías disponibles, complementarias entre sí: (a) desarrollo y aplicación de planteamientos teóricos-conceptuales, referidos a contenidos específicos sobre los que verse la evaluación o a aspectos metodológicos de formatos con los que evaluar cualquier contenido y (b) obtención y análisis de datos empíricos que informen sobre la conveniencia de cualquier decisión adoptada en la construcción. El presente simposium, que alcanza ya su novena edición en los congresos de AEMCCO, presenta una muestra de trabajos que vienen realizándose sobre el objetivo aquí planteado y siguiendo las dos vías mencionadas para alcanzarlo. Atendiendo a la vía teórico-conceptual, uno de los trabajos considera los aspectos de contenido de vocabulario emocional en la adaptación de pruebas a otros idiomas, evaluando su validez de contenido; el otro muestra desarrollos que para la generación de ítems y predicción de sus propiedades psicométricas permite una clasificación de formatos de ítems basada en las teorías de la comunicación y la información. El resto de trabajos atienden más a la vía empírica. Dos de ellos estudian determinados formatos: uno aborda las influencias de las variables cognitivas de los sujetos sobre el formato Likert, y el otro el comportamiento del formato elección múltiple en el contexto universitario. Los dos trabajos restantes abordan cómo determinados aspectos de formato o contenido de los ítems afectan a determinadas propiedades: uno se centra en el grado de dependencia local, y otro en la detección de patrones atípicos de respuesta.

Concordancia en el uso de términos emocionales y validez de contenido. Ana R. Delgado¹, Gerardo Prieto¹ y Debora I. Burin²

¹Universidad de Salamanca, ²Universidad de Buenos Aires-CONICET

Es común contar con expertos para adaptar el contenido de los ítems de una prueba a otros idiomas o culturas. En el caso de la investigación sobre las aptitudes emocionales, centrada en constructos que se solapan y cuyos indicadores pueden ser de difícil adaptación, es de interés contar con procedimientos menos subjetivos. El vocabulario emocional tiene la ventaja de que puede ser medido objetivamente. Con la finalidad de hallar evidencia de la validez de contenido de los ítems de un test de vocabulario emocional, y partiendo de un extenso corpus, hemos contrastado la predicción de consenso sobre la frecuencia de uso de 40 términos emocionales en castellano en las 8 principales áreas de habla hispana: España, México-América Central, Rioplatense, Caribe Continental, Andina, Antillas, Chilena y Estados Unidos. Nuestros resultados proporcionan evidencia de apoyo para el uso de estos ítems de vocabulario emocional en cualquiera de las áreas lingüísticas del español.

Clasificación y construcción sistemáticas de ítems basadas en las Teorías de la Comunicación y la Información. Rafael Martínez¹, Rafael Moreno¹ y José Muñiz²

¹Universidad de Sevilla, ²Universidad de Oviedo

Moreno, Martínez y Muñiz (2018) entienden los ítems como dispositivos de comunicación entre examinador y examinado, cuya función es la obtención de información sobre algún dominio psicológico del examinado. Acorde con esa propuesta, han planteado una taxonomía de formatos de ítems identificando cuatro compo-

nentes funcionales en todos ellos: la estructura del contenido del ítem, y los dispositivos que sirven para transmitir lo preguntado, recoger la respuesta, y aportar instrucciones sobre cómo entender el ítem y responderlo. En este trabajo pretendemos mostrar las posibilidades que el primero de tales componentes tiene para la clasificación y generación de ítems. A tal fin, se presentan las cuatro facetas constitutivas de dicho componente – dominio, marco, target y alternativas –, se formula el espacio muestral o universo de posibilidades de formatos que generan y se ilustran tales posibilidades mediante algunos ejemplos de ítems. Por último, y como complemento, se muestra cómo calcular la incertidumbre asociada a distintos formatos desde el punto de vista de la teoría de la información, como herramienta formal útil para la predicción cuantitativa de las propiedades métricas de los ítems.

El formato tipo Likert no funciona igual con todas las personas: el papel de las diferencias individuales. José Muñoz¹, Marcelino Cuesta¹, Javier Suárez¹, Alvaro Postigo¹ y Eduardo García Cueto¹

¹Universidad de Oviedo

Los ítems tipo Likert son ampliamente utilizados en la mayoría de los campos de las ciencias sociales y de la salud, y sus propiedades psicométricas se han investigado ampliamente. El presente trabajo trata de aportar datos sobre el comportamiento psicométrico de los ítems tipo Likert cuando se tienen en cuenta las diferencias individuales de las personas en inteligencia verbal y general. Se utilizó una muestra de 374 participantes con edades comprendidas entre 18 y 73 años, con media de 33,98 años y desviación típica de 14,12, el 62,60% mujeres. Se aplicó una escala de autoeficacia con tres condiciones: todos los ítems tipo Likert formulados de forma positiva, todos negativos y un combinado de ambos. También se evaluó la comprensión verbal y la inteligencia general. Las propiedades psicométricas se analizaron utilizando la teoría clásica de los test y los modelos de teoría de respuesta a los ítems. Para cada una de las tres condiciones de la escala de autoeficacia se obtuvo la estructura factorial, medias, coeficiente de fiabilidad, índices de discriminación de los ítems, parámetro a , función de información y DIF. Los mismos análisis se hicieron controlando el efecto de la comprensión verbal y la inteligencia general, mediante correlaciones parciales y análisis de covarianza. Ello permite comprobar en qué medida las variables cognitivas modulan las propiedades psicométricas de los ítems con formato Likert. Se discuten los resultados y se establecen algunas conclusiones, tanto de carácter básico como aplicado.

Estudio psicométrico de los exámenes de opción múltiple en la Universidad Autónoma de Madrid. David Cabeza¹, Vicente Ponsoda¹ y Carmen García¹

¹Universidad Autónoma de Madrid

El objetivo de esta investigación es el análisis psicométrico de todos los exámenes de opción múltiple procesados por el Gabinete de Estudios y Evaluación Institucional de la Universidad Autónoma de Madrid durante el primer semestre del curso 2017/18. De los 507 exámenes procesados, se analizaron todos (91) los respondidos por, al menos, 70 estudiantes. Los resultados muestran un coeficiente alfa de Cronbach promedio de 0,63 (0,71, si el examen tuviese 40 ítems), significativamente mayor que el obte-

nido en un estudio similar realizado también en la UAM (Sierra et al., 2010). No se obtuvieron diferencias significativas en la consistencia interna media de los exámenes por centros. Se realizó un análisis de conglomerados de todos los ítems analizados, a partir de los índices de discriminación de cada una de las opciones de respuesta y de la omisión. Se obtuvieron tres conglomerados de ítems, que hemos denominado “buenos” (30% del total), “aceptables” (31%) y “revisables” (39%). Se comenta la importancia y algunas consecuencias de no prestar la debida atención a la calidad de los exámenes aplicados en contextos universitarios.

Dependencia local y formato de los ítems. Gerardo Prieto¹ y Ana R. Delgado¹

¹Universidad de Salamanca

Para evaluar la comprensión lectora en los exámenes de segundas lenguas (Cambridge Assesment English, TestDaF del Goethe Institut, DELE del Instituto Cervantes) se emplean varios formatos de ítems: elección múltiple, emparejamiento, completamiento (*cloze*), etc. Todos los modelos TRI mantienen el supuesto de independencia local. Su violación puede inflar los estadísticos de fiabilidad y producir problemas en la validez de constructo. Con el fin de explorar la relación entre el formato de los ítems y el grado de dependencia local en pruebas reales, presentamos el análisis de 13 exámenes de comprensión lectora de nivel B1. Se han analizado mediante el estadístico Yen's Q3, el habitualmente usado para detectar la dependencia local, los cinco formatos de respuesta incluidos en cada examen: elección múltiple, emparejamiento, emparejamiento extendido, completamiento extendido y completamiento con elección múltiple. En todos los formatos se detectó dependencia local. No se encontraron diferencias significativas entre formatos, lo que puede atribuirse a que los ítems de cada tarea se derivan de un texto común. Se concluye planteando recomendaciones para la construcción de ítems.

Influencia de las características de la prueba en la presencia consistente de patrones atípicos de respuesta. Eduardo Doval¹, M. Dolors Riba¹ y Jordi Renom²

¹Universitat Autònoma de Barcelona, ²Universitat de Barcelona

La coherencia del patrón de respuestas en pruebas de evaluación constituye una de las condiciones necesarias para realizar inferencias válidas sobre el nivel individual de conocimientos a partir de las puntuaciones resultantes. En este sentido la detección de patrones atípicos de respuesta (PAR) constituye un recurso psicométrico imprescindible. Un PAR puede estar vinculado a una gran variedad de anomalías del proceso enseñanza-aprendizaje (E-A), algunas propias del mismo (p. ej., pautas de estudio incorrectas) y otras totalmente ajenas a él (p. ej., presencia de copia en la evaluación). Pensamos que este último tipo de anomalías generarán PAR detectables tanto si se analiza el conjunto de la prueba como una parte representativa en términos de dificultad, mientras que las primeras serán más dependientes del contenido de los ítems. En este trabajo estudiamos la consistencia, intraindividual e intraprueba, de los PAR analizando las respuestas de 425 estudiantes de segundo curso de psicología de la UAB a 24 preguntas de la asignatura Modelos estadísticos y psicométricos. Se ha utilizado el índice U3 de van der Flier para detectar PAR en el conjunto de la prueba y en divisiones de la misma según la

dificultad de sus ítems (dos partes de 12 ítems con dificultades similares) y el contenido de los mismos (dos partes, sobre fiabilidad o validez, de 12 ítems). Se concluye en términos de la importancia que tiene para la detección de PAR considerar los aspectos formales y de contenido a la hora de diseñar los ítems y configurar la prueba.

SIMPOSIO

Modelos y técnicas para la estimación de sistemas complejos de conectividad cerebral con señal fMRI

Coordina: Joan Guardia Olmos

Resumen. En los últimos veinte años se han impulsado muchos trabajos que pretenden relacionar el comportamiento de registros de señal cerebral con variables psicológicas que describen desde el rendimiento neuropsicológico, hasta perfiles de personalidad. No solo se han elaborado diversas aproximaciones al estudio de la señal cerebral a partir de los postulados de la Teoría de Detección de Señales, sino que además se han propuesto algunos heurísticos para el análisis de datos en este tipo de diseños. Evidentemente, ello ha generado una multitud de opciones y versiones para el estudio de situaciones parecidas y, por ende, existe una cierta confusión de procedimientos a aplicar. No se trata de una discusión muy metodológica, pero sí muy técnica desde el punto de vista estadístico aplicado. Curiosamente, existe escasa tradición desde la metodología psicológica para el estudio de los patrones de conectividad cerebral y esta propuesta pretende mostrar algunas de las opciones que hemos desarrollado para el estudio de conectividad funcional y efectiva con señal fMRI.

Analizar señal cerebral fMRI puede ser una caja oscura. Mari-bel Perú Cebollero¹

¹Universitat de Barcelona

El análisis de datos de señal fMRI implica una preparación previa de los registros llevados a cabo en el resonador, lo que comúnmente se conoce como preprocesamiento. Una de las funciones del preprocesamiento consiste en la optimización de la ratio señal-ruido y esta optimización se puede llevar a cabo en diferentes pasos del preprocesamiento, a saber, en el realineamiento, en la corrección por tiempo en que se registra el corte, en el suavizado e incluso aplicando filtros de banda alta y banda baja para filtrar la señal y así eliminar el ruido. Además, en el análisis de datos también se puede optimizar la ratio señal-ruido. Son diferentes los procedimientos que se pueden utilizar como eliminar por regresión de los registros temporales de señal BOLD las series temporales de los indicadores de movimiento establecidos en el realineamiento, o bien la aplicación de la estrategia analítica del análisis de componentes independientes (ICA), con el fin de determinar que componentes se pueden atribuir a ruido y que componentes se pueden atribuir a señal. Pero de toda esta serie de estrategias comentadas, no existe un estándar que permita afirmar cual es la mejor estrategia para la optimización de la ratio señal-ruido, es por este motivo que en la presente comunicación se pretende realizar una revisión de los diferentes procedimientos que se aplican y su adecuación.

Estimación de redes de conectividad funcional fMRI en situación de reposo en personas con síndrome de Down. María Carbó Carreté¹, Cristina Cañete Massé¹, Marc Montalà Flaquer¹ y Laia Farràs Permanyer¹

¹Universitat de Barcelona

En la población con Síndrome de Down (SD), la presencia de la enfermedad de Alzheimer (AD) se detecta clínicamente a partir de los 30 o 40 años de edad y es esencial identificar los factores protectores y sus efectos. Muchos estudios han demostrado las ventajas de usar técnicas de señalización cerebral no invasivas, incluido el uso de fMRI. Los estudios realizados hasta ahora con esta técnica y población son comparaciones entre grupos de activaciones estadísticamente significativas obtenidas de los registros de fMRI. Nuestro objetivo es evaluar el uso de las redes de conectividad cerebral como biomarcador del estado cognitivo de las personas con SD. Método: la muestra estuvo compuesta por 16 personas con SD entre 16 y 35 años ($M = 27,25$ y $SD = 5,12$) y el 18,8% eran mujeres. Se usó un resonador Siemens de 3 Tesla para la adquisición de imágenes de resonancia con los parámetros habituales con una imagen estructural ponderada en T2 y T1 de alta resolución y también un conjunto de datos de fMRI en estado de reposo de 6 minutos. Además, otras variables, como la gravedad de la ID, las actividades diarias o la intensidad de la actividad física se registraron para ver su efecto (si lo hubiera) en la red. Resultados: todas las imágenes fueron preprocesadas y se aplicaron una serie de técnicas de filtrado y estimación estadística para la mejor detección de la señal. Cada una de las redes de conectividad se caracterizó por la segregación de conectividad, obteniendo correlaciones significativas con las variables antes mencionadas. Los resultados indican que los sujetos con menos gravedad y mayor actividad diaria tienen más patrones de conectividad armónica que aquellos con mayor afectación y menos actividad diaria. Conclusiones: es factible considerar las propiedades de la red de conectividad como un biomarcador para una aproximación del estado cognitivo de las personas con SD.

Una aproximación al concepto de reserva cognitiva: *small data* y síndrome de Down. Cristina Cañete Massé¹, María Carbó Carreté¹, Guillermo Rubén Oviedo², Nuria Hidalgo Torres¹, Laia Farràs Permanyer¹, Neus Estanyol Torres¹, Casimiro Javierre Garcés¹, Roser González Domínguez¹, Myriam Guerra Balic² y Cristina Andrés Lacueva¹

¹Universitat de Barcelona, ²Universitat Ramon Llull

Introducción: la reserva cognitiva (RC) juega un importante rol en las diferencias individuales que existen con la aparición clínica y neuropatológica de las demencias. Es por ello que en los últimos años ha ido ganando mucho interés. Sin embargo, es un concepto que aún no ha quedado totalmente definido, sobre todo por la gran cantidad de variables que pueden participar en su operativización. Los adultos con síndrome de Down (SD) presentan una clara tendencia a desarrollar la enfermedad de Alzheimer (EA) hacia los 40 años. Es por ello que resulta una población interesante para poder estudiar el concepto de RC. En este estudio se pretende operacionalizar el concepto de reserva cognitiva en SD mediante diferentes técnicas aplicadas a *small data*. Básicamente es el proceso por el cual se hace uso y se analizan datos de menor tamaño y simplicidad, con el objetivo de poder caracterizar

a una población a través de diferentes indicadores. Método: la muestra está compuesta por 15 personas con SD entre 16 y 35 años de edad ($M = 27,25$ y $SD = 5,119$) y 18.8% mujeres. El muestreo ha sido accidental y el reclutamiento ha sido realizado por varios centros de SD. Se han registrado muchas variables con el objetivo de adquirir información específica de lo que en los últimos años se ha conocido como reserva cognitiva. Los participantes fueron evaluados por un cuestionario *ad hoc* compuesto por una evaluación neuropsicológica, de calidad de vida, de actividad física y estado nutricional. Todos los participantes fueron evaluados en dos sesiones con investigadores especializados. Resultados: la covarianza fue estimada entre las distribuciones observadas de las variables evaluadas que indican, aun con el tamaño de muestra, la existencia de una estructura de similitudes entre las diferencias medidas cercanas al concepto teórico de RC. Conclusiones: una medida óptima de estructura para la operacionalización de la RC puede ser identificada en personas con SD. Además, las técnicas aplicadas a small data permiten identificar los indicadores para caracterizar a los sujetos con SD en el concepto de reserva cognitiva. Finalmente, esto implica el primer paso para llegar a la generalización y a la identificación de grandes poblaciones.

Conectividad funcional en reposo en pacientes con trastorno depresivo mayor en remisión fMRI. Patricia Duarte Rosas¹, Esteban Gudayol Farré², Núria Mancho Fora¹, Cristina Cañete Massé¹, Marc Montalà Flaquer¹, Daniel Zarabozo Hurtado³, Antonio González Garrido⁴ y Joan Guàrdia Olmos¹

¹Universitat de Barcelona, ²Universidad Michoacana San Nicolás de Hidalgo, ³Grupo Rio, Guadalajara (México), ⁴Universidad de Guadalajara (México)

Introducción: diversos estudios con resonancia magnética han descrito patrones de conectividad funcional anormales en pacientes con trastorno depresivo mayor (TDM) que se relacionan con los síntomas neuropsicológicos característicos de esta enfermedad. Aunque el tratamiento antidepressivo ha sido asociado con cambios en la conectividad cerebral que tienden a la normalización de las alteraciones, se ha sugerido que esta normalización no se da de manera completa ni de manera homogénea a través de las diferentes redes aun cuando los pacientes se encuentran en fase de remisión sintomática. El objetivo de este estudio es identificar redes de conectividad funcional durante el estado de reposo en pacientes con TDM en fase de remisión para determinar si existen diferencias en los patrones de conectividad funcional en comparación con personas sanas. Material y Método: la muestra está conformada por 16 pacientes con diagnóstico TDM en fase de remisión (escala Hamilton 17-items = 3,6 (2,36)) con edad media de 33,12 (11,09) y 16 controles pareados por sexo, edad y escolaridad. Se adquirieron imágenes por medio de resonancia magnética funcional durante 10 minutos de estado de reposo con ojos abiertos y punto de fijación. Todas las imágenes fueron preprocesadas, se aplicaron filtros y técnicas estadísticas de estimación para garantizar una mejor detección de señal. Resultados: se obtuvo la matriz de correlación entre las regiones de interés (ROIs) para cada participante y se aplicaron diversas opciones de segregación de conectividad y estimaciones de comunidades. Conclusiones: se aprecian diferencias entre los pa-

trones de conectividad que indican efectos de sincronización diferentes entre grupos de pacientes y controles. En general, los sujetos control muestran patrones concordantes con redes más estables y menos intensas.

Conectividad funcional dinámica en el deterioro cognitivo leve mediante análisis de grafos. Núria Mancho Fora¹, Marc Montalà Flaquer¹, Laia Farràs Permanyer¹, Esteban Gudayol Farré², Geisa Gallardo Moreno³, Daniel Zarabozo Hurtado⁴, Maribel Peró Cebollero¹ y Joan Guàrdia Olmos¹

¹Universitat de Barcelona, ²Universidad Michoacana San Nicolás de Hidalgo, Morelia (México), ³Universidad de Guadalajara (México), ⁴Grupo Rio, Guadalajara (México)

Estudios recientes con población adulta han detectado patrones no estacionarios de conectividad funcional (FC) en *resting-state* fMRI (rs-fMRI). A través del análisis de grafos de las redes de FC, se ha constatado un decremento de la variación en la participación de los nodos *hub* asociado al envejecimiento, así como patrones distintos en la fluctuación del *clustering coefficient* al comparar participantes sanos con pacientes diagnosticados con deterioro cognitivo leve (MCI). Se analizó la conectividad funcional dinámica de las secuencias de 10 minutos de rs-fMRI de 10 pacientes mexicanos con diagnóstico de MCI y 10 controles apareados por edad, género y años de escolarización. Las imágenes funcionales T2* (BOLD) se obtuvieron mediante adquisición en plano transversal (TR = 2 ms, TE = 30 ms, FOV = 230 x 230 x 160, tamaño del vóxel = 2.4 x 2.4 x 4 mm, 29 cortes) con un resonador Philips Ingenia 3.0-T. Para cada participante, se parcelaron las imágenes funcionales mediante el atlas AAL para obtener las series de activación de 90 regiones corticales. Los puntos de cambio en los patrones de las redes de FC entre las 90 regiones se detectaron mediante *spectral clustering* y se validaron con *stationary bootstrap*. Estos puntos de cambio se utilizaron para segmentar la secuencia fMRI de cada paciente y se analizaron las fluctuaciones de indicadores de densidad, integración, segregación y modularidad de las redes de FC. No se detectó efecto de grupo, crudo o ajustado, en los modelos lineales mixtos para el *strength* total de la red, el *strength* medio, la transitividad, la longitud media del camino más corto de la red o el índice *small-worldness* de las redes analizadas. El análisis de la modularidad de las redes indica cambios intra-paciente en la configuración de las comunidades.

SIMPOSIO

Metodología observacional: aplicaciones

Coordina: M. Teresa Anguera y Marta Castañer

Resumen. La observación sistemática no solamente se ha consolidado durante las últimas décadas, sino que ha ampliado considerablemente los ámbitos de aplicación, revelándose como flexible, útil, y de gran rigor, que constituyen sus virtudes fundamentales. Su carácter de método científico, y su perfil focalizado en el estudio de comportamiento espontáneo en contextos naturales la hacen idónea en un amplio espectro de posibilidades investiga-

doras y profesionales del psicólogo. Cada vez se publican más trabajos que utilizan la metodología observacional en revistas de alto factor de impacto y progresivamente se intensifica la cultura que la envuelve, siendo la única metodología posible en un gran número de situaciones. En esta comunicación se presentan seis trabajos, que se refieren a ámbitos educativos, de deporte, de salud ocupacional, y de desarrollo infantil, y se realiza un especial énfasis en la obtención de la información, su gestión y análisis de datos.

Evaluación a través de metodología observacional de la eficacia de un programa para la mejora de la planificación en alumnado con trastorno del espectro autista: diferencias en función de la severidad de los síntomas del alumnado. Marian Acero Ferrero¹, Elena Escolano Pérez¹ y María Luisa Herrero Nivelá¹

¹Universidad de Zaragoza

Bajo los dos dominios en que se agrupan los criterios diagnósticos del trastorno del espectro autista (TEA) en el DSM-5 (APA, 2013), subyacen determinados procesos ejecutivos deficitarios que marcan unos determinados estilos cognitivos y conductuales en las personas con TEA. Uno de los déficits ejecutivos más universales y que más impacto negativo causa en la vida de las personas es el que afecta a la capacidad de planificación. A pesar de su importancia, son escasas las investigaciones en este campo. Esta investigación pretende evaluar la eficacia de un programa para trabajar la planificación en seis niños con TEA de 5 a 12 años. Se establecieron dos grupos de niños en función de la severidad de sus síntomas: Nivel 1 (requiere apoyo) y Nivel 2 (requiere apoyo substancial). El programa constaba de distintas tareas que trabajaban la planificación y que estaban planteados en forma de juego de dificultad creciente. Para evaluar la eficacia del programa se empleó la metodología observacional puesto que es la que más posibilidades ofrecía con niños de estas características en entornos ecológicos. Se utilizó un diseño de baja intervención nomotético, de seguimiento y multidimensional. Los resultados obtenidos tras la aplicación del programa muestran su eficacia puesto que, aunque se observaron ciertas diferencias entre los dos grupos, ambos mejoraron sus procesos de planificación. Esta investigación supone un avance en la evaluación de programas de intervención en funciones ejecutivas en TEA. La metodología observacional proporciona un nuevo enfoque metodológico e incrementa la validez ecológica de la investigación.

Impulsividad y deporte: detección de patrones motrices y de interacción proactivos. Marta Castañer^{1,2}, Oleguer Camerino^{1,2}, Queralt Prat¹, Juan Andueza¹ y Berta Echávarri³

¹INEFC - Universidad de Lleida, ²IRBLLEIDA - Universidad de Lleida, ³Universidad de Navarra

La impulsividad se suele asociar a desórdenes tanto cognitivos como motrices, aunque estudios que relacionen la impulsividad y el deporte son escasos y suelen centrarse en la toma de decisiones en deportes de riesgo. Presentamos dos estudios previos que están en prensa (Castañer et al., 2019; Prat et al., 2019) en que hemos demostrado que, si bien los perfiles de jóvenes y atletas con alta impulsividad muestran poco nivel de interacción y respuestas motrices reiterativas y aisladas, los perfiles con im-

pulsividad moderada muestran capacidad de iniciativa, alto nivel de interacción y respuestas motrices variadas, creativas y pertinentes al contexto motor en que se circunscriben. Un estudio se aplicó a 71 jóvenes adolescentes (M = 18,5 años; DS = 1,72) y el otro estudio a 68 jóvenes atletas (M = 19,8 años; DS = 2,3). Ambos estudios fueron aprobados por el comité ético del Consell català de medicina de l'esport y se obtuvieron los consentimientos informados. En ambos estudios planteamos situaciones con tareas motrices abiertas con relación al uso libre del material, del espacio y de la interacción entre los participantes. Primeramente se administró el SHORT UPPS-P para obtener los perfiles de impulsividad de los participantes. Se utilizó el *Observational System of Motor Skills, Space, Time and Interaction, OSMOSTI* (Castañer et al., 2019) para observar las habilidades motrices específicas, el uso del espacio, del tiempo y de la interacción. Para codificar las conductas del OSMOSTI se usó el programa de codificación LINCE (Gabín et al., 2012) y para obtener los resultados en forma de *T-patterns* se volcaron los datos en el programa THEME (Magnusson et al., 2016). Los resultados mostraron alta similitud en ambos estudios: los participantes sin impulsividad y con impulsividad moderada muestran capacidad de iniciativa, interacción con los demás participantes y variedad de respuestas motrices, aspectos que se relacionan directamente con las funciones ejecutivas de capacidad de inhibición de respuestas, memoria de trabajo y capacidad de adaptación mental que, a nivel de la esfera motora se traducen en: capacidad de parar, de reelaborar o de cambiar respuestas motrices. En cambio, los participantes con alta impulsividad mostraron *T-patterns* pobres puesto que contenían poca interacción y respuestas motrices reiterativas y poco variadas. Concluimos que se debería reconsiderar la etiquetación de la impulsividad y profundizar en los aspectos proactivos de perfiles con impulsividad moderada con relación a que promueven respuestas motrices variadas, creativas y adecuadas a cada situación motriz dada. Consideramos que el OSMOSTI es un Sistema de observación muy adecuado y fácil de aplicar para observar las respuestas motrices, de uso del espacio, del tiempo y de la interacción.

Instrumentos de evaluación e intervención en procesos formativos en salud ocupacional: contribuciones desde la metodología observacional. Marion Portell¹, Anna M. Sene-Mir², M. Teresa Anguera³, Gudberg K. Jonsson⁴ y José L. Losada³

¹Universitat Autònoma de Barcelona, ²Universitat de Vic - Universitat Central de Catalunya, ³Universidad de Barcelona, ⁴University of Iceland

La formación en técnicas de manipulación manual de cargas (MMC) en el lugar de trabajo se espera que contribuya a la reducción de trastornos musculoesqueléticos (TME), pero los resultados que suelen obtenerse no son satisfactorios. El proyecto que denominamos *SsObserWork (Systematic Self-Observation of Work)* parte del análisis, en diferentes áreas no laborales, de aplicaciones de la observación (autoobservación y heterobservación) para crear *feedback* dirigido a modificar comportamiento. Partiendo de este análisis los objetivos del estudio son: (1) desarrollar un instrumento de heterobservación (H-O) para evaluar los cambios en los patrones de comportamiento del trabajador con un nivel de análisis conveniente para derivar una versión para la autoobservación sistemática (S-O) durante entrenamiento en MMC; (2) proporcionar evidencia sobre la fiabilidad del instrumento H-O y sobre la usabilidad/utilidad del instrumento S-O; y (3) proporcio-

nar evidencia sobre el uso del instrumento H-O para crear *feedback* con base al análisis de *T-patterns* y de coordenadas polares. En las diferentes fases de desarrollo de los instrumentos han participado 69 trabajadores del sector industrial. Tanto los resultados del estudio de concordancia entre observadores para el instrumento H-O, como los relativos a la utilidad y usabilidad percibida para el instrumento S-O son muy satisfactorios. Con respecto a la forma de utilizar el instrumento H-O para crear *feedback*, se muestra el gran potencial del análisis de *T-patterns* y de coordenadas polares para ofrecer al trabajador un *feedback* sobre el proceso de MMC y no únicamente sobre el resultado.

“Aprender a aprender” en Educación Infantil: aportaciones educativas desde tres técnicas de análisis de datos observacionales complementarias. Elena Escolano-Pérez¹, M^a Luisa Herrero-Nivela¹ y M. Teresa Anguera²

¹Universidad de Zaragoza y ²Universidad de Barcelona

El currículum de Educación Infantil contempla la competencia de “aprender a aprender”. Ello significa que ya desde edades tempranas el alumnado debe aprender a regular sus propios procesos de aprendizaje, es decir, debe aprender a planificar, supervisar y evaluar la propia conducta para lograr resolver tareas. El objetivo de este estudio fue analizar la estructura secuencial y asociativa de los procesos regulatorios que ponen en marcha niños de 3º de Educación Infantil durante la resolución de una tarea lúdica (un puzle). Más concretamente, interesaba conocer si existían estructuras ocultas diferentes en la utilización de estos procesos regulatorios en los niños que resolvían el puzle y los que no, centrándonos en sus conductas de planificación, supervisión y evaluación. Para ello se usó la metodología observacional, considerada es sí misma como un mixed method dada la integración de elementos cualitativos y cuantitativos que se produce en su proceso. El diseño observacional utilizado fue nomotético/puntual/multidimensional. Se registró la actividad de 44 niños mientras resolvían el puzle individualmente. Los registros obtenidos, una vez controlada la calidad del dato, se analizaron complementariamente desde tres técnicas cuantitativas de análisis de datos observacionales: análisis secuencial de retardos, análisis de coordenadas polares y detección de *T-Patterns*. El uso de estas tres técnicas ofreció un análisis pormenorizado y profundo de la actividad infantil que mostró diferencias en las conductas de planificación, supervisión y evaluación de los niños que resolvieron y no el puzle. Estos resultados tienen importantes implicaciones para la práctica educativa.

Instrumento de observación táctico para esgrima: el instrumento táctico de esgrima (ITE). João José¹, Ana Teresa Campaniço¹, Xavier Iglesias² y Jorge Campaniço¹

¹University of Trás-os-Montes and Alto Douro, ²Universidad de Lleida-Barcelona

Este estudio describe el proceso de desarrollo de un sistema de observación *ad hoc* para analizar el contexto de las acciones tácticas y la interacción entre atletas en esgrima competitiva. La metodología observacional permitió la caracterización y el análisis del comportamiento y los patrones tácticos de juego de esgrima que nos llevó la construcción del Instrumento (ITE). Una aplicación

desarrollada en Visual Studio 2017 nos permite recopilar los datos de forma más pronta y eficiente, y con menos probabilidad de cometer errores, mediante el análisis realizado con el software THEME (versiones 5.0 o 6.0). La observación del video diferido nos permitió identificar cuatro criterios vertebradores y los comportamientos respectivos, que se validaron a través de un manual del instrumento y un cuestionario. Se logró un índice de concordancia intraobservador a través de la kappa de Cohen del 100% y 88% para concordancia interobservador. Se analizaron 62 combates del campeonato mundial de esgrima de Moscú 2015, revelando muchas estructuras complejas que se repiten con el tiempo. El posterior análisis con THEME, *T-patterns*, reveló una ocurrencia en el combate de 37,1% durante el primer período, 46,8% durante el segundo y 12,9% durante el tercero. Los toques de esgrima ocurrieron 54,8% en la zona intermedia de la pista, 38,7% en las zonas finales y 6,5% en las zonas medias. Solo se detectaron 2 patrones temporales completos (*T-patterns*), lo que confirma la naturaleza altamente variable de soluciones de juego. Este instrumento de observación demuestra que puede detectar la ocurrencia de una amplia gama de acciones tácticas altamente dinámicas y rara vez repetibles. Y el posterior análisis mediante THEME demuestra que puede analizar los patrones del esgrimista a lo largo de los múltiples juegos de una o más competiciones. En definitiva, este estudio permite aclarar las preguntas y procedimientos estratégicos del juego para la optimización de los métodos de entrenamiento técnico-táctico.

Una estrategia basada en coordenadas polares y análisis complementario de los datos: El estudio Longitudinal de Coyoacán. Carlos Santoyo¹, M. Teresa Anguera², Andrea Allegro³, Lúgía Colmenares¹ y G. Yamillet Torres¹

¹Universidad Nacional Autónoma de México, ²Universidad de Barcelona, ³Università Degli Studi di Padova

Desde una perspectiva de análisis complementario de los datos, y con base en la metodología observacional, se propone una estrategia cuyo centro son las coordenadas polares para el desarrollo de un análisis complementario de datos del Estudio Longitudinal de Coyoacán, que conjunta análisis molares, *T-Patterns* y coordenadas polares, con la finalidad de extender la comprensión de la persistencia académica en escenarios preescolares. En este estudio se presentan datos de tres casos de participantes preescolares que fueron observados durante tres bloques de observación de seis meses cada uno. Se enfatizan las relaciones sociales entre los niños, su persistencia académica, la influencia de la interacción con la profesora. Se enfatiza la discusión sobre la estabilidad, cambio y diferencias individuales tal y como se recomienda en la Ciencia del Desarrollo.

SIMPOSIO

Meta-investigación: crisis metodológica y promoción de mejores prácticas científicas

Coordina: Dolores Frías Navarro

Resumen. Los trabajos se enmarcan en la autocrítica permanente de la ciencia sobre la fiabilidad y validez de los resultados y abar-

ca reflexiones desde la falta de comprensión y mal uso de las pruebas de inferencia estadística hasta la denominada crisis de la reproducibilidad y la replicación donde se abordan cuestiones como las prácticas de investigación cuestionables, la cultura de la investigación (y los incentivos a los académicos), la alta tasa de resultados positivos en la literatura que contrasta con la escasa potencia estadística de los diseños, el sesgo de publicación, el uso del pre-registro y la ciencia en abierto y las funciones de la meta-investigación dirigidas a cuantificar y visibilizar los efectos de las prácticas de investigación cuestionables e identificar y promover las mejores prácticas científicas. Las aportaciones se centran en ese contexto y ámbito de estudio aportando reflexiones de la situación de la ciencia junto al análisis de posibles estrategias que ayuden a la construcción de hallazgos fiables y válidos. El debate en la psicología (y en otras áreas) está siendo intenso. Si esta situación equivale a una crisis o no es un tema también debatible, pero, lo que sí parece claro es que la comunidad científica está intercambiando ideas y reflexiones que probablemente conduzcan a un cambio importante en las prácticas de investigación y en la cultura de la publicación, mejorando la conducta del investigador y, probablemente, cambiando la forma de plantear, de hacer y de informar sobre la ciencia. Las aportaciones presentan innovaciones metodológicas para analizar y detectar la fiabilidad y validez de los resultados de la investigación.

Encuesta sobre la crisis de la replicación y las prácticas de investigación cuestionable. Un análisis aplicando metodologías severas. Marcos Pascual Soler¹, Jose Perezgonzalez², Juan Pascual Llobell³ y Dolores Frías Navarro³

¹ESIC Business & Marketing School of Valencia, ²Massey University, Business School (New Zealand), ³Universidad de Valencia

En el trabajo se realiza un análisis general de la situación de la denominada crisis de la replicación y las prácticas de investigación cuestionables en el diseño del estudio, en la recogida de los datos, en su análisis e interpretación y en la elaboración del informe. En primer lugar, se presentan los resultados de las investigaciones que han comprobado los usos, malos usos, errores e interpretaciones incorrectas del proceso de inferencia estadística. En segundo lugar, se detallan los análisis de las encuestas de opinión que se han realizado sobre la actual crisis de la replicación en diferentes países. En tercer lugar, se aportan datos de las opiniones de los investigadores españoles sobre dichas cuestiones y sobre cómo mejorar la práctica estadística y el diseño de investigación para ofrecer datos fiables y válidos. Se presenta la aproximación de Deborah Mayo sobre las metodologías severas dirigidas a controlar o eliminar errores metodológicos. El método requiere un procedimiento serio: bien evaluado, con múltiples pruebas capaces de converger en una inferencia dada, razonamientos formales e informales tanto antes como después de obtener los resultados. Cualquier inferencia debe estar basada solo en aquellas hipótesis que han pasado pruebas severas con métodos altamente válidos. Si no se encuentran errores a pesar de haberlos buscado muy bien, se podrá inferir que dichos errores no existen. La convergencia de medidas provenientes de pruebas múltiples permitirá un argumento por coincidencia; es decir, una inferencia mucho más fidedigna y precisa que la que permite cualquiera de las pruebas por separado.

Más allá de la prueba *t*. Contrastes de equivalencia con el procedimiento *Two One-Sided Tests*. Raimundo Aguayo Estremera¹

¹Universidad de Málaga

Cuando se quiere averiguar si dos grupos o tratamientos difieren entre sí, la técnica frecuentista del contraste de hipótesis basado en la significación de la hipótesis nula (en inglés NHST) ha sido preponderante en psicología. En esta técnica la hipótesis nula establece la igualdad entre los grupos. Su rechazo informa de diferencias estadísticamente significativas entre los grupos, asumiendo determinado nivel de error Tipo I. Sin embargo, el NHST no informa del tamaño de las diferencias ni de su importancia. Aún más, tampoco permite dar apoyo a la hipótesis nula, ya que su aceptación no implica su confirmación. Estas características del NHST son frecuentemente olvidadas por los investigadores. Ello ha contribuido al abuso de ésta, al desarrollo de prácticas cuestionables de investigación y a ignorar técnicas complementarias como el tamaño del efecto y los intervalos de confianza. A pesar de estas limitaciones, el NHST permite rechazar estadísticamente diferencias de una magnitud concreta. Se puede rechazar una hipótesis que establezca que la diferencia entre dos grupos es más grande que un intervalo concreto de valores. Cuando se rechaza una hipótesis de este tipo se puede actuar como si la diferencia fuera equivalente a cero, y concluir que es irrelevante en la práctica. Este procedimiento se conoce como contrastes de equivalencia. En este trabajo se presenta el procedimiento *Two One-Sided Tests* para realizar contrastes de equivalencia, con el propósito de ayudar a superar las limitaciones del NHST, sin salirse de la aproximación frecuentista, y a cambiar algunas prácticas cuestionables de investigación.

Estrategia de análisis de datos en dos pasos para controlar la probabilidad de cometer errores Tipo I. José Manuel Caperos¹, Ricardo Olmos² y Antonio Pardo²

¹Universidad la Nebrija, ²Universidad Autónoma de Madrid

Un problema habitual en la inferencia estadística deriva de la realización simultánea de varios contrastes dentro con los mismos datos. La probabilidad de cometer errores Tipo I aumenta respecto del nivel nominal adoptado (normalmente, 0,05) a medida que aumenta el número de contrastes. Hemos evaluado el funcionamiento de una estrategia basada en segmentar la muestra en dos mitades, una para buscar relaciones en un primer paso y otra para contrastar los resultados positivos en un segundo. Entre las estrategias utilizadas en el segundo paso para contrastar los resultados positivos hemos empleado tanto NHST como una aproximación bayesiana que permite aprovechar los resultados del primer paso como información *a priori*. Hemos comparado ambos resultados con los obtenidos sin controlar la tasa de error y controlándola mediante la corrección de Bonferroni. Los resultados muestran que ambas estrategias pueden potencialmente controlar la tasa de error. Sin embargo, mientras que la potencia del contraste por medio de NHST se ve penalizada por la segmentación de la muestra, la aproximación bayesiana presenta un mejor funcionamiento a medida que se incrementa el número de contrastes.

Hacia un análisis de datos exploratorio transparente. Eduardo García Garzón¹

¹Universidad Autónoma de Madrid

La creciente preocupación por conseguir una ciencia psicológica transparente ha generado un alto interés por buscar fórmulas con las que reducir los grados de libertad del investigador (i.e., o cómo las decisiones de los investigadores pueden determinar el resultado de pruebas estadísticas). En investigación confirmatoria, donde existen hipótesis concretas relativas a los efectos de interés, el efecto de estos grados de libertad se ha controlado mediante la implementación de pre-registros e informes registrados. Sin embargo, estas opciones son poco adecuadas en contextos de investigación exploratorios. En estas circunstancias, el análisis multiverso (y otras variantes como el análisis de curvas de especificación) han sido propuestos como alternativas válidas para la evaluación de resultados empíricos y la elaboración de informes científicos transparentes. Pese a ser aplicadas con éxito en multitud de áreas, estas metodologías permanecen relativamente desconocidas para muchos investigadores. Esta presentación introducirá la lógica detrás del análisis multiverso y las diferentes aproximaciones estadísticas para su implementación, así como los beneficios que aportan a los investigadores que decidan utilizarlas. Por último, se revisarán tanto los diferentes recursos disponibles para su uso y su potencial aprovechamiento dentro de la investigación metodológica.

La técnica del análisis de las curvas p : utilidad y limitaciones. Fernando Blanco Bregón¹

¹Universidad de Deusto

Durante las últimas décadas, la psicología (al igual que otras ciencias) se ha visto inmersa en una crisis de credibilidad. En ocasiones, los resultados publicados no pueden ser replicados, o bien sospechamos que son fruto de un falso positivo. En este contexto, urge dotarnos de herramientas que nos permitan diagnosticar los problemas en la literatura científica y evaluar la solidez de la evidencia obtenida. Entre estas herramientas encontramos, por ejemplo, los meta-análisis, pero también otras técnicas más sencillas que pueden ser de gran utilidad. En esta presentación, trataremos la técnica de las curvas p . Por medio de esta técnica, podemos representar la distribución de los p -valores obtenidos en un área de investigación y después analizar las propiedades de esa distribución para diagnosticar algunos problemas, como la publicación selectiva, o algunas formas de *p-hacking*. Finalmente, también es posible obtener a partir de la curva p una estimación aproximada de la calidad de la evidencia recogida. Analizaremos los detalles de la técnica, comentando algunos casos reales en los que se ha empleado y también examinaremos sus limitaciones como una herramienta más al servicio de los investigadores.

El papel del meta-análisis en la crisis de la reproducibilidad de la investigación psicológica. Julio Sánchez Meca¹, Rubén López Nicolás¹, Desirée M^a Blázquez Rincón¹ y Carmen López Ibáñez¹

¹Universidad de Murcia

En los últimos años estamos asistiendo a una crisis de confianza en los resultados de las investigaciones psicológicas (y de otras

disciplinas afines) provocada por los recientes fracasos en los intentos de replicación. Estos fracasos se han observado en dos tipos de investigaciones: (a) en los meta-análisis que se publican rutinariamente integrando los estudios realizados sobre una determinada pregunta de investigación y (b) en diversas iniciativas internacionales dirigidas a intentar reproducir o replicar resultados de investigaciones concretas. Respecto del primer tipo, es preciso considerar los sesgos que pueden sufrir las estimaciones de los efectos, especialmente la sobreestimación del efecto debido a la existencia de sesgo de publicación y a las expectativas del investigador. Respecto del segundo tipo, dependiendo de cómo se lleve a cabo el estudio de replicación, el meta-análisis constituye la metodología idónea para evaluar los resultados. En concreto, el meta-análisis está especialmente indicado como metodología para integrar los resultados de un conjunto de intentos de replicación de una misma investigación. Aun así, su aplicación debe hacerse teniendo en cuenta todas las fuentes potenciales de variabilidad que pueden afectar a los intentos de replicación. En esta comunicación se discuten las ventajas y las limitaciones del meta-análisis como el método para valorar la replicabilidad de los resultados de las investigaciones, así como los criterios objetivos para determinar cuándo se alcanza éxito en un intento de replicación.

COMUNICACIONES ORALES

Validación de escalas

Coordina: Urbano Lorenzo Seva

Recuperación de parámetros en el análisis factorial bajo el sesgo de aquiescencia: un estudio de simulación. Almudena López Cediel Verdú¹ y Carmen Ximénez¹

¹Universidad Autónoma de Madrid

El estudio de los efectos del sesgo de aquiescencia en las propiedades psicométricas de los test de personalidad ha recibido una considerable atención en la literatura previa. El presente trabajo tiene como objetivo analizar las consecuencias de cometer errores en relación al "no balanceo" o desequilibrio del contenido de los ítems en el contexto del análisis factorial confirmatorio (AFC). Para ello, se han llevado a cabo varias simulaciones en modelos AFC de tres factores. En base a las recomendaciones de estudios previos (Ferrando, Condon y Chico, 2004), se ha trabajado con un modelo teórico que incluye un factor de aquiescencia y dos factores de contenido. Las condiciones del estudio incluyen: (1) el tipo de balanceo de la escala, (2) la magnitud de las saturaciones en el factor de aquiescencia y (3) el nivel de sesgo de aquiescencia en los ítems. Se ha estudiado la recuperación de parámetros y la bondad de ajuste, tanto bajo un modelo que incluye el factor de aquiescencia como bajo otro modelo erróneo que no contempla dicho factor. En la misma línea que en investigaciones previas (Billiet y McCleden, 2010; Ferrando y Seva, 2010; Savalei y Falk, 2014), los resultados apuntan que se debe analizar el sesgo de respuesta para hacer evaluaciones más fiables. Para ello, se recomienda ser minuciosos en el diseño de los test no solo redactando la mitad de los ítems como inversos, sino buscando un buen equilibrio a nivel de contenido entre los ítems directos y los inversos.

Imagen corporal, redes sociales y satisfacción personal. Ignacio Cuesta López¹ y Eduardo García Cueto¹

¹Universidad de Oviedo

Introducción: la satisfacción con la imagen corporal es una de las variables más importantes a la hora de relacionarse con el entorno social. Esta percepción influye en la satisfacción con la vida social, en el uso de redes sociales e incluso en el consumo de alcohol. En el presente estudio se explorará cómo es la relación entre la satisfacción con la imagen corporal y las variables mencionadas, así como las posibles diferencias entre sexos y edad. Método: se estiman los parámetros psicométricos de los test utilizados, todos contruidos *ad hoc* para la presente investigación. Se estudian las diferencias en función del sexo y la edad en las distintas variables y se propone un modelo matemático mediante ecuaciones estructurales para explicar la relación entre las diversas variables. Resultados: las pruebas utilizadas se han mostrado fiables y con garantías de validez factorial, así como buena validez convergente y divergente. Se encontraron diferencias entre hombres y mujeres en cuanto a la satisfacción con la imagen corporal, aunque no en el resto de las variables. Al comparar las variables estudiadas en función de la satisfacción con la imagen corporal se hallaron diferencias en todas las variables. Las correlaciones entre los test empleados resultaron ser significativas. Conclusiones: los test presentan buenas propiedades psicométricas y las variables del estudio están claramente relacionadas y es posible establecer un modelo matemático explicativo.

¿Limitaciones o cambio de objetivos? Lecciones aprendidas comparando dos opciones para reportar medidas con baja consistencia interna. Ariadna Angulo Brunet¹ y Carme Viladrich¹

¹ Universitat Autònoma de Barcelona

Tal como se recomienda en los textos normativos, actualmente está consolidada la práctica de estudiar la consistencia interna de las medidas basadas en cuestionarios antes de usarlas para comprobar la hipótesis principal de un estudio. En este trabajo nos planteamos la disyuntiva de qué hacer cuando, una vez obtenidos todos los datos del estudio principal, se observa que la consistencia interna de algunas medidas no es aceptable. En algunas publicaciones, la baja consistencia interna se trata como una limitación del trabajo y se invita a tener precaución al interpretar los resultados principales; mientras que otra opción consiste en modificar los objetivos del trabajo para aportar datos a la clarificación del problema psicométrico detectado. El dilema se ilustra con un caso práctico cuyo objetivo inicial fue comparar, utilizando *path analysis*, la red nomológica entre los constructos registrados mediante dos cuestionarios, uno formado por medidas de un solo ítem y otro formado por medidas de múltiples ítems. Para ello se dispuso de respuestas dicotómicas obtenidas en una muestra de 128 mujeres y hombres mayores de 60 años que contestaron ambos cuestionarios. Aunque el ajuste del *path analysis* con las medidas de múltiples ítems resultó adecuado y con las medidas de un solo ítem fue parecido aunque más discutible ($[df] = 6,44[3]$, $p = 0,09$, $CFI = 0,99$, $TLI = 0,89$, $RMSEA [90\% IC] = 0,08 [0 - 0,17]$), la consistencia interna de tres de las siete medidas de múltiples ítems no fue aceptable (omega categórica; rango 0,57 a 0,88). A la luz de estos resultados se discuten las implicaciones de utilizar variables calculadas a partir de la suma

o promedio de otras variables en un *path analysis* cuando no se ha alcanzado un umbral mínimo de fiabilidad.

Roles de género: ¿Existen en el siglo XXI? Pelayo Montes Álvarez¹, Francisco José González Primo¹ y Eduardo García Cueto¹

¹Universidad de Oviedo

Antecedentes: la medición de los roles de género supone un problema en sí mismo debido a la gran dependencia de circunstancias psicosociales en el momento de realizar dicha medición. Haciendo de este constructo algo variable en el tiempo, pues hombres y mujeres llegan a adoptar roles previamente asociados al sexo opuesto. Método: el principal objetivo de este trabajo ha sido la elaboración y exploración de las propiedades psicométricas de un test de masculinidad y feminidad, así como el análisis de las diferencias entre hombres y mujeres de una muestra compuesta por 612 personas en base a su sexo, edad, tipo de estudios y zona geográfica de residencia en España. Para ello se administró una batería compuesta por ítems de masculinidad, feminidad, amabilidad, motivación, tolerancia al estrés, apertura a la experiencia, toma de riesgos y locus de control interno. Resultados: se obtuvo un test válido y fiable que mide masculinidad y feminidad y que se adapta a un modelo de dos factores. Los hombres obtuvieron puntuaciones significativamente más altas que las mujeres en masculinidad y tolerancia al estrés, mientras que las mujeres obtienen puntuaciones significativamente más altas que los hombres en amabilidad, motivación y locus de control interno. A su vez, existen diferencias en base a la edad y el tipo de estudios, pero no en cuanto a la zona geográfica de residencia. Conclusiones: los hombres siguen asumiendo roles masculinos en mayor medida que las mujeres, las cuales asumen roles femeninos en mayor grado que los hombres. Las actitudes y conductas propias de un sexo u otro, puede variar con la edad y el nivel de estudios, ya que las mujeres jóvenes y con estudios universitarios puntúan más alto en masculinidad que aquellas más mayores y con niveles de estudio bajos.

Estudio de las propiedades de los bancos de ítems y los métodos de estimación en Test Adaptativos Informatizados en contextos de falseamiento en test. Susana Sanz Velasco¹, Rodrigo Schames Kreitchmann¹ y Pablo Nájera Álvarez¹

¹Universidad Autónoma de Madrid

Introducción: actualmente, los candidatos a un proceso de selección pueden completar ciertas pruebas de forma telemática sin vigilancia, por lo que podrían utilizar métodos fraudulentos para aumentar su puntuación. Se han propuesto algunos métodos de detección para estos casos, siendo uno de los métodos más usados el de Guo y Drasgow (2010), consistente en realizar un contraste estadístico entre la prueba telemática y una posterior vigilada de verificación (VT). En los últimos años, estos métodos se han aplicado con test adaptativos informatizados y parece que la función de información del banco de ítems y el procedimiento de estimación afectan a la tasa de errores Tipo I y a la potencia. Objetivo: analizar cómo afecta la función de información del banco de ítems al funcionamiento de los métodos de estimación WLE, MAP y EAP cuando son aplicados en contextos de falseamiento, a fin de seleccionar el que mejores resultados proporcione dependiendo del objetivo de la aplicación. Método: se simulan dos bancos, uno con ítems que se distribuyen normalmente y otro unifor-

memente. Para cada uno de ellos se realizan tres simulaciones (una por cada procedimiento de estimación) con 25.000 simulados honestos y 300.000 falseadores y valores en el rasgo de -3 a 3 a intervalos de 0,25. El UIT cuenta con 30 ítems; el VT, con 10 ítems. Los honestos contestan a ambas en función de su nivel real de rasgo. Los falseadores son sinceros en el VT, pero se simulan distintos niveles de falseamiento para la primera fase. Después, se realizan los contrastes de hipótesis y se obtienen la tasa de errores Tipo I y la potencia. Resultados esperados: se espera que los procedimientos bayesianos proporcionen los menores errores de estimación y el menor sesgo para el banco normal, y WLE un menor sesgo en la condición de banco uniforme.

cdcatR: Un paquete de R para la aplicación de test adaptivos informatizados basados en modelos de diagnóstico cognitivo. Miguel A. Sorrel¹ y Francisco José Abad¹

¹Universidad Autónoma de Madrid

Los modelos de diagnóstico cognitivo son modelos de clase latente confirmatorios con implicaciones importantes para educadores y otros profesionales. Permiten la clasificación en varias variables latentes discretas denominadas típicamente atributos. Recientemente, se ha trabajado en la aplicación computarizada de pruebas adaptativas basadas en estos modelos. En concreto, se han propuesto varias reglas de selección y se han estudiado las condiciones de aplicación de estas pruebas mediante estudios de simulación, lo que ha resultado en un alto número de publicaciones durante la última década. Las aplicaciones empíricas son, sin embargo, aún escasas. Hasta la fecha los investigadores han empleado códigos ad-hoc escritos en distintos lenguajes de programación. Para facilitar la investigación y la aparición de aplicaciones empíricas en esta área, hemos desarrollado el paquete de R *cdcatR*. El objetivo de la presente comunicación es ilustrar las diferentes funciones incluidas en este paquete. A través de las diversas funciones, *cdcatR* permite trabajar con los modelos de diagnóstico cognitivo más comunes, aplicar de manera adaptativa bancos de ítems calibrados según estos modelos, comparar el funcionamiento de diferentes pruebas adaptativas en términos de fiabilidad y uso del banco de ítems y simular bancos de ítems según la longitud del banco, complejidad de la matriz-Q y calidad de los ítems para los distintos modelos mencionados. Se incluyen para ello distintas reglas de selección y para el control de la exposición. El procesamiento en paralelo permite una alta eficiencia computacional. En la ilustración se mostrarán los resultados obtenidos en distintas publicaciones de los miembros de nuestro grupo de investigación.

COMUNICACIONES ORALES

Metodologías

Coordina: Jesús M. Alvarado

Construir desde el territorio, saberes colectivos sobre el uso de las plantas en habitantes insulares. Eduardo Aburto¹ y José Flores Cárdenas¹

¹Universidad de los Lagos

Desde una perspectiva cualitativa y con enfoque etnometodológico, se buscó estudiar los elementos particulares en la vida coti-

diana de los habitantes de un territorio insular en la región de los Lagos, Chile, a través del saber popular sobre los usos y prácticas vinculadas al contacto con diversas especies botánicas; en particular, con los ámbitos referidos a: (1) la alimentación, (2) el uso terapéutico y/o medicinal, (3) la dimensión tecnológica en el ámbito de la creación de objetos y herramientas así como de materiales e insumos para la artesanía, (4) una cultura inmaterial en su relación a componentes místicos, festivos, mágicos y míticos, y (5) la inclusión del paisaje o manejo integrado mediante la presencia de bioindicadores. Todo lo anterior, orientado a conocer los significados desde las experiencias vitales de los propios actores, debido a que resulta importante el acceder desde estas perspectivas, a las acciones locales que permiten enfrentar, desde los aspectos que configuran el mundo interno de la comunidad insular, los desafíos de la expansión y transformación territorial. La estrategia que permitió cumplir con el propósito establecido consistió en un abordaje etnometodológico a través de la observación y familiarización del contexto, el cual permitió acceder a los cultores de estos saberes, para luego, elaborar un mapeo colectivo, el cual dio paso a la realización de entrevistas semiestructuradas que versaron sobre los aspectos específicos y asociados a la adquisición y reproducción de este conocimiento etnobotánico. Luego de la producción de los datos, se inició un proceso de análisis orientado hacia la comprensión de los elementos que permiten a los sujetos la trascendencia del conocimiento de este mundo; finalizando con la creación de un mapa que contiene los saberes categorizados y localizados en el territorio para uso de la comunidad.

Un análisis narrativo de las emociones morales sobre el pasado nacional en estudiantes españoles. Margarita Gómez Márquez¹ y César López Rodríguez²

¹Universidad Europea de Madrid y ²Universidad Autónoma de Madrid

La investigación narrativa es uno de los subtipos de metodologías cualitativas más empleadas junto a la teoría enraizada, la etnografía, el estudio de casos o la fenomenología. Su objetivo fundamental es establecer asociaciones entre aquello que la persona cuenta y el significado que le otorga a dicha narración. Esta metodología es especialmente útil para el análisis de narrativas nacionales, relatos en los que los individuos hacen referencia al pasado de su nación y en los que se conecta este pasado con el presente de la misma (López, Carretero, Rodríguez-Moneo, 2014). La investigación narrativa permite dilucidar no sólo cómo las personas entienden dicho pasado sino también cómo se sienten en relación al mismo, es decir, si se dan o no emociones colectivas. El objetivo de esta presentación es analizar, por medio de investigación narrativa, cómo una muestra de 56 estudiantes universitarios españoles entiende el llamado "Descubrimiento de América" en una serie de entrevistas semi-estructuradas e individuales. Se detallará en la exposición el proceso de categorización inductiva y deductiva que se llevó a cabo para clasificar el contenido de las respuestas dadas por los participantes. Asimismo, se discuten las implicaciones metodológicas de este trabajo analizando la utilidad de la investigación narrativa para analizar la asociación entre identificación nacional y emociones grupales y morales.

Factores asociados a trayectorias académicas de estudiantes con orientación a la educación superior en Puerto Montt. Damaris Opazo Vega¹, Sandra Sandoval Barrientos¹, Patricio Quezada Carrasco¹, Ana Mujica Hoevelmayer¹ y Eduardo Aburto¹

¹Universidad de los Lagos

El presente estudio se enmarca en la psicología educacional, trabajando con estudiantes de educación secundaria con orientación a la educación superior para determinar los factores que inciden en una trayectoria académica exitosa. Para el presente estudio se utilizarán fuentes primarias y secundarias; como fuentes secundarias se emplearán bases de datos institucionales con información sociodemográfica y socioeconómica de los estudiantes, desempeño académico en educación secundaria y puntajes para la selectividad en educación superior. Como fuente primaria se utilizará una escala construida *ad hoc* denominada "Factores asociados a trayectorias académicas de estudiantes con orientación a la educación superior en la comuna de Puerto Montt". A partir de los datos, se construirá una matriz de doble entrada variables por sujetos, la que será depurada para conservar aquellos sujetos que presentan al menos el 90% de la información de las fuentes secundarias y que abordaron el 100% de las preguntas de la escala construida. En una primera etapa se realizará un estudio descriptivo de la totalidad de las variables, conservando aquellas que presenten una dispersión distinta de 0 y que posean niveles de asimetría y curtosis adecuados. Con los datos de la escala se realizará el mismo análisis descriptivo, manteniendo aquellos ítems que conserven los valores psicométricos estándar. Con la base de datos ya depurada se procederá a realizar un análisis factorial exploratorio para determinar si existen sub-escalas o el instrumento mide una única dimensión. Con las fuentes secundarias se realizará un análisis de conglomerados jerárquicos con todas las variables que presenten el nivel de medición adecuado. A partir de las variables conservadas y mediante modelos de ecuaciones estructurales se determinará la posible existencia de un modelo explicativo de las causas que favorecen el desempeño académico y la permanencia en la educación superior.

Construcción y validación de una escala al sujeto de formación integral basada en competencias. Damaris Opazo Vega¹, Eugenia Avendaño¹, Valentina Vera¹ y Gabriela Mora¹

¹Universidad de los Lagos

El presente estudio tuvo por objetivo construir y validar una escala al sujeto para medir el nivel de desempeño de las competencias de Formación Integral de los y las estudiantes de la Universidad de los Lagos, Chile. Se tomó como base para la construcción de los ítems la Propuesta Institucional de Formación Integral, en la cual se describieron los niveles de las competencias. Posteriormente, mediante entrevista semi-estructurada a expertos de la institución que imparten las asignaturas ligadas a la competencia, se determinaron los niveles de ejecución operado para cada uno de los niveles de competencia en los que respecta al saber, saber hacer y saber ser. A partir de la descripción de cada nivel, se generó una matriz de evaluación, que se utilizó como base para generar los ítems que componen la escala. La primera versión de la escala compuesta por 45 ítems se aplicó a una muestra intencionada de 185 Estudiantes de la Universidad de los Lagos, Campus Osorno, matriculados en las carreras de Educación Diferencial, Educación Parvularia, y Nutrición y Dietética que se encontraban

cursando asignaturas en el periodo académico 2018, dado que son carreras reformuladas que han incorporado el nuevo modelo de formación integral basado en competencias. A partir de la aplicación se construyó una matriz variables por sujetos que se utilizó para los análisis. En una primera etapa se realizó un análisis descriptivo, eliminándose aquellos ítems con varianza 0 y aquellos que presentaban una curtosis o asimetría alta. Para determinar la confiabilidad por consistencia interna se utilizó el alfa de Cronbach en cada una de las dimensiones de la prueba. Para la validez de constructo se utilizó un análisis factorial. En una segunda etapa se espera obtener mediante una muestra representativa las normas para la interpretación de la escala.

Validación de una escala de caracterización de emprendedoras. Miguel Ripoll¹, Cristian Bucarey¹, Camila Martínez¹, Rodrigo Huanel¹ y Eduardo Aburto¹

¹Universidad de los Lagos

El presente estudio tuvo por objetivo construir una escala que busca caracterizar el perfil emprendedor de las mujeres que participan de la escuela de emprendimiento "Mujer, Asociatividad y Emprendimiento", impartido por el Servicio Nacional de la Mujer y Equidad de Género, en la Región de Los Lagos, Chile. Se tomó como base para la construcción de los ítems las dimensiones propuestas en las teorías de "Perfil emprendedor" (Alcaraz) y "Liderazgo transformacional" (Montero, 2006), considerando las características sociodemográficas de la población en estudio. Se utilizaron dos muestras, una compuesta por 10 alumnas de la generación 2018 para descentrar el instrumento y una segunda muestra de 60 sujetos formada por alumnas de primer ($n = 30$) y tercer ($n = 30$) curso del programa en el año 2019 seleccionadas de manera aleatoria. A la primera muestra se les aplicó una versión preliminar y una vez que terminaron la aplicación se les pidió revisar los ítems y que reportaran aquellos que les costó comprender, a partir de esto se construyó una nueva versión la que se utilizó para realizar el análisis psicométrico. A partir de la escala final resultante se procedió a correlacionar los resultados con los datos demográficos para realizar una caracterización de los distintos tipos de emprendedora que asisten a estos programas como una manera de focalizar los contenidos en las características de las beneficiarias y las áreas del emprendimiento que se busca desarrollar.

La construcción de exámenes de elección múltiple en el ámbito universitario. Sara Villadonga Barril¹ y Marcelino Cuesta Izquierdo¹

¹Universidad de Oviedo

En el ámbito académico es importante realizar una evaluación adecuada del rendimiento de los alumnos. En el contexto universitario es bastante frecuente emplear para esto pruebas de elección múltiple. En este trabajo se pretende analizar en qué medida los profesores que emplean este tipo de prueba siguen algunas de las pautas más comunes citadas en la literatura sobre la construcción de este tipo de ítems. Para ello se ha construido un cuestionario *ad hoc* en el que se pregunta sobre el uso de estas directrices. La muestra ha consistido en 376 docentes universitarios contactados vía correo electrónico, con una edad media de 47 años de los cuales un 53,5% son mujeres. Los resultados nos permitirán describir las prácticas habituales (buenas y malas) en la

construcción de estos instrumentos de evaluación académica tan populares y ver si existen diferentes perfiles en función de ciertas características de los docentes como el sexo, rama, experiencia docente, etc.

COMUNICACIONES ORALES

Validación de Escalas: Aplicaciones

Coordina: África Borges del Rosal

Validación de criterio y de constructo del SCIP-S del funcionamiento cognitivo en adultos mayores mexicanos. Ana L. González Celis¹, Edgardo Ruiz Carrillo¹, M. Jesús Cruz Peralta¹, Violeta A. Mendoza Madrigal¹, Juana Gómez Benito² y Georgina Guilera²

¹Universidad Nacional Autónoma de México, ²Universidad de Barcelona

Para entender el envejecimiento saludable es fundamental tener en cuenta que la capacidad intrínseca medida por el funcionamiento físico y mental tiende a disminuir con la edad, y que las opciones de vida e intervenciones en diferentes momentos del ciclo de vida determinarán la trayectoria de cada individuo. Es de gran relevancia la medición del funcionamiento cognitivo en adultos mayores debido a que en dicha población se encuentra una alta prevalencia de trastornos cognitivos y su importancia en la detección oportuna es contribuir en su prevención. El Mini-Mental de Folstein (1985) es considerado el instrumento estándar de evaluación cognitiva, el cual ha sido empleado en población geriátrica y validado a población española por Lobo y colaboradores (2002); a diferencia del SCIP-S (Screen for Cognitive Impairment in Psychiatry-Spanish) en su versión validada española por Pino y colaboradores (2014), se ha utilizado para población psiquiátrica, sin embargo es un instrumento breve, y se propone examinar las propiedades psicométricas del SCIP-S en una muestra de adultos mayores mexicanos. Por ello, se aplicó una batería de 5 instrumentos: Ficha de datos sociodemográficos, SCIP-S, Mini-Mental, Cuestionario de calidad de vida WHOQOL-OLD validado a población anciana mexicana (González Celis y Gómez Benito, 2013), y la Escala de Bienestar Subjetivo, validada a población española (Díaz et al, 2006). Se obtuvo alfa de Cronbach ($> 0,85$, $p < 0,05$) y una validación de criterio, r de Pearson ($p < 0,01$) con el Mini-Mental. Un análisis factorial exploratorio con cinco dimensiones, aprendizaje verbal inmediato, memoria de trabajo, fluidez verbal, aprendizaje verbal diferido, y velocidad de procesamiento. Se concluye que los datos de validez son los suficientemente aceptables para emplearse el SCIP-S en una muestra geriátrica y sin influencias culturales en México, ya que parece que los datos de validez son comparables con criterios internacionales.

Personalidad en adolescentes de Latinoamérica y España: revisión rápida de las propiedades psicométricas de instrumentos de evaluación. Nadia Soria Miranda^{1,2}, José Luis Padilla², M. Carmen Navarro González² y Carolina Díaz Piedra²

¹Universidad de Guayaquil, ²Universidad de Granada

La personalidad en la adolescencia incorpora elementos complejos para su evaluación, rasgos que pueden confundirse con alteraciones, escasa colaboración o dar una impresión positiva. Esto

demanda el uso de instrumentos de calidad. En España, se suelen adaptar al español instrumentos de evaluación generalmente del inglés (Muñiz, Fernández, Fonseca, Campillo y Peña, 2011). En Latinoamérica, principalmente Argentina, Chile, Colombia y México, han aportado estudios de adaptación y validación de tests clásicos de la personalidad (Wenger y Andrés-Pueyo, 2016). El objetivo de la revisión es conocer los estudios psicométricos sobre instrumentos de evaluación de la personalidad adolescente, que sirvan de referencia a profesionales, para una evaluación psicológica objetiva de las características y comportamientos en la adolescencia. Se realizó la búsqueda de estudios publicados entre 2010 y 2019 en cuatro bases de datos. Para la revisión de referencias y extracción de datos se utilizó el software COVIDENCE, con la participación de varios revisores. En la revisión de propiedades métricas se utilizó una versión adaptada del Cuestionario de Evaluación de Tests-Revisado (CET-R). Se encontraron 1.037 referencias y 36 estudios se analizaron a texto completo. Esta revisión permite valorar las evidencias de validez y la calidad de los procesos de adaptación de los instrumentos, desde las recomendaciones de los Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, y National Council on Measurement in Education, 2014), y las Guidelines for Translating and Adapting Tests (International Test Commission, 2017). El uso de software especializado aporta transparencia al proceso de inclusión de estudios, registra las decisiones de los revisores y las mantiene visibles a los miembros del equipo para la resolución de conflictos. También se sugieren líneas de investigación para mejorar la calidad de las adaptaciones de tests y cuestionarios entre diferentes grupos lingüísticos y culturales.

Equivalencia métrica de una escala de sobrecualificación a través de distintos países: consideración de algunos factores relevantes. Ana María Hernández Baeza¹, Inés Tomás Marco¹, Zahide Karakitapoglu-Aygun², Berrin Erdogan³ y Vicente González¹

¹Universidad de Valencia, ²Bilkent University, ³Portland State University

La sobrecualificación es un problema importante a la hora de comparar la calidad del trabajo en distintos países. Entre las escalas propuestas para medir sobrecualificación, la propuesta por Maynard, Joseph y Maynard (2006), de 9 ítems, es la más utilizada en la actualidad (Erdogan, Bauer y Karaeminogullari, 2017). Sin embargo, antes de comparar los niveles de sobrecualificación de distintos países, y/o las relaciones entre sobrecualificación y un número relevante de variables a través de distintos países, es necesario mostrar que las medidas tienen equivalencia métrica (o, en otras palabras, están libres de DIF). Además, en algunas ocasiones, los resultados sobre dicha equivalencia pueden depender de factores como el número de categorías de respuesta empleadas por los distintos países, o las diferencias existentes en variables sociodemográficas relevantes. Teniendo en cuenta estas consideraciones, el objetivo principal del presente estudio es evaluar la equivalencia métrica de la escala de sobrecualificación de Maynard et al (2006) a través de dos países: España y EEUU. Asimismo, se ilustra cómo las conclusiones pueden depender de distintos factores. La muestra de estudio consiste en 440 estudiantes y graduados recientes de dos universidades (201 de una universidad española y 239 de una universidad americana), em-

pleados en el momento de la encuesta. La equivalencia métrica se evalúa mediante análisis factorial confirmatorio multigrupo con estructura de medias y covarianzas latentes (AFC-MACS) (con y sin variables covariadas), a través de Mplus 8 (Muthén y Muthén, 1998, 2015). Los resultados permiten ilustrar cómo los resultados y conclusiones pueden variar dependiendo de (a) el modo en que se equiparan distintas alternativas de respuesta a través de los grupos y (b) si se controlan o no variables potencialmente relacionadas con la sobrecualificación, como el sexo y la edad.

Adaptación española de la Triada Oscura. Francisco José González Primo¹, Pelayo Montes Álvarez¹ y Eduardo García Cueto¹

¹Universidad de Oviedo

El objetivo del presente trabajo fue diseñar una versión adaptada al español del instrumento *Short Dark Triad* que evalúa la Triada Oscura conformada por tres rasgos de la personalidad: psicopatía, narcisismo y maquiavelismo. La necesidad de la presente validación fue el hecho de no contar con este instrumento en castellano dificultando la evaluación de la Triada Oscura. Para cumplir con este cometido se cumplieron las normas de la *International Test Commission* (ITC). Para poder presentar evidencias de validez divergente y convergente se aplican además los siguientes cuestionarios: Cuestionario de Clase Social (CCS), Inventario Neo Reducido de Cinco Factores (NEO-FFI), la Escala de Maquiavelismo (MAQ-IV), *Hypersensitive Narcissism Scale* (HSNS) y *Self-reported Psychopathy Inventory III-Short Form* (SRP-III-SF). La aplicación se llevó a cabo "on line" a una amplia muestra de toda España y de edades comprendidas entre los 18 a los 70 años. El muestreo fue de "bola de nieve" y se propagó utilizando las redes sociales: WhatsApp, Facebook, Twitter y el correo electrónico.

Estudio piloto sobre las propiedades psicométricas de la versión española del Inventario de "Autointerés e Interés por otros". Raimundo Aguayo Estremera¹, Manuel Correa Gómez² y Alberto Ruiz Villaverde²

¹Universidad de Málaga, ²Universidad de Granada

El Inventario de "Autointerés e interés por otros" pretende medir, por una parte, el comportamiento autointeresado y, por otra, el comportamiento de interés hacia otras personas. Ha sido desarrollado en dos versiones, una para adultos y otra para estudiantes universitarios. En la versión de estudiantes, cuenta con 19 ítems, de los cuales ocho evalúan conductas autointeresadas y otros ocho, conductas de interés hacia otras personas. El formato de respuesta es tipo Likert y evalúa el grado de acuerdo en una escala de siete puntos, todas con etiqueta verbal, que va desde 1 hasta 7 (completamente de acuerdo). El objetivo de este estudio es analizar el funcionamiento de la adaptación española de este inventario en estudiantes universitarios. Tras el proceso de adaptación a la lengua y cultura española mediante procedimientos analítico-rationales, se usaron procedimientos empíricos para la obtención de evidencias sobre algunas propiedades psicométricas. Después de realizar unos primeros análisis y haber modificado ciertos aspectos que fueron insatisfactorios, en este trabajo se presenta el estudio piloto que fue realizado con una muestra de 165 estudiantes universitarios. Según los resultados, todos los ítems disponen de índices de discriminación adecuados, por enci-

ma de 0,30. Las medias fueron moderadamente elevadas, en el sentido de los constructos, y las desviaciones típicas resultaron en torno a uno. En cuanto al inventario completo, se obtuvieron índices de fiabilidad adecuados, por encima de 0,70 en todas las subescalas. Para obtener evidencias de validez de constructo se utilizó el análisis factorial exploratorio. Se pusieron a prueba varios modelos, tanto esperables según investigaciones previas como plausibles según los datos obtenidos. Los resultados indicaron que la mejor solución es la de tres dimensiones.

COMUNICACIONES ORALES

Validación de Escalas: procedimientos

Coordina: M^a Dolores Hidalgo

Recursos para calcular los coeficientes alfa y omega de fiabilidad de consistencia interna. Carme Viladrich¹, Ariadna Angulo Brunet¹ y Eduardo Doval¹

¹Universitat Autònoma de Barcelona

En los últimos años han sido muchos los esfuerzos para facilitar la estimación correcta del coeficiente de fiabilidad de consistencia interna. Por una parte, se ha alcanzado un amplio consenso sobre las condiciones de aplicación de los coeficientes alfa y omega y, por otra, se ha difundido su uso apropiado no solo en revistas técnicas sino también en otras con orientación aplicada. Por ello, los efectos del bien conocido divorcio entre la psicometría básica y sus aplicaciones podrían considerarse paliados en este ámbito. Sin embargo, faltan todavía algunos pasos para conseguir la plena reconciliación. El artículo de Viladrich, Angulo-Brunet y Doval se posicionó en la línea de facilitar la incorporación rutinaria de este conocimiento al trabajo de autores y editores que usan medidas basadas en cuestionarios del ámbito de las Ciencias Sociales y de la Salud. En la publicación citada se dio un primer paso en esta dirección desarrollando cuatro casos paradigmáticos, así como la sintaxis necesaria para tratarlos completamente utilizando el programa de uso libre R. En el presente trabajo pretendemos dar un paso más en la misma dirección compartiendo los recursos que hemos desarrollado posteriormente en formatos diversos como los programas comerciales Mplus y Stata. Tanto los conjuntos de datos como las diversas sintaxis desarrolladas hasta el momento presente están disponibles en el enlace permanente <https://ddd.uab.cat/record/173917> (<https://ddd.uab.cat/record/173917>) y en esta presentación explicamos sus bases y damos indicaciones prácticas para su uso correcto.

Uso de la regresión logística en la detección del DIF en instrumentos de evaluación educativa: una revisión sistemática de los últimos 5 años. M^a Dolores López Martínez¹, María Dolores Hidalgo Montesinos¹ y Juana Gómez Benito²

¹Universidad Murcia y ²Universidad de Barcelona

Los estándares para la medición educativa y psicológica recogen la importancia de considerar el análisis del funcionamiento diferencial del ítem (DIF) como una fuente de evidencia de validez. Conscientes de la necesidad de obtener evidencias de la validez de los instrumentos de medida que se emplean en la práctica profesional, se encuentran cada vez, con mayor frecuencia, trabajos

en los que se evalúa la presencia de DIF en éstos, respecto a diferentes versiones idiomáticas del instrumento o a diferentes variables de agrupamiento tales como el sexo de los examinados. Existen numerosos procedimientos y técnicas de análisis para detectar el DIF (Hidalgo y Gómez, 2010; Sireci y Rios, 2013), aunque el análisis de Regresión Logística (RL) es por su flexibilidad, relativa simplicidad y eficacia en la detección de diferentes tipos de DIF uno de los procedimientos recomendados. El objetivo del presente trabajo es realizar una revisión sistemática de aquellos estudios que usan RL como técnica de detección del DIF en instrumentos de evaluación, habitualmente empleados en la práctica profesional, de tipo aptitudinal y de rendimiento académico. La búsqueda de trabajos se realizó a través de ERIC, PsychInfo y Web of Science, seleccionando para este estudio artículos que usan RL como técnica de detección del DIF en instrumentos de evaluación de tipo aptitudinal y de rendimiento académico. Los artículos fueron seleccionados por dos investigadores del estudio; los casos de discordancia se resolvieron por consenso. Un total de 23 artículos que usaron RL fueron identificados. Se expondrán los principales resultados obtenidos en los trabajos revisados, así como las diferentes variables consideradas en la aplicación de la RL (tamaño de los grupos comparados, método de RL, procedimiento de purificación del test, matching criterion, tamaño del test, porcentaje de ítems con DIF o formato de ítems). Se establecerán las fortalezas y debilidades de los estudios aplicados de DIF usando RL.

Procedimientos para la evaluación diagnóstica en contextos educativos. Pablo Nájera Álvarez¹, Susana Sanz Velasco¹ y Rodrigo Schames Kreitchmann¹

¹Universidad Autónoma de Madrid

Introducción: en los últimos años se han desarrollado algunos modelos estadísticos que facilitan la aplicación de la evaluación educativa diagnóstica o formativa. Entre ellos están los modelos de diagnóstico cognitivo (MDC), que son modelos multidimensionales de rasgos latentes categóricos, llamados habitualmente atributos. Los MDCs han experimentado un gran desarrollo teórico, aunque presentan ciertas limitaciones prácticas que dificultan su aplicación en contextos educativos. Una de ellas es la recomendación de trabajar con muestras numerosas para garantizar una estimación fiable de los parámetros. Otra limitación es la necesidad de tener cierto nivel de formación en el funcionamiento de dichos modelos y del software necesario para su aplicación. Todo esto puede generar rechazo en los docentes de cara a usar los MDCs en contextos educativos, que pueden optar por realizar la evaluación diagnóstica basándose en las puntuaciones directas. **Objetivo:** la presente investigación pretende estudiar el rendimiento de los MDCs (paramétricos y no paramétricos) bajo condiciones realistas de contextos educativos, y compararlo con las puntuaciones directas por atributos. **Método:** se realizó una simulación exhaustiva de tipo Monte Carlo, empleando como factores el tamaño muestral, el número de atributos, la calidad de los ítems y la calidad de la matriz-Q, entre otros. La variable dependiente fue la precisión en la clasificación de los atributos. **Resultados y conclusiones:** los MDCs presentaron, en la mayoría de las condiciones evaluadas, un rendimiento adecuado y sustancialmente superior al de las puntuaciones directas por atributos. Por tanto, los MDCs son útiles en contextos aplicados en la medida en que permiten

estimar con precisión qué atributos dominan o no dominan los alumnos. Se discuten pasos a seguir en el desarrollo de un software que facilite la evaluación diagnóstica en contextos escolares y universitarios.

¿Cómo cuantificar un Test de Juicio Situacional sin opciones correctas? Estudio comparativo de cuantificaciones basadas en TCT y TRI. Alicia Franco-Martínez¹, Diego Fernández Regueras¹ y Silvia Gregori Labarta¹

¹Universidad Autónoma de Madrid

En los *tests de juicio situacional* (SJT) todavía es necesario un consenso sobre la forma óptima de obtención de puntuaciones (*scoring*) cuando existe incertidumbre respecto a qué opción es correcta, o cómo evaluar cada una de las posibles opciones de respuesta. Estas dificultades describen el SJT que analizamos en este estudio, centrado en la medición de los efectos cognitivos del uso de distintas formas de genérico en español (i.e., "todos", "todos/as", "todxs" y "todes"), pues las opciones de respuesta representan estereotipos cualitativos. Presentamos dos estudios comparativos entre distintas formas de cuantificar con datos simulados (estudio 1) y reales (estudio 2). La calidad de la cuantificación se evalúa relacionando las puntuaciones asignadas a cada simulado con sus niveles de rasgo reales, sus puntuaciones en una variable criterio, y mediante el coeficiente alfa de las puntuaciones en el test. En el estudio 1 obtuvimos las puntuaciones simuladas en un test ($J = 10$, $K = 3$) y comparamos 14 cuantificaciones basadas en la TRI paramétrica y no paramétrica (modelo nominal de respuesta y *kernel smoothing*) y en la TCT. En éstas, cada ítem se ha cuantificado como ítem de opción múltiple (1, la opción correcta y 0 las dos incorrectas), o como ítem tipo Likert (1, 2 o 3, según la corrección de la opción). Para determinar la opción correcta exploramos dos reglas basadas en el consenso: considerar correcta la opción más elegida o puntuar cada opción con su porcentaje de elección. En algunos métodos se ponderó cada ítem de distintas formas: solo la opción correcta o todas ellas, usando diferentes índices de discriminación. En el Estudio 2 pusimos a prueba las anteriores cuantificaciones con los datos de una muestra real ($n = 800$), además de evaluarlas frente a otra cuantificación obtenida mediante dos criterios de expertos (frecuencia de elección y estereotipia de las opciones).

Métodos gráficos para la detección del funcionamiento diferencial del ítem usando medidas del tamaño del efecto basadas en Mantel-Haenszel y regresión logística. María Dolores Hidalgo¹, Brian French² y Juana Gómez Benito³

¹Universidad de Murcia, ²Washington State University, ³Universidad de Barcelona

El uso de medidas de tamaño del efecto en la detección del funcionamiento diferencial de los ítems es recomendable para comprender la magnitud de las diferencias que se detectan a través de una prueba de significación estadística. Se han sugerido diversas medidas del tamaño del efecto en función del método de detección del DIF, así como diferentes guías para su interpretación. El estadístico Mantel-Haenszel y el análisis de regresión logística son dos de los métodos más empleados para detectar el DIF para los que disponemos de medidas del tamaño del efecto (Mohanán et al., 2007). En el caso del estadístico Mantel-Haenszel podemos

usar una transformación del cociente de razones (α -MH); en el caso del análisis de regresión logística se dispone del estadístico delta-R2 y de medidas derivadas de la *odds*-ratio. El objetivo de este trabajo es describir el uso de métodos gráficos meta-analíticos para tomar decisiones tanto del funcionamiento diferencial de los ítems de un test (DIF) como del funcionamiento diferencial del test (DTF) cuando se utilizan medidas de tamaño de efecto del DIF basadas en el estadístico de Mantel-Haenszel y en el análisis de regresión logística. Se pretende establecer pautas para cuantificar y comparar la cantidad de DIF y DTF entre diferentes ítems de un mismo test y entre diferentes tests.

COMUNICACIONES ORALES

Aplicaciones de la Metodología I

Coordina: Roser Bono

Un modelo bayesiano para examinar curvas de aprendizaje a nivel individual. Fernando Blanco Bregón¹ y Joaquín Moris Fernández²

¹Universidad de Deusto, ²Universidad Internacional de La Rioja

En el campo del aprendizaje asociativo, la inferencia estadística se realiza usualmente sobre datos que agregan grupos de individuos o de sesiones experimentales, como es tradicional en psicología. Sin embargo, la interpretación de los datos agregados no siempre coincide con la conducta de los individuos que forman parte del mismo grupo. Es decir, algunas propiedades que se observan en los datos agregados no aparecen al examinar individuos concretos. En concreto, esta limitación es problemática cuando investigamos curvas de aprendizaje. Aunque las teorías actuales predicen curvas de aprendizaje graduales y negativamente aceleradas, este patrón sólo emerge al combinar curvas de varios participantes. Por el contrario, la conducta de los individuos muestra cambios bruscos y alta variabilidad entre participantes y situaciones. Por otro lado, el estudio de curvas individuales ha sido hasta ahora muy limitado, utilizando herramientas estadísticas insuficientes. Presentamos un modelo bayesiano de detección de cambio que permite analizar la evolución del aprendizaje de los individuos a lo largo del experimento, produciendo información complementaria al análisis tradicional. Gracias a este análisis, concluimos que el comportamiento de las curvas de aprendizaje de los individuos no se ajusta a las predicciones teóricas predominantes, ya que éstas incluyen saltos bruscos y cambios de dirección, que por otro lado pueden ser predichas a partir de parámetros manipulados en el experimento (como la tasa de reforzamiento). Estas observaciones requieren la incorporación de nuevos enfoques teóricos que reflejen más claramente la conducta individual y permitan entenderla.

Un análisis de redes del Cuestionario Mindfulness de Cinco Facetas (FFMQ). Oscar Lecuona de la Cruz¹ y Raquel Rodríguez Carvajal¹

¹Universidad Autónoma de Madrid

El Cuestionario Mindfulness de Cinco Facetas es una de las medidas autoinformadas de atención plena (*mindfulness*) más popu-

lares en la literatura. Sin embargo, estudios recientes aportan evidencia ambigua sobre su fiabilidad y validez. Además, un estudio de replicación reciente (Lecuona, García-Garzón, García-Rubio y Rodríguez-Carvajal, 2018) encuentra problemas para replicar la estructura latente, incluso tras fuertes modificaciones. Sin embargo, los problemas radicaban principalmente en los factores generales y de método, mientras que las facetas se replicaban adecuadamente. Por tanto, es posible que otros enfoques permitan aportar soluciones eficientes. En este estudio exploratorio se aplica el análisis de redes, una técnica novedosa que permite relacionar los ítems entre sí e identificar comunidades o *clusters* de los mismos. Una muestra de 800 participantes fue usada para estimar los modelos. Las redes estimaron correlaciones positivas entre todos los ítems, a la vez que los ítems quedaron colocados de manera parecida a las facetas originales. Sin embargo, un análisis de comunidades reveló un total de 6 comunidades de ítems, replicando cuatro de las cinco facetas originales del cuestionario menos "actuar con consciencia", que se dividió en dos posibles sub-facetos. Como añadido, no se encontraron diferencias apreciables de género o por experiencia meditativa en la estimación de las redes. Por tanto, los datos sugieren que el FFMQ puede contar con estructuras latentes alternativas, además de aportar información sobre el comportamiento del autoinforme en un marco de análisis de redes.

¿Pueden predecirse los rasgos de personalidad a partir del lenguaje? Un estudio meta-analítico sobre modelos computacionales. José David Moreno¹, José Ángel Martínez Huertas¹, Manuel Aenlle¹, Álvaro López¹, Guillermo Jorge Botana² y Ricardo Olmos Albacete¹

¹Universidad Autónoma de Madrid, ²Universidad Nacional de Educación a Distancia

Durante los últimos años ha surgido un fuerte interés en la relación que existe entre el lenguaje y los estilos de la personalidad. De hecho, hay estudios que trabajan bajo el supuesto de que el lenguaje puede predecir patrones de personalidad de los individuos. Los métodos de análisis del lenguaje para el estudio de la personalidad pueden dividirse en dos: los que se basan en criterios de jueces humanos expertos y los que se derivan de modelos lingüísticos computacionales (p. ej., LIWC, LSA o Topic Model). Sobre estos últimos, la variabilidad es muy grande (Carvalho y Louwerse, 2017; Kwantes et al., 2016). Con esta idea, se presenta un meta-análisis a partir de una muestra de 83 artículos con el objetivo de estudiar la capacidad predictiva de los Cinco Grandes (Big Five) desde los modelos computacionales. En particular, se analiza la estimación del tamaño del efecto que se obtiene al predecir los distintos rasgos de personalidad junto con posibles variables moderadoras tales como el tipo de método computacional empleado, el idioma analizado o las características del lenguaje estudiadas (sintáctico vs. semántico). A su vez, se intenta dar respuesta a qué rasgos de personalidad quedan mejor predichos por estos modelos y qué variables moderadores tienen más impacto en las estimaciones del tamaño del efecto. Respecto a los resultados, se espera encontrar estimaciones combinadas del tamaño del efecto medias, en torno a 0,40 para los cinco rasgos de personalidad analizados, tal y como se ha podido observar

en las primeras estimaciones realizadas en una submuestra extraída del total de estudios. Así, los resultados del presente meta-análisis pretenden comprobar qué rasgos de personalidad se predicen mejor a partir del lenguaje y cuáles son las variables moderadoras que presentan una mayor influencia sobre las estimaciones del tamaño del efecto.

Efecto de la rehabilitación cardíaca sobre la actividad del sistema nervioso autónomo en pacientes con enfermedad arterial coronaria: un meta-análisis. Agustín Manresa Rocamora¹, José Manuel Sarabia Marín¹, Manuel Moya Ramón¹, Javier Ibias² y Francisco José Vera García¹

¹Universidad Miguel Hernández (Elche), ²Universidad Nacional de Educación a Distancia

Los pacientes con enfermedad arterial coronaria presentan disfunción autónoma, la cual puede ser medida a partir de diferentes índices, como la frecuencia cardíaca de recuperación (FCR), la variabilidad de la frecuencia cardíaca (VFC), medida como la raíz cuadrada media de los intervalos R-R (RMSSD) o alta frecuencia (HF), o la sensibilidad barorefleja (SBR). El efecto de la rehabilitación cardíaca basada en el ejercicio (RC) sobre la disfunción autónoma es controvertido, por lo que fue analizado en este meta-análisis aplicando modelos de efectos aleatorios y mixtos. La búsqueda bibliográfica nos permitió incluir 29 estudios en la síntesis cualitativa y 23 en la síntesis cuantitativa, dando un total de 29 comparaciones independientes entre el grupo tratado y el grupo control. La medida del tamaño del efecto utilizada fue la diferencia entre cambios medios pre-post (DCM) para la FCR y la SBR. Para las variables relacionadas con la VFC se utilizó la forma estandarizada (DCME), la cual se interpretó como trivial (2,0). El efecto estimado alcanzó significación estadística para la FCR (DCM+ = 5,35; 95% IC = 4,08 - 6,61 lpm) y la SBR (DCM+ = 1,98; 95% IC = 0,60 - 3,37 ms·mmHg-1) a favor del grupo tratado. Además, los análisis mostraron diferencias estadísticamente significativas en RMSSD entre ambos grupos a favor del grupo tratado con un tamaño del efecto pequeño (DCME+ = 0,30; 95% IC = 0,12 - 0,49). Los análisis no mostraron diferencias estadística o clínicamente significativas en HF entre ambos grupos (SMD+ = 0,14; 95% IC = -0,12 - 0,40). Las relaciones de las variables extrínsecas, de sujeto, de tratamiento y metodológicas con el tamaño del efecto fueron analizadas, se llevó a cabo un análisis de sensibilidad y de sesgo de publicación. Finalmente, fueron discutidas las implicaciones clínicas de los resultados.

Aplicación del concepto de “validez estructurada” para planificar una investigación cuasi-experimental que combina el feedback y autorregulación. Paula Fernández García¹, David Martín Espinosa¹, María Belén San Pedro Veledo¹, Feliciano Ordóñez Fernández^{2,3} y Carmen González González-Mesa¹

¹ Universidad de Oviedo, ²Facultad Padre Ossó, ³Universidad Internacional de La Rioja

Durante tres cursos académicos se realizó llevó a cabo un proyecto de innovación docente en una asignatura de metodología de un curso de grado en la Universidad de Oviedo. La finalidad

fue incrementar el interés por la asignatura como medio para incrementar la satisfacción con lo aprendido, la nota académica, y la conciencia de utilidad de la asignatura para su desarrollo profesional. Los resultados fueron satisfactorios y se atribuyen al *feedback* administrado por el profesor. Se planifica una investigación cuasi-experimental de cohortes para poner a prueba la eficacia de la misma intervención, pero matizada y enriquecida, con el objetivo de incrementar la autorregulación de los alumnos de bachiller y conseguir elevar la nota académica no sólo en la asignatura que recibe la intervención, sino en todas las asignaturas del curso. De la mano del concepto de “validez estructurada” se realizó un protocolo para garantizar la integridad del tratamiento, para contener la pérdida de datos, para registrar higiénicamente las variables dependientes de interés, y para medir y registrar todas las fuentes de variabilidad que afectan a la conducta ajenas a la intervención. En estos momentos la investigación está en marcha. Diariamente se realiza un inventario de todas las dificultades encontradas en la aplicación y el modo en que se ha decidido atenderlas. De este modo se pretende contener las amenazas a la validez de constructo, validez interna, validez de conclusión estadística y conocer el alcance de la validez externa.

¿Cómo afecta, en los modelos de crecimiento, la relación entre el estatus y la pendiente de crecimiento en la fiabilidad de las estimaciones? Delia Arroyo Resino¹, María Castro Morera¹ y Enrique Navarro Asencio¹

¹Universidad Complutense de Madrid

Cada vez son más los países que utilizan las evaluaciones como indicador de la calidad de los sistemas educativos. Por ello, las evaluaciones longitudinales basadas en el seguimiento de una misma cohorte de estudiantes a lo largo de más de dos ocasiones de medida están adquiriendo cada vez más importancia, sin embargo, se desconoce cómo la relación entre el estatus (rendimiento en la primera medición) y la pendiente de crecimiento puede afectar en la fiabilidad de las estimaciones del logro en este tipo de diseños. Partiendo de dicha problemática, se plantea conocer el número de medidas necesarias para obtener estimaciones fiables. Para ello, a partir de los datos obtenidos en una evaluación longitudinal, se realiza una simulación Monte-Carlo basada en un modelo jerárquico lineal, con tres niveles de anidamiento (tiempo, estudiante y aula) y nueve ocasiones de medida, y se desarrollan dos modelos. En el primero de ellos se considera que el rendimiento de partida de los estudiantes no se encuentra muy relacionado con el adquirido a lo largo del tiempo (se fija la correlación entre estatus y pendiente a 0,2) y en el segundo modelo lo contrario, es decir que el rendimiento que obtienen los estudiantes en la primera medición se encuentra muy relacionado con el adquirido en el resto de las aplicaciones (correlación entre estatus y pendiente igual a 0,9). Los resultados muestran que, mientras que en el primer modelo se requieren unas 4-5 mediciones para obtener estimaciones fiables del rendimiento de los estudiantes, en el caso del segundo modelo se pueden obtener estimaciones muy precisas con solo 3 aplicaciones. Así, se concluye que la relación entre el punto de partida y la pendiente de crecimiento es un factor importante a la hora de decidir el número de mediciones a realizar en estos diseños longitudinales.

COMUNICACIONES ORALES

Aplicaciones de la Metodología

Coordina: Encarnación Sarriá

Efectos de un procedimiento de evaluación formativa sobre el rendimiento en la asignatura "Estadística aplicada a la psicología II" de la UCM. Sergio Escorial Martín¹

¹Universidad Complutense de Madrid

La adaptación al Espacio Europeo de Educación Superior (EEES) supuso un cambio de reglas en el proceso de enseñanza-aprendizaje, trasladando la responsabilidad de 'enseñar' del profesor al 'aprender' del alumno. Este cambio afectó a las programaciones docentes. Uno de los aspectos que se modificaron con esta adaptación es que se buscó incrementar el número de actividades docentes que potencien un papel más activo del alumno en su aprendizaje. El objetivo es promover actividades formativas, que motiven al alumno a hacer un esfuerzo más distribuido a lo largo del período de impartición de la asignatura. En este trabajo se presentan los efectos para el rendimiento en la asignatura de "Estadística aplicada a la psicología II" de una serie de actividades de aprendizaje activo que se pueden englobar dentro de lo que los expertos en nuevas metodologías denominan "Evaluación Formativa" (p. ej., Prieto, 2017). Durante el curso 2017-2018 se utilizaron para realizar esta investigación los resultados de dos grupos cuya docencia era impartida por el mismo profesor. En un grupo se desarrolló un procedimiento de evaluación formativa (aprendizaje activo) y en el otro grupo no se realizó esta actividad. El resto de actividades docentes y prácticas fueron idénticas en los dos grupos, los materiales, procedimientos de evaluación y el examen también fueron los mismos. El grupo que realizó la evaluación formativa alcanzó una nota media más elevada en el examen ($M = 6,79$, $DT = 2,16$) que el grupo que no realizó esta actividad ($M = 4,76$, $DT = 1,98$). Esta diferencia resultó estáticamente significativa, $t(113) = 5,25$, $p < 0,001$, $d = 0,97$. Por otra parte, no se encontraron diferencias significativas en el grado de satisfacción que los alumnos de ambos grupos informaban para valorar la actividad docente del profesor.

¿El foco de referencia afecta a los índices de agregación y a los resultados de los análisis multinivel? Inés Tomás Marco¹, Charlotte Merrett², Ana María Hernández Baeza¹, Paul Appleton² y Joan Duda²

¹Universitat de València, ²University of Birmingham

Las técnicas de análisis multinivel han adquirido gran relevancia en las últimas décadas ya que muchos fenómenos que se estudian en los diferentes ámbitos de la psicología (p. ej., organizacional, deportiva, educativa) son de naturaleza multinivel y los datos recogidos implican estructuras anidadas. Antes de llevar a cabo un análisis multinivel es necesario realizar la agregación de los datos, y previamente, estimar diferentes índices de agregación para justificar dicha agregación. Entre los índices de agregación más frecuentemente utilizados podemos mencionar: el índice de desviación promedio (Average Deviation Index, ADI; Burke, Finkelstein y Dusig, 1999), el rWG(J) (James, Demaree y Wolf, 1984) y los coeficientes de correlación intraclase ICC1 e ICC2 (intraclass correlation coefficients; Bliese, 2000). Cuando medimos fenóme-

nos colectivos en el contexto del análisis multinivel, encontramos dos aproximaciones en función de cuál sea el foco de referencia: el individuo o el equipo. En la aproximación donde el referente es el individuo, se le pregunta al sujeto sobre su percepción individual (p. ej., "me siento apoyado por el entrenador"). Mientras que en la aproximación donde el referente es el equipo, se le pregunta al sujeto sobre la percepción del equipo (p. ej., "nos sentimos apoyados por el entrenador"). En ambos casos, las puntuaciones individuales (con el foco en el individuo o con el foco en el equipo) son agregadas para obtener la puntuación de nivel equipo. El propósito de este estudio es analizar si el uso de las diferentes aproximaciones (referenciada al individuo vs referenciada al equipo) puede afectar a la estimación de los diferentes índices de agregación (ADI, rWG(J), ICC1 e ICC2), así como a los resultados de la estimación de un modelo teórico multinivel. Se extraerán conclusiones sobre las implicaciones derivadas de usar una u otra aproximación.

¿Es el GFP algo distinto de la DS? Una aproximación basada en el método de Ferrando et al. (2009). Sergio Escorial Martín¹, David Navarro González², Fabia Morales-Vives² y Andreu Vigil Colet²

¹Universidad Complutense de Madrid, ²Universitat Rovira i Virgili

El factor general de personalidad estaría apresando varianza común a los factores de personalidad del modelo Big-Five. Propuesto inicialmente por Digman (1997), su existencia y su estatus empírico fue muy controvertido hasta el meta-análisis de Van der Linden et al. (2010). A partir de ese momento, su existencia fue aceptada, pero el debate se trasladó a la naturaleza de esa varianza común. Para algunos autores, el GFP representa una varianza sustantiva, mientras que para otros representa sesgos de respuesta, fundamentalmente DS. En este estudio se empleará el método propuesto por el equipo de la URV coordinado por Pere Joan Ferrando (Ferrando, Lorenzo-Seva y Chico, 2009). Ese método permite controlar la DS empleando un conjunto de marcadores (ítems de DS) y, mediante procedimientos basados en escalas parcialmente balanceadas, limpiar las puntuaciones de AQ. Se empleará para el objetivo de este estudio una base de datos de 3838 sujetos (49,69% mujeres, con un rango de edad entre los 13 y los 95 años, $M = 32,28$ y $DT = 15,32$) a los que se les administró el OPERAS, un test de personalidad Big-Five. Los resultados alcanzados indican que el GFP representa fundamentalmente varianza sustantiva debido a que: (1) tras eliminar los sesgos (DS y AQ) se sigue identificando el GFP sin que su estructura factorial se debilite o quede comprometida (el GFP da cuenta en ambos casos de algo más del 34% de la varianza); (2) la correlación entre los factores de GFP (con y sin sesgos) es muy intensa ($r = 0,82$); y (3) la correlación entre los factores de DS y GFP (con sesgos) es sustancialmente menor que la observada entre los dos factores de GFP (con y sin sesgos) ($r = 0,42$).

Análisis de tres interpretaciones normativas diferentes sobre un mismo conjunto de puntuaciones en distintas pruebas neuropsicológicas. Raquel Rodríguez Fernández¹, Carmen Díaz Mardomingo¹, Sara García Herranz¹ y César Venero Núñez¹

¹Universidad Nacional de Educación a Distancia

La interpretación del resultado que una persona obtiene en una prueba, desde el punto de vista normativo, se realiza comparando

dicha puntuación con la de su grupo de referencia en esa misma prueba, es decir, las normas o baremos desarrollados a partir de esta orientación se realizan en función del grupo normativo del sujeto. Dentro de dicha orientación hay diversas maneras de transformar las puntuaciones directas de los participantes en otras que expresen su posición relativa respecto a su grupo de referencia. En el presente estudio comparamos las puntuaciones transformadas de los resultados obtenidos en 4 pruebas neuropsicológicas ampliamente utilizadas en la práctica clínica: Trail Making Test, fluidez semántica (animales), fluidez fonética (palabras que empiezan por P) y figura compleja de Rey a la copia, de una muestra de 367 personas mayores (rango de edad entre 60 y 90 años). Concretamente, obtuvimos 3 transformaciones diferentes de las puntuaciones directas: (a) a partir del baremo desarrollado por el grupo de Peña-Casanova et al (2009), o proyecto NEURONORMA, en el cual se aportan tablas para convertir las puntuaciones brutas en escalares y tablas con los ajustes pertinentes por edad y escolaridad a partir de una regresión lineal; (b) calculando, siguiendo la misma metodología empleada en el proyecto normativo NEURONORMA, nuestras propias tablas de transformación utilizando dicha metodología; y (c) hallando las puntuaciones *Z*, tomando para ello las medias y desviaciones típicas obtenidas en cada una de las submuestras que representan los rangos de edad y niveles de estudio de la población de mayores española, controlando así el efecto de ambas variables al igual que hacen las otras dos transformaciones. Los resultados obtenidos se interpretan en función de sus posibles equivalencias, así como de su utilidad en la toma de decisiones dentro del contexto clínico.

Discapacidad auditiva: conciencia de estigma, calidad de vida e identidad social. Adaptación de tres cuestionarios a personas con discapacidad auditiva. Jesús Privado Zamorano¹ y Luis Francisco Carrasco¹

¹Universidad Complutense de Madrid

Introducción: el objetivo de esta investigación fue adaptar y validar cuestionarios de bienestar y calidad de vida en personas con discapacidad auditiva. Los cuestionarios adaptados fueron: "Conciencia de estigma" de Pinel, "Calidad de vida" de Ruiz y Baca (1993) e "Identidad social" de Cameron (2004). Método: se evaluaron a 216 discapacitados auditivos de 38,22 años (DT: 12,97), el 53,2% eran mujeres, con diferente nivel educativo, estado civil y ocupación. Los instrumentos fueron adaptados por expertos en discapacidad auditiva y se estudió su evidencia de validez de contenido. Después fueron aplicados por evaluadores entrenados a discapacitados auditivos. Se analizó la distribución de las medidas, las evidencias de validez interna y de criterio y la fiabilidad de las medidas. Resultados: la fiabilidad de las tres pruebas fue adecuada: superior a 0,70. Mediante análisis factorial confirmatorio se obtuvo: un factor en la prueba "Conciencia de estigma", cuatro factores en la de "Calidad de vida" (satisfacción general, tiempo libre, bienestar físico y psicológico, y apoyo social) y tres factores en la de "Identidad social" (centralidad, lazos intergrupo y afecto intragrupal). En todos los casos las pruebas de bondad de ajuste fueron adecuadas. Además, se encontraron pruebas de evidencia de validez convergente, discriminante y de criterio. Conclusiones: en el factor de la prueba "Conciencia de estigma" hemos encontrado resultados coincidentes con estudios previos en personas no discapacitadas (Pinel, 1999). En la escala "Calidad de vida" los cuatro factores obtenidos son similares a los obtenidos por los autores de la escala (Ruiz y Baca, 1993). La estructura interna de la prueba de "Identidad social" reflejó la presencia de tres factores correspondientes a las tres subescalas planteadas por Cameron (2004). Por lo tanto, se ha logrado adaptar a discapacitados auditivos tres medidas pensadas para no discapacitados, ofreciendo adecuados índices de fiabilidad y validez.

dad de vida" los cuatro factores obtenidos son similares a los obtenidos por los autores de la escala (Ruiz y Baca, 1993). La estructura interna de la prueba de "Identidad social" reflejó la presencia de tres factores correspondientes a las tres subescalas planteadas por Cameron (2004). Por lo tanto, se ha logrado adaptar a discapacitados auditivos tres medidas pensadas para no discapacitados, ofreciendo adecuados índices de fiabilidad y validez.

Evaluación psicométrica de la Escala de evaluación de los Activos del Barrio en una muestra de adolescentes chilenos.

José Luis Gálvez Nieto¹ e Ítalo Trizano Hermosilla¹

¹ Universidad de La Frontera

El concepto de activos del barrio se define como la percepción del adolescente respecto de los activos vecinales, tales como la valoración positiva y el empoderamiento de los jóvenes por parte adultos, la seguridad del barrio, la existencia de actividades recreativas y de ocio dirigida a la población adolescente. Para medir este constructo se desarrolló la Escala de Evaluación de los Activos del Barrio la cual ha demostrado adecuados niveles de validez y fiabilidad para su uso en España, sin embargo aún no se presentan evidencias para su uso en Chile. Esta investigación tuvo dos objetivos: adaptar y validar la EEAB y analizar la equivalencia factorial entre hombres y mujeres en una muestra de adolescentes chilenos. La muestra quedó conformada por 841 estudiantes (53,5% mujeres) con una edad media de 16,05 años (DE = 1,41). Los resultados obtenidos mediante análisis factorial exploratorio y confirmatorio, evidenciaron que la EEAB mostró una estructura de cinco factores relacionados. El análisis de invarianza factorial reflejó que el grado de equivalencia entre hombres y mujeres se presenta hasta el nivel de invarianza escalar. Estos resultados implican que la EEAB puede ser utilizada en contextos educativos chilenos con leves modificaciones de la versión original.

Factores protectores y de riesgo en la retención y rendimiento académico de estudiantes de pregrado. Miguel Ripoll¹, Claudio Rivera¹, Claudia Rubio¹ y Valentina Perez¹

¹Universidad de Los Lagos

A partir de la cohorte de pregrado 2018 de la Universidad de Los Lagos, Región de Los Lagos, Chile, se ha iniciado seguimiento de los alumnos por la duración de sus carreras mediante la captura de fuentes secundarias de información de al menos un nivel ordinal de medición, compuesta por una base demográfica de los alumnos, desempeño en pruebas de selección universitaria, que incluye su desempeño en la enseñanza secundaria mediante calificaciones y ranking en su establecimiento de origen, pruebas de diagnóstico en matemática, lenguaje y ciencia rendidas al matricularse en la universidad. Encuesta aplicada al matricularse en la universidad sobre conocimiento y antecedentes que lo llevaron a la selección de la carrera en que se matricula y, por ahora, desempeño académico en su primer año de estudio, compuesto por calificaciones y número de créditos, promedio general por cuatrimestre y anual, y porcentaje de créditos aprobados por cuatrimestre y año de las asignaturas cursadas. Se construyó una matriz sujetos por variable y se depuró, conservando aquellos sujetos que tuvieran más de un 75% de celdas con información válida. En una primera instancia se realizó un estudio descriptivo de las va-

riables, con la finalidad de descartar aquellas variables que presentaran dispersiones cercanas a 0. A partir de la matriz resultante, se realizó un análisis de conglomerados jerárquico basado en distancias euclídeas al cuadrado. La presente investigación se continuará hasta que el total de los alumnos considerados en la muestra inicial pierda su condición de alumno regular, por deserción, eliminación o titulación, que será una última variable a considerar en la matriz de análisis, con lo que se buscará construir un modelo explicativo mediante ecuaciones estructurales.

COMUNICACIONES ORALES

Modelos psicométricos

Coordina: Carme Viladrich

RETAM: un procedimiento en el contexto de la rotación factorial para refinar de forma objetiva matrices diana. Urbano Lorenzo Seva¹, Pere J. Ferrando¹ y Andreu Vigil Colet¹

¹Universitat Rovira i Virgili

Una dificultad común en el análisis factorial de ítems diseñados para medir constructos psicológicos es que las estructuras factoriales obtenidas mediante el análisis factorial exploratorio tienden a ser rechazadas si se evalúan estadísticamente con un modelo factorial confirmatorio. Una alternativa al análisis factorial confirmatorio es el análisis factorial no restringido basado en la rotación de Procrustes, que minimiza la distancia del patrón factorial rotado respecto a una matriz diana propuesta por el investigador. Esta presentación se centra en la situación en la que el investigador propone una matriz diana parcialmente especificada, si bien está dispuesto a permitir que su hipótesis inicial sea refinada. Nuestra propuesta metodológica para esta situación es RETAM, un procedimiento para refinar una matriz diana utilizando un criterio objetivo. Hasta la fecha, se ha recomendado que este tipo de refinamiento sea guiado por el juicio humano. Por el contrario, nuestro enfoque es objetivo porque el valor de umbral se calcula automáticamente y no precisa que el investigador deba repetir el análisis de los mismos datos en un número arbitrario de ocasiones. El nuevo procedimiento ha demostrado su utilidad en un extenso estudio de simulación. Su aplicabilidad en la práctica se ilustra con un ejemplo empírico en el dominio de la personalidad.

Análisis bayesiano del modelo beta-factorial para datos continuos acotados. Beatriz Hidalgo Sanz¹ y Javier Revuelta Menéndez¹

¹Universidad Autónoma de Madrid

El análisis factorial con datos continuos acotados ha recibido poca atención. Estos datos suelen presentar asimetría positiva o negativa como consecuencia de que existen límites que no pueden exceder. Los métodos robustos de estimación para datos que no cumplen el supuesto de normalidad no producen resultados satisfactorios en este caso. Este trabajo propone un modelo beta-factorial adecuado para el análisis de datos continuos acotados con distintos tipos de asimetría. Se han empleado métodos de inferencia bayesiana debido a que dada la novedad del modelo beta-factorial, no existen métodos de estimación y contraste en el marco clásico. La inferencia bayesiana permite estimar y evaluar el

modelo, y por tanto tomar una decisión sobre el número adecuado de factores a retener. Se ha realizado un estudio de simulación para comparar el funcionamiento del modelo propuesto frente al modelo factorial con variables normales en datos con asimetría positiva, negativa y mixta. Se utilizan distintos métodos de evaluación y de selección de modelos (DIC, WAIC y LOO-CV) propios de la inferencia bayesiana y que hasta ahora han sido poco explorados en la literatura psicométrica. Además, aplicamos el modelo factorial beta a una base de datos reales que motiva el estudio de simulación. Los resultados indican que el modelo factorial beta es una alternativa adecuada para la estimación de datos con asimetría, en particular es el único método que proporciona resultados satisfactorios ante la presencia simultánea de observaciones negativas y positivas.

Análisis de bondad de ajuste del modelo logístico de tres parámetros mediante el test de la razón de verosimilitudes condicional con grados de libertad estimados. Javier Revuelta Menéndez¹ y David Coba Castellano¹

¹Universidad Autónoma de Madrid

El modelo logístico de tres parámetros de la teoría de respuesta al ítem incluye un parámetro de adivinación que representa la probabilidad de acertar el ítem cuando el valor de la variable latente evaluada por el ítem es mínimo. La bondad de ajuste de dicho modelo se contrasta mediante un estadístico de razón de verosimilitudes frente a un modelo de dos parámetros, que asume que la probabilidad de adivinación es cero. A dicho estadístico se le llama *naive* en la literatura psicométrica porque ignora el hecho de que la distribución de la razón de verosimilitudes solamente es chi-cuadrado cuando todos los parámetros del modelo verdadero se encuentran dentro de su espacio paramétrico, cosa que no sucede en el modelo logístico de dos parámetros. Para abordar este problema se ha propuesto el *chi-bar test*, que tiene el inconveniente de que solo puede aplicarse a un ítem a la vez y requiere realizar comparaciones múltiples. Una solución más práctica a este problema es contrastar la bondad de ajuste mediante el test de la razón de verosimilitudes condicional (CLRT). En este test, el número de grados de libertad del estadístico chi-cuadrado es el número de parámetros de adivinación estimados en el 3PL que se encuentran dentro del espacio paramétrico. En esta comunicación se comparan los resultados del CLRT con los del chi-cuadrado *naive* y con los del *chi-bar test* mediante simulación de Monte Carlo. Se han generado datos del 2PL y el 3PL en distintas condiciones para estimar la probabilidad de cometer errores Tipo I y la potencia estadística de los tres métodos de contraste de hipótesis. Los resultados muestran que el CLRT produce mejores resultados que los otros estadísticos, además de ser muy sencillo de aplicar en la práctica.

Modelos explicativos de respuesta al ítem para la medición del aprendizaje durante el test. José Héctor Lozano Bleda¹ y Javier Revuelta Menéndez¹

¹Universidad Autónoma de Madrid

Se presentan tres modelos explicativos de respuesta al ítem para dar cuenta del aprendizaje que tiene lugar durante la administración de un test psicométrico debido a la práctica repetida de las

operaciones implicadas en la resolución de los ítems. Los modelos incluyen un componente de dificultad derivado de las operaciones cognitivas implicadas en el ítem, así como un componente de aprendizaje derivado del uso de dichas operaciones en ítems anteriores. Los modelos difieren en el tipo de respuesta en el que el modelo establece que tiene lugar el aprendizaje: respuestas correctas e incorrectas indistintamente, respuestas correctas exclusivamente (aprendizaje contingente) y respuestas correctas e incorrectas en diferente medida (aprendizaje contingente diferencial). Se presenta un estudio de simulación que permite examinar el rendimiento del método de estimación y los estadísticos de bondad de ajuste bayesianos en la recuperación de parámetros y la identificación del modelo utilizado para la generación de los datos, respectivamente. La estimación se llevó a cabo mediante cadenas de Markov Monte Carlo utilizando el algoritmo NUTS. La evaluación del ajuste se realizó por medio de comprobaciones predictivas posteriores con base en el índice de discrepancia SGDDM, mientras que la comparación de modelos se basó en tres índices de devianza bayesianos: DIC, WAIC y LOO. Los resultados muestran el impacto del aprendizaje en la recuperación de parámetros, así como el buen rendimiento de las medidas de ajuste y comparación de modelos. Adicionalmente, se presenta un estudio empírico que permite ilustrar la aplicabilidad de los modelos a datos reales. Concretamente, se analizaron datos procedentes del test de sustracción de fracciones de Tatsuoka. Los resultados proporcionan evidencias de aprendizaje contingente diferencial en los datos. Agradecimientos: la estimación fue realizada con los recursos del Centro de Computación Científica de la Universidad Autónoma de Madrid (CCC-UAM).

Evaluación de la dimensionalidad en ítems dicotómicos: Comparación entre el modelo de escalamiento de Mokken y el análisis gráfico exploratorio. María Dolores Hidalgo Montesinos¹, Gomaa Said Mohamed Abdelhamid² y Juana Gómez Benito²

¹Universidad de Murcia y (2) Universitat de Barcelona

Antecedentes: la determinación del número de factores es un paso crítico en el desarrollo de cuestionarios. El modelo de Mokken utiliza un procedimiento de selección de ítems automatizado (AISP) para identificar ítems que miden el mismo rasgo (Mokken, 1971). Straat et al. (2013) propusieron un procedimiento basado en una función objetivo que utiliza un algoritmo genético (GA) para encontrar la partición de ítems óptima en MSA. De manera similar, los análisis de redes son modelos cada vez más atractivos para investigar la complejidad dimensional de los fenómenos psicológicos. El análisis de gráficos exploratorios (EGA), en el marco del análisis de redes, se ha propuesto recientemente para identificar el número de factores latentes (Golino y Epskamp, 2017). Están disponibles dos métodos de EGA: glasso gráfico (EGA glasso) con el parámetro de regularización especificado mediante EBIC y el enfoque de gráfico triangulado con filtro máximo (EGAtmfg). Objetivo: comparar, usando datos simulados, la eficacia de los cuatro métodos (AISP, GA, EGAglasso, EGAtmfg) para determinar el número de factores subyacentes en ítems de respuesta dicotómica. Método: las variables manipuladas fueron: número de factores (2); correlación entre factores ($\rho = 0$), tamaño de la muestra (250), número de ítems por factor (5). Los datos se simularon utilizando un modelo de TRI logístico 2-p compensa-

torio multidimensional. Se simularon 1.000 réplicas por condición manipulada. Para evaluar la eficacia de los diferentes métodos se calculó: porcentaje de factores correctamente identificados, sesgo del número de factores, raíz de la desviación cuadrática media y criterio de clasificación errónea por ítem. Resultados: los cuatro métodos mostraron una alta precisión en la estimación del número de factores y los sesgos fueron bajos. El método más preciso fue EGAtmfg. Conclusión: el estudio destaca la importancia de utilizar MSA y EGA para determinar las dimensiones latentes en ítems dicotómicos.

A tutorial on the recovery of the optimal number of latent factors with traditional and modern techniques. María Dolores Nieto¹, Luis Eduardo Garrido², Hudson Golino³, Dingjing Shi³ y Francisco José Abad¹

¹Universidad Autónoma de Madrid, ²Pontificia Universidad Católica Madre y Maestra, ³University of Virginia

Identifying the number of underlying dimensions is a critical phase in the development of theory. For this purpose, Horn's Parallel Analysis is currently considered the gold standard method because it has systematically shown to be superior over other traditional techniques. More recently, the Exploratory Graph Analysis (EGA) has emerged within the framework of network psychometrics as a promising approach for the same goal. A distinctive feature of the EGA is that it produces a visual guide – network plot – that not only indicates the number of dimensions to retain, but also which items cluster together and their level of association. The combination of both techniques in practice provides valuable complementary information in dimensionality assessment. With the aim of providing further insight and specific guidance about the use of both techniques, a tutorial will be presented to show how different population factor model parameters (factor loadings, variables per factor, etc....) and data characteristics (sample size, number of response options, etc.) impact the recovery of the optimal number of factors in each case. In addition, other highly popular dimensionality methods will be included to make further comparisons. For this endeavor, a Shiny application based on free R code is presented. This application allows users to manipulate the factor model parameters and the data characteristics to see how well the population dimensionality can be recovered by the different methods. Additionally, the Shiny application allows the users to conduct simulations for user-specified conditions so that they can ascertain the behavior of the methods across a specified number of replicates. In closing, the main purpose of this work is that users with different profiles (students, teachers, and researchers) can benefit in different ways from the PA and the EGA techniques, and provide users with a tutorial and Shiny application to this end.

UNIVAL: un paquete de R para evaluar la dimensionalidad utilizando criterios externos. David Navarro González¹, Pere J. Ferrando¹, Urbano Lorenzo Seva¹ y Andreu Vigil-Colet¹

¹Universitat Rovira i Virgili

La evaluación de la dimensionalidad de las medidas psicométricas mediante análisis factorial es un aspecto básico en la construcción y desarrollo de tests. Sin embargo, las prácticas utilizadas habitualmente para decidir entre soluciones múltiples o esencialmente

unidimensionales son bastante mejorables, y se basan fundamentalmente en información interna, como la proporcionada por los índices de bondad de ajuste. Ferrando y Lorenzo-Seva han propuesto un procedimiento para determinar la unidimensionalidad basado en cómo los factores se relacionan con un criterio externo. El método considera dos facetas diferenciadas de validez: diferencial e incremental. La diferencial asume que los factores primarios están relacionados con la variable externa de la misma manera que se relacionan con el factor general. La validez incremental hace referencia a la hipótesis de que la predicción realizada usando las puntuaciones factoriales del factor general no puede ser mejorada usando las puntuaciones de los distintos factores primarios. Así pues, se contrastan dos hipótesis, para determinar si la pérdida de información al utilizar el modelo unidimensional es demasiado sustancial en términos de validez. Para facilitar la utilización de dicho procedimiento, se creó un paquete del software R llamado *unival*, disponible a través del CRAN (<https://CRAN.R-project.org/package=unival> (<https://cran.r-project.org/package=unival>)), que pone a disposición del usuario el método anteriormente mencionado. En esta comunicación realizaremos una demostración de software, donde introduciremos el método, las características del paquete *unival* e ilustraremos su utilidad a través de un ejemplo práctico usando datos reales.

COMUNICACIONES ORALES

Análisis de datos

Coordina: Antonio Pardo

Avances en modelos longitudinales para variables latentes en tiempo continuo: aplicación a diseños de cohortes acelerados. Eduardo Estrada^{1,2} y Emilio Ferrer²

¹Universidad Camilo José Cela, ²University of California

En las últimas décadas, numerosos estudios longitudinales han empleado modelos de cambio latente para estudiar el desarrollo psicológico en distintas fases del ciclo vital. Estos modelos dinámicos ofrecen gran versatilidad para describir y explicar el cambio en una o más variables, e identificar influencias recíprocas entre ellas. Sin embargo, aplicarlos implica asumir que el intervalo entre medidas es constante para todos los participantes y entre todas las medidas de cada participante. Recientemente se han desarrollado modelos en tiempo continuo (Chow, Ho, Hamaker, y Dolan, 2010; Voelkle, Oud, Davidov, y Schmidt, 2012) que combinan las ventajas del modelado de variables latentes y de ecuaciones diferenciales, y permiten estimar sus parámetros independientemente del intervalo temporal entre observaciones. Nuestro objetivo fue comparar la utilidad de un modelo dinámico tradicional en tiempo discreto y otro análogo en tiempo continuo para analizar datos obtenidos mediante un diseño de cohortes acelerado (*accelerated longitudinal design*, ALD). En estos diseños, participantes de distintas cohortes son medidos repetidamente, pero las medidas de cada uno cubren solo una parte del rango temporal del estudio. Se asume que la trayectoria común se puede recuperar agregando la información de las diferentes cohortes. Debido a la propia naturaleza de un ALD, cada individuo es evaluado en momentos distintos. Por tanto, un modelo en tiempo continuo parece idóneo para estos datos. Los resultados de nues-

tro estudio de simulación apoyan la pertinencia de los modelos en tiempo continuo bajo diversas condiciones de muestreo. Todos los parámetros de la trayectoria se estimaron sin sesgo en las condiciones con cohortes provenientes de la misma población. Sin embargo, diferencias entre cohortes dieron lugar a estimaciones sesgadas en parámetros relevantes. Discutimos nuestros hallazgos en el contexto del desarrollo psicológico, y proponemos varias recomendaciones sobre como diseñar y analizar este tipo de estudios.

Modelos mixtos, una alternativa a los modelos ANOVA de medidas repetidas en diseños con dos efectos aleatorios cruzados: Un estudio de simulación. José Ángel Martínez Huertas¹ y Ricardo Olmos Albacete¹

¹Universidad Autónoma de Madrid

Diferentes estudios de simulación con múltiples medidas del mismo sujeto han mostrado la superioridad de los modelos multinivel frente a otros modelos estadísticos clásicos como los ANOVAs con medidas repetidas. Sin embargo, estos estudios se han centrado en simulaciones de diseños experimentales excesivamente simples como diseños pre-post, mientras que los estudios reales suelen responder a diseños más complejos (como los diseños *split-plot*). Una gran ventaja de los modelos mixtos se da en los diseños experimentales que tienen dos (o más) efectos aleatorios, diseños muy comunes en el ámbito educativo (donde los dos efectos aleatorios pueden ser los estudiantes y los materiales didácticos), en psicolingüística (participantes y palabras) o en psicología clínica (pacientes y síntomas). El problema de los modelos ANOVA de medidas repetidas es que son demasiado rígidos e ignoran la información (variabilidad) que se da en los dos efectos aleatorios simultáneamente, mientras que los modelos multinivel son capaces de modelar ambos efectos aleatorios conjuntamente. Por ello, el objetivo de este trabajo es estudiar la significación estadística que ofrecen los modelos multinivel y los ANOVAs de medidas repetidas mediante un estudio de simulación en el que se ha manipulado el tamaño del efecto, el número de participantes, el número de ítems y la presencia de distintos efectos aleatorios. Los resultados analizan la potencia estadística y la tasa de errores Tipo I bajo las distintas condiciones simuladas. Los resultados sobre la significación estadística de ambos modelos mostraron un gran efecto de las condiciones de simulación y también notables diferencias entre los efectos intra e inter sujetos. Específicamente, las condiciones simuladas afectaron a la potencia estadística, mientras que el modelo estadístico aplicado fue determinante en la tasa de los errores Tipo I. Las implicaciones que tiene esta nueva orientación de análisis estadístico para los investigadores aplicados son claramente relevantes.

Baremos ajustados por edad y escolaridad en la evaluación del funcionamiento cognitivo en adultos mayores mexicanos. Ana L. Gonzalez-Celis¹, Edgardo Ruiz-Carrillo¹, Karina Reyes Jarquin², Ana Gabriela Magallanes-Rodríguez³, María Refugio Acuña Gurrola² y Margarita Chavez Becerra¹

¹Universidad Nacional Autónoma de México, ²Universidad Autónoma del Estado de Hidalgo, ³Universidad Autónoma de Baja California

Para la evaluación del funcionamiento cognitivo en los adultos mayores es importante contar con instrumentos de medida valida-

dos tomando en cuenta las características de la población a la que va dirigida, se deben tomar en cuenta los criterios de edad y escolaridad, además de los aspectos culturales. De ahí que el propósito del estudio es resaltar la importancia de los baremos adaptados por grupos de edad y nivel educativo alcanzado, al comparar los resultados con dos instrumentos de medición del funcionamiento cognitivo, el Mini-Mental de Folstein (1985), en su versión española validada por Lobo y colaboradores (2001) y el SCIP-S (Screen for Cognitive Impairment in Psychiatry-Spanish) también validada al español (Pino et al., 2014). Mientras el MMSE ha sido diseñado para población geriátrica, el SCIP-S solo presenta tablas de baremos para población general de 18 a 55 años; por lo que al emplearla en población adulta mayor de 60 años y mayores, se presenta un problema al momento de interpretarla. De ahí que la necesidad de realizar los ajustes para adaptarla a población adulta mayor. En una muestra de adultos mayores mexicanos de 60 años y más, a quien se le aplicaron los dos instrumentos de funcionamiento cognitivo MMSE y SCIP-S, además de una ficha de datos sociodemográficos que contemplaba la edad y el nivel educativo alcanzado; se presentan las diferencias obtenidas entre ambos instrumentos, con el fin de evidenciar la importancia de contar con los baremos específicos para la edad y escolaridad. La distribución por rangos cualitativos considerando los puntos de corte establecidos en los manuales de ambos instrumentos discrepa dependiendo de las puntuaciones obtenidas en el MMSE. Es evidente que se deben utilizar instrumentos de evaluación traducidos y validados tomando en consideración la edad y escolaridad para su interpretación y comparación internacional.

Estimación de un índice de precariedad mediante PLS-PM Ordinal en estudiantes chilenos de 4º Básico. Horacio Miranda¹, Sonia Salvo¹ y Edgardo Miranda¹

¹Universidad de La Frontera

La condición de precariedad en la etapa estudiantil es una característica que expresa el grado de escasez de bienes materiales, recursos económicos y educativos que afectan a los estudiantes para su aprendizaje. Cada año el Ministerio de Educación de Chile recolecta datos de los municipios y del sistema sanitario con la finalidad de otorgar un certificado de alumno prioritario que categoriza la condición de precariedad estudiantil y permite obtener un subsidio del Estado. El presente trabajo surge de un avance de un proyecto sobre resiliencia socio-educativa en Chile. El objetivo de esta fase fue desarrollar un índice compuesto basado en información recolectada en los cuestionarios de contexto de la prueba estandarizada del sistema de medición de la calidad de la educación (SIMCE), empleando como variable de verificación la categorización de alumno prioritario del MINEDUC. Las variables fueron seleccionadas desde los cuestionarios de contexto mediante análisis de segmentación jerárquica e incluida en un modelo de trayectorias formativo-reflectivo estimado mediante mínimos cuadrados parciales con variables indicadoras ordinales. Las estimaciones del modelo y la tabla de confusión para evaluar la concordancia de la predicción del modelo con la categorización de alumno prioritario se realizaron mediante los paquetes *plsrm* y *DMSTools* de R respectivamente. El análisis de datos empleó

una base censal de 67.216 alumnos prioritarios y 9.408 no prioritarios. El modelo presentó una bondad de ajuste (GoF) de 0,75 y una predictibilidad con una sensibilidad de 0,79 y una especificidad de 0,60 con una exactitud de 0,61 y un valor kappa de 0,18. Estos resultados evidencian la posibilidad de continuar con el estudio de la modelación mediante PLS-PM a partir de los cuestionarios de contexto SIMCE, como un procedimiento complementario a las herramientas que emplea el MINEDUC basadas en información proveniente de los municipios y el sistema de salud. Agradecimiento: este estudio reporta resultados parciales del proyecto FONDECYT N°1160429 denominado "Factores protectores y de riesgo asociados a la resiliencia socioeducativa en estudiantes vulnerables de cuarto a octavo año de enseñanza básica: un modelo multidimensional explicativo de las trayectorias de evolución del rendimiento escolar en Chile", financiado por CONICYT y desarrollado entre marzo de 2016 y marzo de 2020.

Una comparación del efecto del sesgo de respuesta en formato Likert y de respuesta expandida. Francisco José Abad¹, Luis Eduardo Garrido², María Dolores Nieto¹, Miguel A. Sorrel¹ y Eduardo García Garzón¹

¹Universidad Autónoma de Madrid, ²Pontificia Universidad Católica Madre y Maestra

El uso de ítems inversos en escalas psicológicas con formato Likert resulta controvertido. Por un lado, el uso de ítems inversos parece necesario para controlar el denominado sesgo de aquiescencia. Por otro lado, la inclusión de estos puede dar lugar a escalas aparentemente menos fiables y a distorsión de la estructura factorial. Se ha sugerido que la presencia de estos efectos puede venir mediada por las habilidades lingüísticas de los respondientes. Para resolver el problema del sesgo de respuesta se han propuesto nuevos modelos psicométricos, que añaden directamente un factor de aquiescencia, pero también el uso del formato de respuesta expandido, que reemplaza cada opción de respuesta con una sentencia completa. De acuerdo con Zhang y Savalei este formato resultaría robusto al sesgo de aquiescencia. El objetivo del presente estudio es valorar críticamente esta última aseveración. Para ello se aplicaron a una muestra de 609 estudiantes universitarios una prueba de aptitud verbal y tres versiones (dos en formato Likert y una con formato expandido) de la escala de autoestima de Rosenberg. Se comparó la fiabilidad de las puntuaciones en los distintos formatos y el ajuste de los modelos factoriales sin y con factor de aquiescencia, en grupos que diferían en aptitud verbal. Los resultados indican que el formato expandido también se veía afectado por el factor de aquiescencia, aunque el modelado psicométrico permitió recuperar la estructura correcta para ambos formatos. El efecto moderador de la aptitud verbal era pequeño, pero significativo en algunos casos. Finalmente, se establecen algunas recomendaciones en relación al uso de ítems inversos, del formato de respuesta expandida y el modelado del factor de aquiescencia.

COMUNICACIONES ORALES

Metodologías

Coordina: Eduardo García Cueto

Modelos de redes neuronales artificiales y sesgos de respuesta*. Eduardo García Garzón¹ y Álvaro Lozano¹

*Trabajo asociado a la ayuda AEMCCO-Julio Olea a jóvenes investigadores (otorgada al primer autor en 2019).

¹Universidad Autónoma de Madrid, ²Universidad de Salamanca

La expansión de las redes sociales y los estudios internacionales basados en cuestionarios han generado un importante interés por investigar la existencia de relaciones complejas entre variables psicológicas. Dada a su alta capacidad predictiva, técnicas de *machine learning* como las redes neuronales artificiales han comenzado a tener un papel muy relevante en dichos contextos. Sin embargo, el conocimiento psicométrico en lo relativo al tratamiento de las puntuaciones de ítems ha sido ignorado por los investigadores que aplican dichas técnicas estadísticas. En esta presentación se expondrá el efecto de la presencia del error de medida y/o sesgos de respuesta a la hora de estimar modelos de redes neuronales artificiales. Específicamente, se calculó qué impacto tiene el uso de puntuaciones directas o factoriales en la calidad de la predicción de las puntuaciones en un test de depresión. Para ello, se usaron los datos de las rondas tres y seis de la European Social Survey ($n \approx 90.000$). Los resultados demostraron que la aplicación puntuaciones corregidas permitía aumentar la capacidad predictiva tanto de redes neuronales como otras técnicas tradicionales de regresión (p. ej., regresión polinómica). Por otro lado, se obtuvo evidencia de que las redes neuronales artificiales mejoraban los resultados obtenidos por técnicas tradicionales (p. ej., regresión polinómica) en función del grado de no-linealidad existente. En esta presentación se mostrarán ejemplos de cómo los resultados de las redes neuronales artificiales pueden ser interpretados gracias a nuevas herramientas disponibles en el programa R. Se espera que este estudio permita una aproximación a futuros desarrollos estadísticos que combinen psicometría y técnicas de *machine learning* en el ámbito de las ciencias del comportamiento.

Una perspectiva crítica sobre el análisis de potenciales evocados basado en promedios. Juan Carlos Oliver Rodríguez¹ y Mirjam Moerbeek¹

¹Universitat Jaume I, ²Utrecht University

El ANOVA de medidas repetidas y el MANOVA son frecuentemente utilizados en el análisis de potenciales cerebrales evocados. Son típicamente aplicados a los promedios de un conjunto de ensayos con el fin de aumentar la fiabilidad de la señal electroencefalográfica. Sin embargo, el promediado puede implicar la pérdida de información sobre la matriz de covarianzas de las diferencias individuales aleatorias en los efectos del tratamiento y del tiempo. Estudios previos han mostrado que la falta de una adecuada especificación de la matriz de covarianzas puede conducir a errores de inferencia. El objetivo del presente estudio es comparar la ejecución de este procedimiento tradicional con el de un

análisis multinivel a partir de los ensayos individuales que permite una representación explícita de estos efectos aleatorios. El número de estímulos y participantes, así como la magnitud de los componentes de varianza y covarianza entre participantes, de interacción entre participantes y tratamiento, y de tiempo fueron manipulados en una simulación de un experimento sobre percepción facial. Un sesgo en el error Tipo I fue observado en el análisis de promedios en la condición de interacción aleatoria entre participantes y tratamiento. Se expondrán implicaciones más generales para el análisis de potenciales evocados.

Transparencia y calidad en la investigación en psicología: recursos para la mejora del diseño y el informe de investigación. Hilda Gambará¹ e Ignacio Montero¹

¹Universidad Autónoma de Madrid

Con el fin de hacer una psicología más transparente y facilitar la replicación de los trabajos se debe hacer un esfuerzo por equiparar la calidad del informe con la calidad de la investigación. El objetivo de esta comunicación es presentar los recursos actualmente disponibles para mejorar la calidad de los estudios cuantitativos y cualitativos y sus respectivos informes. Si bien se suele distinguir entre calidad de un informe y la calidad de los estudios, en la mayoría de las ocasiones, el informe es la única fuente de la que se dispone para evaluar la calidad de un estudio. En las últimas décadas se han publicado una serie de recomendaciones, escalas y listas de comprobación que pueden ser de utilidad para los investigadores. Más allá de las recomendaciones APA este tipo de herramientas no solo sirven a la hora de escribir el informe, sino también a la hora de diseñar e implementar el estudio. Se presentarán estos recursos tanto para estudios primarios como meta-analíticos y tanto para investigaciones cuantitativas como cualitativas. La presentación irá acompañada de ejemplos y aplicaciones, discutiendo la necesidad de su uso en diferentes ámbitos de la psicología.

Fundamentos teóricos para la construcción de test. Rafael Jesús Martínez Cervantes¹, Rafael Moreno Rodríguez¹ y José Muñiz Fernández²

¹Universidad de Sevilla, ²Universidad de Oviedo

Para que la construcción de tests y cuestionarios psicológicos supere procedimientos basados casi exclusivamente en la depuración de ítems a posteriori, se hace necesaria una mayor fundamentación teórica sobre la medición psicológica, acorde con perspectivas científicas y filosóficas contemporáneas. Para ello, se revisan la propuesta de Ryle sobre los constructos psicológicos, y la de Hubbard (2014) sobre la teoría de la información para fundamentar la medición científica. Estos planteamientos se desarrollan en coherencia con la propuesta de Moreno, Martínez y Muñiz (2018), que entienden los ítems de medición psicológica como dispositivos de comunicación entre examinador y examinado, y cuya función es la obtención de información sobre algún dominio psicológico del examinado. Se concluye sobre las potencialidades de esta fundamentación teórica en la construcción y uso de instrumentos de medida psicológicos.

Propiedades psicométricas de la versión española de la Escala de Crianza. Gemma López Fernández¹, Maite Barrios² y Juana Gómez Benito²

¹Corporación Sanitaria Parc Taulí, ²Universitat de Barcelona

La "Escala de crianza" de Arnold (O'Leary, Wolff y Acker de 1993) es uno de los instrumentos más utilizados en todo el mundo para evaluar las conductas de crianza disfuncionales. Sin embargo, su estructura factorial no está clara. Mientras que Arnold et al. (1993) encontraron una estructura de 3 factores, otros autores han hallado estructuras de 1 factor, 2 factores y 3 factores. El objetivo de este estudio es comparar diferentes modelos de estructura factorial de la "Escala de crianza" así como valorar las diferentes propiedades psicométricas de la estructura que mejor se ajusta a nuestra población. Para ello se utiliza el análisis factorial confirmatorio con el fin de valorar el ajuste de las diferentes estructuras factoriales propuestas hasta la fecha, incluyendo la estructura factorial original, en una muestra de 662 madres españolas con hijos entre 2 y 7 años. El modelo de 2 factores propuesto por Reitman et al. (2001) presenta los mejores índices de ajuste. Adicionalmente, sobre este modelo, se evalúa la consistencia interna, la fiabilidad test-retest al mes, la validez convergente y la validez concurrente utilizando el "Cuestionario de capacidades y dificultades" (Goodman et al. 1998), la "Escala de estrés percibido" (Cohen et al., 1983) y la "Escala del vínculo madre-hijo" (Taylor et al., 2005), en su versión en español. Tanto la validez convergente como la fiabilidad presentan valores adecuados; asimismo, la validez concurrente presenta correlaciones significativas en la dirección esperada. Por ello, concluimos que la "Escala de crianza" en su estructura de 2 factores propuesta por Reitman et al. (2001) proporciona puntuaciones fiables y válidas para evaluar el estilo de crianza en mujeres españolas. No obstante, se requieren futuros estudios que exploren la validez y la fiabilidad en población masculina.

Revisión y validación del Inventario de Depresión Estado-Rasgo en una muestra de embarazadas españolas. F. Javier del Río¹, Manuel A. García Sedeño¹, Yolanda Sánchez Sandoval¹, Laura Verdugo¹, María José Flores García² y Javier Ruiz Ruiz²

¹Universidad de Cádiz y (2) Centro de Salud de San Benito (Jerez)

Introducción: el embarazo es considerado uno de los períodos más críticos en la vida de la mujer. Investigaciones previas señalan que, debido a cambios hormonales y a los cambios propios en el estilo de vida, tanto durante el embarazo, como tras el parto, es común en la mujer la presencia de sintomatología depresiva. El objetivo de este estudio fue revisar la fiabilidad y la estructura factorial del "Inventario de depresión estado-rasgo" en una muestra de mujeres embarazadas españolas. Metodología: participaron 152 mujeres embarazadas, con edades comprendidas entre 18 y 43 años. Se utilizó el programa SPSS para realizar el análisis estadístico. Resultados: la correlación ítem-total de los ítems es buena (0,311 - 0,657). La fiabilidad del cuestionario fue alta, obteniéndose un índice alfa de Cronbach de 0,86. La prueba KMO (0,83) y la prueba de Bartlett ($p < 0,001$), aseguran la adecuación de la

realización del análisis factorial. Para el análisis factorial, al igual que realizaron los autores que validaron el IDER al español, se utilizó la técnica de ejes principales y la rotación Promax. Se encontró una estructura factorial de cuatro factores que explican el 55,4% de la varianza (F1 = 29,7%; F2 = 11,7%; F3 = 7,7%; F4 = 6,3%). Conclusiones: el "Inventario de depresión estado-rasgo" ha mostrado propiedades psicométricas adecuadas para evaluar a mujeres embarazadas.

Concordancia entre la versión de la "Escala de Derrota y Atrapamiento" y la versión corta usando modelos de Rasch. Jorge L. Ordoñez Carrasco¹, María Sánchez Castelló¹ y Antonio J. Rojas Tejada¹

¹Universidad de Almería

El "modelo integrado motivacional-volitivo" contempla la ideación suicida como una reacción psicológica a una situación que se percibe como una derrota junto a la percepción de estar atrapado/a y sin posible escape ni rescate. Las escalas diseñadas para medir sendas percepciones conciben la derrota y el atrapamiento como constructos independientes. Hallazgos posteriores sugieren que ambas medidas pueden ser un único constructo. La escala resultante al combinar los ítems de las escalas originales resulta inadecuada para el contexto clínico debido a su longitud. Con el fin de obtener un instrumento breve para medir la percepción de derrota y atrapamiento, Griffiths y colaboradores diseñaron una escala de ocho ítems orientada al contexto clínico. El presente trabajo tiene como objetivo establecer la concordancia entre los ítems de las escalas originales de derrota y atrapamiento y la versión breve. Para ello, se administró un cuestionario online con las escalas y se llevó a cabo un procedimiento de equiparación con el mismo grupo de personas usando un modelo de Rasch (modelo de escalas de clasificación). Participaron 644 personas de 18 a 35 años (51,2% mujeres; edad media = 25,91 y DT = 5,14) reclutados a través de un muestreo por cuotas (edad, sexo y nivel educativo) acorde a la composición de la población de jóvenes adultos españoles. Los resultados mostraron un aceptable ajuste para ambas escalas (versión completa y versión corta). La correlación entre ambas escalas fue 0,97 y el índice de reducción de incertidumbre (RIU) fue 76%. La eficiencia para la versión corta en relación con la versión completa fue adecuada. Las dos versiones fueron capaces de diferenciar las puntuaciones de las personas localizadas en diferentes puntos a lo largo del continuo derrota-atrapamiento. En síntesis, la versión completa y corta de la escala de derrota-atrapamiento han mostrado una alta concordancia.

Scale of sexual self-concept for young people and adults in northern Chile. Rodrigo Ferrer Urbina¹, Diego Henríquez¹ y Gerald Sepúlveda¹

¹Universidad de Tarapacá

Background/objective: STIs and HIV/AIDS are an important public health problem, so it is necessary to identify protective factors associated with such behaviours, such as the role of Self-concept. In the Spanish-speaking context, there are no instruments with sufficient evidence of reliability and validity for this. The objective of this work is the development of an instrument to evaluate sexual self-concept in young people and adults. Method: Non-probabi-

listic sampling, by availability. 792 participants from Arica and Iquique, aged between 17 and 53 years old, with 66.2% women ($n = 500$), 33.6% men ($n = 258$). Results: The final instrument consists of 16 items; sexual self-esteem (items = 4); sexual self-efficacy (items = 4); assertive sexual communication (items = 4); assertive sexual behaviour (items = 4). The structure identified provides adequate levels of reliability ($\omega > 0,8$) and presents evidence of validity based on the internal structure of the test, using ESEM (RMSEA = 0,060; CFI = 0,99; TLI = 0,98). Conclusions: The sexual self-concept instrument has adequate psychometric properties to evaluate sexual self-concept.

Desarrollo de una escala de búsqueda de sensaciones sexuales para jóvenes del norte de Chile. Rodrigo Ferrer Urbina¹ y Patricio Mena Chamorro¹

¹Universidad de Tarapacá

Las conductas sexuales de riesgo son la principal vía de transmisión de VIH/SIDA e ITS. Se ha evidenciado que las CSR poseen una fuerte relación con aspectos específicos de la personalidad, particularmente, la búsqueda de sensaciones sexuales (BSS). Pese a la utilidad de las BSS para estudiar CSR, en la literatura hispanohablante no existen instrumentos con evidencia psicométrica para su uso y, en otros idiomas, solo se cuenta con la escala de BSS de Kalichman y colaboradores (1994), la cual se encuentra desactualizada en contenidos y en evidencias de validez. En este contexto, este estudio tuvo por propósito desarrollar un instrumento de medida, acorde con estándares psicométricos contemporáneos, para evaluar BSS en jóvenes y adultos residentes del Norte Grande de Chile. El estudio fue transversal de tipo instrumental, con un muestreo tiempo-espacio correspondiente a 813 estudiantes de educación superior de las ciudades de Arica e Iquique. La escala quedó compuesta por 8 ítems, destinados a evaluar dos dimensiones: (1) búsqueda de emociones y (2) tendencia al aburrimiento. Los resultados evidencian adecuados niveles de fiabilidad ($\omega > 0,80$) y evidencias de validez basada en la estructura interna del test, mediante ESEM (TLI = 0,974; CFI = 0,988; RMSEA = 0,053) y CFA (TLI = 0,966; CFI = 0,977; RMSEA = 0,060).

Análisis factorial de variables asimétricas con el modelo de respuesta graduada y bajo el supuesto de normalidad multivariante. Noelia Yoelina Minaya Ventura¹ y Javier Revuelta Menéndez¹

¹Universidad Autónoma de Madrid

Cuando las variables observadas en un análisis factorial presentan asimetría, al estimar el modelo mediante máxima-verosimilitud se asume que las variables observadas siguen la distribución normal multivariante. Olsson demostró que cuando se combinan variables con asimetría positiva y negativa, dicho modelo sobrestima el número de factores que realmente subyacen a los datos. Una situación práctica en la que se puede dar este caso es cuando se trabaja con cuestionarios de actitudes o personalidad en los que se combinan ítems directos e inversos para controlar la aquiescencia. Esto puede originar datos con asimetría mixta. El problema de la asimetría puede abordarse aplicando el modelo de

respuesta graduada (MRG) de la teoría de respuesta al ítem en lugar del modelo factorial con variables normales. El MRG asume que las variables observadas se obtienen categorizando variables latentes distribuidas normalmente. En este trabajo se comparan los resultados del modelo factorial y el MRG aplicados a datos de opiniones de estudiantes universitarios acerca de sus estudios y perspectivas de futuro. El cuestionario combinaba ítems directos e inversos. Se analizó el número de factores estimados por ambos modelos con los datos originales y los datos recodificados para que todos los ítems tuvieran la misma orientación. Los resultados muestran que el modelo factorial produce más factores que el MRG.

Un modelo AFC bifactor para el compromiso laboral: planteamiento y un estudio empírico. Raul Castañeda Vozmediano¹ y Carmen Ximénez¹

¹Universidad Autónoma de Madrid

El compromiso laboral es una de las variables que ha generado mayor investigación en el contexto del trabajo y las organizaciones. Es ampliamente reconocido que se trata de un constructo multidimensional y que está relacionado con otras actitudes como la satisfacción laboral y la intención de abandono. El modelo de Meyer y Allen (1991) de tres factores (compromiso afectivo, compromiso normativo y compromiso de continuación) es el marco teórico más aceptado y que ha generado un mayor número de estudios empíricos tal y como se documenta en el meta-análisis de Meyer, Stanley, Herscovitch y Topolnytsky (2002). En este póster se presenta un estudio empírico con datos para una muestra de 591 participantes que respondieron a una encuesta de opinión sobre diferentes actitudes hacia su organización del trabajo. Se estudió la bondad de ajuste de diferentes modelos de análisis factorial confirmatorio (AFC) para los ítems de compromiso laboral, encontrándose que el modelo *bifactor* es el que ofrece mejores resultados, lo que sugiere que tendría sentido plantear un factor general de compromiso laboral, además de los tres factores específicos de Meyer y Allen. Este tipo de modelo no ha sido estudiado en la investigación previa y puede ser especialmente interesante ya que los modelos *bifactor* tienen ventajas respecto a otros modelos cuando el interés reside en la predicción de variables externas (Chen, West y Sousa, 2006) y este es el caso en el constructo del compromiso laboral que es el principal antecedente de la intención de abandono.

Métodos de Simulación Monte Carlo y Remuestreo: un estudio comparativo evaluando el ajuste de modelos de escalas de Mokken. Goma Said Mohamed Abdelhamid¹, Juana Gómez Benito¹ y María Dolores Hidalgo Montesinos²

¹Universitat de Barcelona, ²Universidad de Murcia

Antecedentes: los métodos de simulación Monte Carlo y de remuestreo desempeñan un papel fundamental en la investigación metodológica. El método de remuestreo selecciona aleatoriamente una gran cantidad de muestras de una población para obtener ciertas conclusiones sobre dicha población. En la simulación Monte Carlo se generan un gran número de muestras independientes para extraer las conclusiones. El análisis de escalamiento

de Mokken se ha utilizado para determinar la dimensionalidad de un conjunto de ítems bien usando el procedimiento de selección de ítems automatizado (AISP) o el algoritmo genético (GA). En este estudio se compara la eficacia de AISP y GA en la determinación del número de factores en ítems dicotómicos y se comparan dos métodos de simulación (Monte Carlo y remuestreo). Método: las variables manipuladas fueron: número de factores (2); correlación entre factores (ortogonales; $\rho = 0$), tamaño de la muestra (250), número de ítems por factor (5). En el método de remuestreo para una condición dada: (1) se simula un conjunto de datos grande (100.000); (2) se seleccionan aleatoriamente 1.000 submuestras de tamaño muestral fijado en las condiciones del estudio ($n = 250$); y (3) se analizan usando AISP y GA. En el método Monte Carlo: (1) se simulan 1.000 bases de datos aleatorias e independientes y (2) se analizan los datos utilizando AISP y GA. En ambos casos las matrices de datos fueron simuladas usando un modelo TRI logístico 2-p compensatorio multidimensional. Resultados: los métodos AISP y GA proporcionaron estimaciones casi idénticas. Los resultados fueron muy similares independientemente del método de simulación usado. Conclusión: en general, ambos métodos de simulación Monte Carlo y remuestreo pueden ser utilizados como una herramienta para comprender los aspectos esenciales relacionados con el comportamiento estadístico de modelos y/o técnicas de análisis estadístico en diferentes condiciones.

Una aproximación lineal logística para la detección de aprendizaje durante el test. José Héctor Lozano Bleda¹ y Javier Reuvelta Menéndez¹

¹Universidad Autónoma de Madrid

Se presenta una variante del modelo de aprendizaje específico de operación, perteneciente a la familia de modelos de Rasch. El modelo consiste en una generalización del modelo lineal logístico de test (LLTM; Fisher, 1973), en la cual la dificultad del ítem depende de las operaciones implicadas en el ítem, así como del número de veces en que dichas operaciones han sido presentadas en ítems anteriores. La nueva variante permite una mejor representación del aprendizaje derivado de la práctica previa cuando la matriz estructural no es binaria. La aplicación del nuevo modelo para la detección de aprendizaje se ilustra con datos reales de un test de evaluación de la aptitud lógica. Los resultados muestran cómo la incorporación de la práctica previa al modelo lineal logístico mejora el ajuste del modelo, así como la predicción de la dificultad de los ítems estimada con el modelo de Rasch. Agradecimientos: la estimación fue realizada con los recursos del Centro de Computación Científica de la Universidad Autónoma de Madrid (CCC-UAM).

Evaluación del emotion regulation checklist para niños y adolescentes mediante el CET-R. Jennifer Pérez Sánchez¹, Gerardo Prieto¹ y Ana R. Delgado¹

¹Universidad de Salamanca

El empleo adecuado de las pruebas psicológicas tanto en ámbitos académicos como en la práctica profesional requiere de procedimientos sistemáticos que informen sobre su calidad y utilidad.

El objetivo de este trabajo es evaluar las propiedades psicométricas del Emotion Regulation Checklist, uno de los instrumentos más empleados en la evaluación de la regulación emocional en niños y adolescentes. Para la valoración de los resultados obtenidos sobre este test en los 72 artículos de investigación empírica publicados en los últimos seis años, se ha empleado el "Cuestionario para la evaluación de los tests - revisado" (CET-R). Se ha hallado que la mayor parte de las investigaciones emplearon diseños de consistencia interna; el promedio de los coeficientes fue elevado, calificable como excelente según el CET-R. En cuanto a la validez, se emplearon principalmente correlaciones entre las puntuaciones del test y marcadores y/o criterios. En líneas generales, las características psicométricas del ERC resultan adecuadas.

Análisis multinivel y medidas del tamaño del efecto en diseños de caso único: un estudio piloto. Cristina Rodríguez Prada¹, Ricardo Olmos Albacete¹ y José Manuel Caperos Montalbán²

¹Universidad Autónoma de Madrid y ²Universidad Nebrija

Durante los últimos años el interés que ha despertado el análisis cuantitativo de los diseños de caso único es palpable. En particular, parte de la investigación se ha centrado en estudiar cuáles de las diferentes medidas de tamaño del efecto son más adecuada en este tipo de diseños (Parker, Vannest, y Davis, 2011) y en la adecuación del uso del análisis multinivel en este tipo de diseños (p. ej., Ferron et al., 2009; Moyaert et al, 2017; o Shadish, 2013). El objetivo de esta investigación es estudiar mediante simulación y en una muestra piloto de artículos de diseño de caso único AB las posibilidades de incorporar alguna medida de tamaño del efecto desde la perspectiva multinivel (i.e., basadas en el coeficiente de correlación intraclase, medidas estandarizadas de los efectos fijos, etc.; Lorah, 2018), aunando por lo tanto estas dos cuestiones que recientemente tienen interés dentro de este tipo de diseños.

Comparison of false positive and false negative rates of two indices of individual reliable change: Jacobson-Truax and Hageman-Arrindell methods. Rodrigo Ferrer¹ Antonio Pardo² y Willem Arrindell³

¹Universidad de Tarapacá, ²Universidad Autónoma de Madrid, ³University of Groningen

The objective of the present study was to learn about the performance of two distribution-based methods designed to evaluate individual change in pre-post designs: the Jacobson-Truax reliable change index [RCI] and the Hageman-Arrindell [HA] approach. The novelty of this study is that reliability has been estimated by the omega_c coefficient rather than with the alpha coefficient or the test-retest correlation. We have evaluated the false positive and false negative rates associated with RCI and HA through an extensive simulation study controlling for, among other things, the effect size, the shape of the distributions, and the sample sizes. Regarding the rate of false positives, only the HA statistic provided acceptable results. Regarding the rate of false negatives, both statistics offered similar results and both could claim to offer acceptable rates when Ferguson's stringent criteria were used to define effect sizes as opposed to when the conventional criteria advanced by Cohen were employed.

Un estudio piloto para evaluar el estigma hacia el trastorno mental desde la metodología implícita IRAP. Blanca González Moreno¹, Ricardo Olmos Albacete¹ e Isabel Cabrera Lafuente¹

¹Universidad Autónoma de Madrid

La deseabilidad social es un problema persistente que dificulta la medición de determinados constructos socialmente sensibles como el racismo, machismo o el estigma. Este problema se acentúa cuando la evaluación es de altas consecuencias. De entre las múltiples soluciones aportadas, en este trabajo se presenta un estudio piloto mediante el procedimiento de evaluación relacional implícita (IRAP). La metodología IRAP evalúa las actitudes teniendo en cuenta las latencias de respuesta a asociaciones entre adjetivos con valencia positiva y negativa, y unos estímulos atributo (viejo vs. joven; gordo vs. delgado). La persona debe responder con latencias de respuesta muy bajas, lo que hace muy improbable que la deseabilidad social participe en la medición. Además, esta metodología contiene mejoras en relación a otras metodologías implícitas clásicas como los test de asociación implícita (IAT) porque permite analizar la direccionalidad de la asociación entre los conceptos al no medirlo de forma relativa. En este trabajo se presenta un estudio piloto en una muestra de estudiantes recién graduados en psicología y en estudiantes pertenecientes a otros grados ($n = 32$). Se diseñó una tarea IRAP para evaluar el estigma hacia el trastorno mental. También se evaluó el estigma con el test explícito CAMI de actitudes hacia la enfermedad mental. Los resultados mostraron diferencias estadísticamente significativas entre los graduados en psicología y los de otros grados (menos estigma hacia el trastorno mental entre los psicólogos; $t = 2,78$, $p = 0,005$). Los resultados detectados con la metodología implícita fueron más elevados en términos del tamaño del efecto que con el test explícito CAMI y, además, permitió detectar efectos adicionales que no son sensibles a los test explícitos. Estos resultados sirven para discutir las implicaciones que tiene la metodología implícita IRAP como una vía novedosa para abordar la evaluación de actitudes socialmente sensibles.

Uso de modelos TRI para el análisis de la batería psicopedagógica Evalúa. Ana Martínez Dorado¹ y David Cabello Gonzalez¹

¹GiuntiEOS PSYCHOMETRICS

El análisis psicométrico de la mayor parte de baterías de test en España se realiza mediante el paradigma clásico de la TCT propuesto por Spearman. Sin embargo, durante los últimos años, se ha desarrollado un nuevo paradigma dentro de la psicometría: la TRI. Es por ello que el objetivo del presente póster es la aplicación de dos modelos de la TRI (modelo de dos parámetros y modelo de respuesta graduada) a la batería psicopedagógica con mayor antigüedad y experiencia en nuestro país, la batería "Evalúa", ideada por García Vidal et al., y publicada por GiuntiEOS Psychometrics. En concreto, en el presente póster analizaremos el ajuste de: (a) una prueba de razonamiento deductivo al modelo de dos parámetros, para comprobar que, como se espera, la función de información del test tiene su máximo en niveles ligeramente inferiores a la media en la escala de habilidad, cumpliendo con su funcionamiento de *screening* de problemas cognitivos; (b) una escala psicosocial al modelo de respuesta graduada de

Samejima, para comprobar su utilidad como *screening* mediante la función de información del test. Así mismo, se espera que los parámetros de los ítems respondan a la lógica del modelo propuesto. Samejima propone un total de parámetros b igual al número de categorías del ítem menos uno. Si el ítem está ordenado, esperamos que para elegir categorías superiores se requiera un nivel de habilidad superior en el rasgo. Es por ello que se ha comprobado, en cada ítem (con cuatro opciones de respuesta), si el parámetro b_{12} es menor al parámetro b_{23} y a su vez éste es inferior al parámetro b_{34} .

VARIABLES CONTEXTUALES Y ESTRESORES PREDICTORES DE LA SOBRECARGA DEL CUIDADOR. F. Javier García-Castro¹ y María J. Blanca¹

¹Universidad de Málaga

Ejercer el rol de cuidador puede tener consecuencias negativas para los cuidadores, siendo la más importante la sobrecarga. Esta ha sido definida como el impacto negativo percibido por los cuidadores en las diferentes áreas de su vida como resultado de las restricciones del rol de cuidador. Según el modelo de proceso del estrés, las variables asociadas a la sobrecarga se pueden clasificar en contextuales, estresores primarios (los relativos a la enfermedad, como el nivel de deterioro cognitivo de la persona enferma) y estresores secundarios (estresores que no se relacionan directamente con el rol de cuidador pero que influirán en este, como dificultades laborales o económicas). Las investigaciones sobre la asociación entre sobrecarga y estas variables arrojan resultados no consistentes. El objetivo del presente trabajo es encontrar aquellas variables contextuales y estresores que predicen mejor la sobrecarga del cuidador. Para ello, se han recogido datos de una muestra de cuidadores de personas con demencia de diferentes asociaciones de familiares de Alzheimer en la provincia de Málaga, y se ha llevado a cabo un análisis de regresión jerárquica, introduciendo tres bloques relativos a variables contextuales, estresores primarios y estresores secundarios. Los resultados indican que los estresores secundarios son los que mejor predicen la sobrecarga del cuidador, siendo las dificultades económicas percibidas, y la limitación del tiempo libre las variables estadísticamente significativas. Estos resultados sugieren que los estresores secundarios, especialmente las dificultades económicas y la limitación del tiempo libre, deberían evaluarse cuando se desarrollen programas de intervención en cuidadores de personas con demencia, ya que aquellos que perciban dificultades económicas y que no dispongan de tiempo para el ocio tendrán un mayor riesgo de presentar sobrecarga.

Comparaciones múltiples post-hoc para el ANOVA de un factor: ¿Cuál utilizar? Álvaro Postigo¹, Joel Juarros Basterretxea¹, Gema Aonso Diego¹, Pelayo Montes Álvarez¹, Álvaro Menéndez Aller¹ y Eduardo García Cueto¹

¹Universidad de Oviedo

Introducción: el análisis de la varianza es una de las técnicas más extendidas y utilizadas en distintas disciplinas. Específicamente, el ANOVA de un factor, se trata de una prueba paramétrica que asume: (a) independencia de los datos, (b) distribución normal de los datos y (c) homocedasticidad. Sin embargo, debido a la varie-

dad de pruebas *post hoc* que nos ofrecen los paquetes estadísticos no hay acuerdo sobre cuál es la prueba que mejor funciona tanto si se cumplen todos los supuestos paramétricos como si se incumple alguno de ellos (desigualdad de varianzas, grupos con tamaños diferentes). El objetivo, atendiendo a la igualdad de varianzas y al equilibrio de los tamaños muestrales, fue analizar qué prueba *post hoc* funciona mejor: HSD Tukey, Scheffé, DMS, Bonferroni, Sidak, Gabriel, Hochberg, Tamhane, T3 de Dunnett, Games-Howell y T de Dunnett. Método: se realizaron cuatro estudios de simulación como resultado del cumplimiento o no de dos condiciones: el equilibrio del tamaño muestral de los grupos y la homocedasticidad. Para cada estudio se hicieron 10.000 réplicas. Además, en cada uno de los estudios se consideraron diferentes tamaños muestrales: $n = 30$, $n = 90$, $n = 150$, $n = 300$, $n = 750$, $n = 1500$ y $n = 3000$. Resultados: la prueba *post hoc* que mejor funciona depende, como era esperable, de las características de las muestras.

Un diseño de cohortes para el estudio de variables que cambian en función del tiempo. Manuel Lage¹ y Alicia Risso²

¹Unidad de tratamiento del alcohol y conductas adictivas, A Coruña, ²Universidad de A Coruña

Los estudios sobre la recuperación espontánea de funciones cognitivas deterioradas por el consumo abusivo de alcohol, una vez que se está en fase de abstinencia-deshabitación, son especialmente complejos y difíciles de llevar a cabo, en buena medida por características propias de la población objeto de estudio. Así, la comorbilidad con otras patologías, físicas o psicológicas, u otros problemas de índole social, por ejemplo, hacen poco viable la adherencia a una investigación de tipo longitudinal. Por este motivo hay muy pocos estudios sobre esta temática y solo algunos van más allá de los 18 meses de abstinencia. En esta investigación se estableció que los datos debían recogerse en intervalos de seis meses, desde el inicio de la abstinencia hasta los 24 meses. Para ello, se solicitó la participación voluntaria de pacientes en tratamiento de deshabitación en una unidad dependiente del Servicio Galego de Saúde y se constituyeron cinco cohortes de 20 personas, correspondientes a los lapsos de tiempo establecidos. Las 27 variables cognitivas que se midieron suministraron datos sobre atención, memoria, leguaje, orientación y comprensión. Para el análisis de los resultados primero se comprobó que las cohortes eran similares entre sí y que no se diferenciaban de la población, en variables de interés como edad, sexo, nivel de estudios, edad de inicio y años de consumo problemático. Esto permitió configurar un grupo control sin historial de alcoholismo, equiparado en dichas variables. Los resultados mostraron que la mejora de las funciones deterioradas observada con un diseño de cohortes es equivalente a la obtenida en estudios longitudinales y, además, permite aportar datos más allá de los 18 meses de abstinencia, que es el punto de corte habitual. Esto indica que se puede proporcionar información sobre variables cuya evolución temporal es de fundamental interés, en menos tiempo y con más control de la situación, sin tener que recurrir a costosos estudios longitudinales, muy difíciles de llevar a cabo sin financiación y en los cuales la mortandad experimental amenaza la validez de los resultados obtenidos.

Malinterpretaciones de los intervalos de confianza en estudiantes universitarios. Laura Badenes Ribera¹, María C. Fuentes¹ y Laura Galiana¹

¹Universitat de València

Desde su introducción en la ciencia, las pruebas de significación de la hipótesis nula (NHST) son el método estadístico más utilizado en psicología para realizar la inferencia estadística. Sin embargo, estudios previos han mostrado consistentemente que los investigadores, profesores y estudiantes en psicología cometen errores de interpretación del valor p . Para contrarrestar estas malas interpretaciones, se han propuesto muchas soluciones. Por ejemplo, complementar los resultados de las NHST con los intervalos de confianza de la estimación. Por lo tanto, saber interpretar los intervalos de confianza es una habilidad fundamental para el realizar una lectura crítica de la evidencia científica. El objetivo de este estudio fue examinar los errores de interpretación de los intervalos de confianza y su interpretación correcta en estudiantes universitarios con formación en estadística del grado de psicología. Se administró un cuestionario con 8 afirmaciones sobre interpretación de los ICs (6 afirmaciones falsas y 2 afirmaciones correctas) a una muestra de 109 estudiantes universitarios de 2º curso del grado de psicología, con una edad media de 23,56 (DT = 7,92). El 81,4% de los participantes fueron mujeres y el 18,6% hombres. Los hallazgos sugieren que la mayoría de los estudiantes encuestados presentan problemas de comprensión de los ICs. Estos resultados van en la línea de estudios previos sobre errores de interpretación de términos estadísticos (ejemplo, interpretación del valor p) y ponen de manifiesto la necesidad de una reeducación estadística.

Evaluación de la calidad metodológica en los meta-análisis publicados en revistas españolas. Hilda Gambara¹, Juan Ignacio Duran Pacheco² y Álvaro Santana¹

¹Universidad Autónoma de Madrid, ²Centro Universitario Cardenal Cisneros

Una de las primeras críticas que recibió el meta-análisis (MA) es la conocida por "garbage in, garbage out", que implica la imposibilidad de realizar buenas síntesis a partir de trabajos de escasa calidad. Desde entonces, la evaluación de la calidad de los estudios primarios (CEP) en los MA es objeto de debate. La consideración de la CEP ha conllevado el desarrollo de múltiples herramientas y de investigaciones en torno a su utilización en las revisiones sistemáticas, así como su repercusión sobre los resultados, fundamentalmente en el entorno médico. En el ámbito de la psicología se han desarrollado menos estudios. Recientemente, Hohn, Slaney y Tafreshi (2018) evaluaron el análisis de la calidad de los estudios en una muestra de 386 MA, encontrando un escaso reporte sobre dicha calidad (55,2% en los MA sobre intervenciones y un 20,1% en el resto de MA) y una gran heterogeneidad entre los instrumentos de evaluación. El objetivo de este poster es presentar un avance de los resultados sobre la consideración de la calidad de los estudios primarios en los MA publicados en las principales revistas españolas en los últimos años. Concretamente, analizamos si la codificación de la CEP en los MA sobre eficacia de intervenciones o tratamientos es una práctica común.

También se estudia el propósito de esta evaluación y la relación entre la CEP y los resultados reportados en dichos MA.

La eficacia de la electro-estimulación neuromuscular en la recuperación tras un ictus: protocolo para una revisión sistemática y meta-análisis en red. Trinidad Sentandreu Mañó¹, Laura Badenes Ribera¹, Julio Sánchez Meca², José Manuel Tomás¹ y José A. López López²

¹Universitat de València, ²Universidad de Murcia

Las técnicas fisioterápicas de electroestimulación neuromuscular se utilizan frecuentemente en los programas de rehabilitación de pacientes que han sufrido un ictus. Sin embargo, existen diferencias importantes dentro de este grupo de intervenciones que pueden afectar a su efectividad y examinar estas diferencias constituye una pregunta clínicamente relevante para los fisioterapeutas. En este trabajo se plantea una revisión sistemática para identificar todos los estudios relevantes de este campo. Además, se integrarán cuantitativamente los resultados de los estudios a través de un meta-análisis en red, una técnica que permite crear una "red" de intervenciones a través de los estudios y que posibilita la inclusión de un mayor número de estudios primarios y un análisis riguroso de la heterogeneidad entre distintas intervenciones. Se incluirán ensayos clínicos aleatorizados donde se hayan evaluado uno o más tipos de electroestimulación neuromuscular aplicada al miembro superior en adultos que hayan sufrido un ictus. Para conseguir una red de intervenciones conectadas entre sí, se incluirán también varios comparadores frecuentemente empleados en estos estudios como estimulación fingida, técnicas fisioterápicas convencionales y no convencionales. Las variables dependientes principales del estudio serán las relacionadas con la recuperación motora de los pacientes, considerando por separado las categorías de deterioro motor y capacidad motora funcional. Además, se extraerán los datos aportados en los estudios acerca de los indicadores de la recuperación cerebral de los pacientes (p. ej., neuroplasticidad). Se evaluará el riesgo de sesgo en los estudios primarios y se llevarán a cabo análisis de sensibilidad para determinar la robustez de los resultados del meta-análisis. Finalmente, se examinará el papel de posibles variables moderadoras, incluyendo el grado de afectación de los pacientes de la muestra, la duración de la intervención, los parámetros de la corriente eléctrica utilizada y el período transcurrido entre el ictus y el comienzo de la intervención.

El método Delphi: un estudio bibliométrico. Georgina Guilera¹, Marina Navarro Gimeno¹, Anna Barquiner Campodarve¹ y Maite Barrios¹

¹Universidad de Barcelona

El método Delphi es un procedimiento que tiene como objetivo obtener información consensuada sobre una temática mediante la consulta a un panel de expertos. Por sus características puede ser empleado en un amplio abanico de áreas de investigación. Desde una perspectiva bibliométrica, este estudio pretende, por un lado, analizar la productividad científica sobre los estudios Delphi y, por otro lado, explorar las áreas temáticas en las que se ha estudiado o empleado esta técnica. Para ello, se realizó una búsqueda en la base de datos Social Sciences Citation Index de la colección principal de la Web of Science, sin restricción idio-

mática. Se analizaron 4.982 artículos científicos. Se aplicaron tres leyes bibliométricas, a saber, la Ley de Lotka para analizar la productividad de los autores, la ley de Bradford para explorar la productividad de las revistas y la ley de Price para examinar el crecimiento de la producción científica. También se realizó un análisis de coocurrencias de términos en los títulos y resúmenes, limitando a 20 el mínimo de ocurrencias. Emergieron cuatro áreas temáticas, una de las cuales de especial interés para la Psicometría. Ésta quedó representada por palabras tales como item, scale, instrument, content, validity, las cuales hacen referencia a la aplicación del método Delphi en la evaluación de la validez de contenido de los instrumentos de medida. En conclusión, este estudio bibliométrico permite caracterizar la literatura científica sobre el método Delphi, destacando la utilidad de este procedimiento para obtener evidencias de validez de los tests.

Violencia sufrida en la pareja adolescente como constructo multifactorial. F. Javier del Río¹, Jerónimo Acosta Quintero¹, Manuel A. García Sedeño¹ y Gabriel de la Torre González¹

¹Universidad de Cádiz

Introducción: la violencia en las relaciones de pareja adultas comienza en la adolescencia y descubrir los factores que influyen en su realización puede servir para diseñar programas de prevención de calidad. El objetivo de la presente investigación es describir el comportamiento de los/as adolescentes del entorno geográfico de la Bahía de Cádiz en relación con la violencia de pareja mediante la combinación de factores referidos a los diferentes niveles ecológicos de interacción. Método: la presente investigación se basa en un estudio descriptivo con metodología correlacional, como un diseño transversal ex post facto, mediado a través de un cuestionario. Participaron en el estudio 629 chicos y chicas, estudiantes de secundaria y bachillerato de tres institutos de Cádiz. Resultados: el modelo formado por las variables pareja violenta en relaciones anteriores, sexismo hostil/benevolente, exposición a violencia en el entorno, amigos/as de la pareja violenta en sus relaciones, ideas desadaptadas sobre el amor y la pareja, malas relaciones familiares en la infancia y recurso a la violencia reactiva/proactiva explican el 40% de la varianza de la variable violencia sufrida en la pareja. Conclusión: puede aceptarse que el modelo explicativo propuesto de violencia sufrida en la pareja tiene un buen ajuste y permite mantener la hipótesis de que la victimización en las relaciones entre adolescentes es mejor explicada por la combinación de factores, desde una perspectiva ecológica.

El nivel comunicativo: la importancia de la comunicación en el proceso de investigación científica. Raimundo Aguayo Estremera¹

¹Universidad de Málaga

La psicología científica ha asumido predominantemente el método hipotético-deductivo. Éste implica una serie de pasos ordenados y convenidos por los científicos, que habitualmente se agrupan en tres niveles: teórico-conceptual, técnico-metodológico y estadístico-analítico. Esta clasificación permite estudiar el proceso de investigación a partir de la descripción, agrupación y diferenciación de las acciones que llevan a cabo los científicos. No obstante, no contempla, al menos no en el mismo grado, todas aquellas acciones destinadas a la comunicación de los hallazgos científicos. En

la actualidad, la forma habitual de comunicar hallazgos es la publicación de artículos en revistas especializadas. Desde hace unos años, la necesidad y la presión por publicar artículos han ido en aumento. Entre los motivos se pueden citar el criterio de evaluación de la ciencia y el sistema de publicación. En cuanto al primero, la cantidad de artículos publicados y la calidad de las revistas en las que se publican se han convertido en los criterios fundamentales para la evaluación de la actividad de los científicos. En cuanto al segundo, el proceso de revisión y aceptación de artículos es percibido como lento o caro. Todo esto, enmarcado en un contexto globalizado, ha incrementado la dificultad para publicar. Por ello, desde el inicio del proceso de investigación, los científicos se preocupan considerablemente por la forma de redactar las investigaciones, los resultados obtenidos y las revistas que más rápida y fácilmente puedan publicarlos. En esta situación en la que las publicaciones se han convertido en el corazón de la ciencia, se hace necesario poner de manifiesto la importancia de estas acciones destinadas a la publicación. Dentro del conjunto de iniciativas para ello, en este trabajo se propone añadir una nueva categoría a la clasificación habitual de los pasos de investigación en la que éstas queden reflejadas. Tal categoría podría denominarse nivel comunicativo.

Modelos computacionales del lenguaje y su relación con la medición de la personalidad: un estudio piloto con el modelo LSA. Manuel Aenlle¹, Ricardo Olmos Albacete¹ y José Manuel Hernández López¹

¹Universidad Autónoma de Madrid

Durante los últimos años, destacados autores han estudiado la relación que hay entre el lenguaje y los estilos de personalidad de las personas. Además, cada vez hay más estudios que utilizan los modelos lingüísticos computacionales como un medio para predecir la personalidad de los individuos (p. ej., Kwantes et al., 2013). En este estudio se indagó en una pequeña muestra de 32 personas el poder predictivo del modelo computacional LSA a la hora de evaluar los rasgos de personalidad extraversión, psicoticismo y neuroticismo, medidos con el instrumento EPQ-R de Eysenck. Para ello se generaron seis situaciones prototípicas de cada uno de los rasgos de personalidad (p. ej., ir a una fiesta o enfrentarse a un examen) y se les pidió a los sujetos que se describiesen en esas situaciones. Para medir la personalidad, se aplicaron tres procedimientos: (1) se les administró el EPQ-R, (2) dos jueces humanos valoraron los tres rasgos a partir de los textos escritos por los sujetos y (3) el modelo LSA calculó sus puntuaciones de personalidad. A su vez, para conseguir que el modelo LSA tuviese un "conocimiento" adecuado de la personalidad, se entrenó con un corpus general de lenguaje común extraído a partir de la Wikipedia y con un corpus de tipo más técnico, con definiciones más específicas de conceptos relevantes, como por ejemplo, los rasgos de personalidad a evaluar. Los resultados mostraron una relación baja entre las puntuaciones de personalidad derivadas de los tres métodos, aunque en el rasgo neuroticismo se encontraron relaciones más elevadas que en los otros dos rasgos. Se discuten los resultados obtenidos, sus implicaciones y las posibles líneas futuras que permiten relacionar la personalidad, el lenguaje empleado por las personas y los modelos computacionales.

Ji-cuadrado para tablas de contingencia 2x2 y múltiples en QB64: JICUADRA.BAS. Ernesto Juan Darías Morales¹ y Pablo García Medina¹

¹Universidad de La Laguna

JICUADRA.BAS v. 1.1 es un programa de ordenador que utiliza el estadístico ji-cuadrado de homogeneidad e independencia para tablas de contingencia múltiple, el test W , escrito para QB64, aunque también corre en otras versiones de BASIC, todos ellos son *freeware* en Internet y pueden traerse por Google. Funciona tanto con MS-DOS (v. 5.0 a v. 6.11) como con varios tipos de Windows (3.11 a 10, 32 y 64 bits), Linux, Mac OSX y Android, en cualquier sistema operativo con VirtualBox (Oracle VM VirtualBox, hasta v. 5.1.18), Portable VirtualBox, Windows Virtual PC y Windows XP mode y PC-BASIC. El estadístico chi-cuadrado de independencia se basa, entre otras cosas, en el cálculo de las frecuencias esperadas, para cada casilla de la tabla de contingencia, asumiendo que la hipótesis nula de independencia es verdadera. La probabilidad de que una observación ocupe la casilla " ij " es el producto entre la probabilidad de pertenecer a la fila i y la probabilidad de pertenecer columna j . La probabilidad de pertenecer a la fila i es el cociente entre la suma de las frecuencias de todas las casillas de la fila i y el total de observaciones de la tabla. La probabilidad de pertenecer a la columna j es el cociente entre la suma de las frecuencias de todas las casillas de la columna j y el total de observaciones de la tabla. El estadístico ji-cuadrado se calcula sumando los cuadrados de las diferencias entre las frecuencias observadas y esperadas de cada casilla después de dividirlos por la correspondiente frecuencia esperada. El programa ocupa poco de memoria RAM y una vez compilado, se ejecutan más rápido (JICUADRA.EXE v. 1.0). Puede ser utilizado por un profesor, estudiante, profesional e investigador.

El valor cuantitativo de la variable y lo medible de lo cualitativo. Laura Elizabeth Cervantes Benavides¹ y Alma Lucero Sosa Blancas²

¹Universidad Autónoma de Ciudad Juárez, ²Universidad Nacional Autónoma de México

A partir de la sociedad del conocimiento, el sistema educativo moderno promueve la enseñanza a la especialización más acentuada; la consecuencia es la desvinculación de las áreas de conocimiento, específicamente en las áreas de metodología. Esta división ha provocado el riesgo en las áreas de investigación, porque la prontitud de enseñar métodos de investigación está provocando que los fenómenos biológicos se pretendan captar desde esta visión, reduciendo los fenómenos físicos y los fenómenos humanos a fenómenos exclusivamente animales o meramente manifestaciones físicas. La poca seriedad y el excesivo descuido, desde la educación, para diseñar los programas en estas áreas generadoras de conocimiento, a la hora de decidir quién debe impartir estas clases en cualquier centro de educación, no se elige al profesional más adecuado para "enseñar" métodos de investigación, esto puede indicarnos lo más grave y peligroso de esta situación en la formación de los futuros investigadores en las ciencias sociales. El aprendizaje adquirido de esta manera tiene como resultado el concebir los fenómenos como lo que no son, el pensarlos como lo que no pueden ser y abordar el estudio de estos "objetos" esperando que se comporten como no se pueden comportar, la simplificación se concreta cuando se empieza a hacer estudiar los fenómenos sociales.

Invarianza factorial de dos versiones lingüísticas de la "Escala de Orientación Emprendedora". Arantxa Gorostiaga Manterola¹, Nekane Balluerka Lasa¹, Jone Aliri Lazcano¹, Goretti Soroa Martínez¹, Alexander Muela Aparicio¹, Aitor Aritzeta Galan¹ y Imanol Ulacia Aramendi¹

¹Universidad del País Vasco UPV/EHU

La orientación emprendedora es la orientación psicológica de las personas que tienen tendencia a plantear ideas innovadoras y creativas para la solución de problemas y a ser proactivas y competitivas en los diversos aspectos de su vida, asumiendo los riesgos de sus propias decisiones y manteniendo una marcada orientación al logro y al aprendizaje. El presente estudio pretende analizar la invarianza factorial de dos versiones lingüísticas de un instrumento elaborado para evaluar la orientación emprendedora, la "Escala de Orientación Emprendedora". La muestra estuvo compuesta por 810 estudiantes pertenecientes a 23 centros de Formación Profesional de la Comunidad Autónoma del País Vasco (46,9% de mujeres y 53,1% de hombres). El 49,8% de los estudiantes cumplieron la versión en euskera del instrumento, mientras que el 50,2% cumplieron la versión en castellano. La edad de los/as estudiantes osciló entre los 16 y los 57 años ($M = 21,77$; $DT = 5,91$) y el 55,9% de la muestra tenía experiencia laboral previa. La "Escala de Orientación Emprendedora" es una escala que consta de 32 ítems que evalúan 6 dimensiones: orientación a la innovación, orientación al riesgo, orientación proactiva, orientación competitiva, orientación al logro y orientación al aprendizaje. El instrumento en su versión en castellano ha mostrado adecuadas propiedades psicométricas. Una vez que el instrumento ha sido adaptado al euskera, resulta de gran interés analizar la equivalencia factorial de ambas versiones lingüísticas. Con este propósito se utilizó el análisis factorial confirmatorio multigrupo. Los resultados mostraron que el modelo que planteaba la equivalencia de la configuración y de los pesos factoriales en ambas lenguas presentó un ajuste aceptable y no mostraba una pérdida significativa del ajuste con respecto al modelo de invarianza configural. Estos resultados se tomaron como evidencia de la invarianza métrica de ambas versiones lingüísticas del instrumento.

Análisis de la dimensionalidad del *Overall Personality Assessment Scale* (OPERAS) en euskera. Jone Aliri Lazcano¹, Alexander Rodríguez-López², Nekane Balluerka Lasa¹ y Arantxa Gorostiaga Manterola¹

¹Universidad del País Vasco UPV/EHU, ²Diputación Foral de Álava

A pesar de la robustez del modelo de los cinco grandes factores de personalidad y de que tales factores se consideran ortogonales, numerosos autores han constatado que los análisis factoriales confirmatorios ortogonales no presentan un ajuste adecuado de los datos a la teoría, por lo que han utilizado otras estrategias como la rotación Procrustes. El OPERAS (Vigil-Colet et al., 2013) es un instrumento basado en dicho modelo que consta de 40 ítems con cinco alternativas de respuesta que, además de evaluar los cinco grandes factores de personalidad, permite controlar sesgos de respuesta tales como la aquiescencia y la deseabilidad

social. El instrumento en su versión original en castellano presenta buenas propiedades psicométricas. A partir de la literatura revisada, el objetivo del presente trabajo consiste en analizar diversos modelos para clarificar la dimensionalidad de la versión adaptada al euskera del OPERAS. La muestra estuvo compuesta por 678 empleados (444 mujeres y 234 varones) pertenecientes a entidades públicas ubicadas en la Comunidad Autónoma del País Vasco con una edad que oscilaba entre los 21 y los 63 años ($M = 44,63$, $DT = 7,66$). Los/as participantes presentaban un nivel medio-alto de euskera, llevaban una media de 11,46 años en la organización y una media de 7,71 años en el puesto que ocupan actualmente. Se analizó el ajuste de diversos modelos de AFC y ESEM (Exploratory Structural Equation Models), ortogonales y oblicuos, con estimador WLSMV y MLR, mediante el software MPlus versión 7,4. Los resultados mostraron que el modelo ESEM ortogonal presenta un ajuste adecuado con cargas factoriales que se ajustan al modelo teórico. Con el presente trabajo se pretenden aportar evidencias sobre la dimensionalidad de la versión en euskera del OPERAS y contribuir a la literatura científica referida a la validez transcultural del modelo de los cinco grandes factores de personalidad.

The *Utrecht Work Engagement Scale for Students* (UWES-9S): Factorial validity, reliability, and measurement invariance in a Chilean sample of undergraduate university students. Marcos Carmona Halty¹

¹Universidad de Tarapacá

This study examines the within-network construct validity of the UWES-9S in a convenience sample of 1,502 Chilean students ranging between 18 and 25 years old. The results of confirmatory factor analysis supported a solution with three related factors that fit significantly better than a one-factor solution. The three subscales (i.e., vigor, dedication, and absorption) and the overall UWES-9S showed satisfactory internal consistency. The results of multiple-group confirmatory analysis supported gender invariance. Overall, the UWES-9S was found to be reliable and valid scale to assess academic engagement in Chilean undergraduate university students.

Estructura bifactorial del contenido de los estereotipos: evidencias de validez en adolescentes autóctonos que evalúan a inmigrantes. María Sánchez Castelló¹, Jorge L. Ordoñez Carrasco¹, Pablo Sayans Jiménez¹ y Antonio J. Rojas Tejada¹

¹Universidad de Almería

La investigación reciente sobre medición del contenido de los estereotipos muestra la existencia de un factor general de evaluación, junto a las dimensiones tradicionales usadas en el estudio de la percepción grupal. El objetivo de este estudio es poner a prueba un modelo bifactorial y proporcionar evidencias de validez que avalen dicha estructura en adolescentes que evalúan a inmigrantes. Participaron 1.120 adolescentes autóctonos ($M = 14,98$, $DT = 1,43$; 52% chicas) de tres submuestras que evaluaron a adolescentes de origen marroquí ($n = 488$), rumano ($n = 314$) y ecuatoriano ($n = 318$). La muestra procede del proyecto "Actitudes prejuiciosas, proceso de aculturación y adaptación de adoles-

centes de origen inmigrante y autóctono" financiado por el Ministerio de Economía, Industria y Competitividad (PSI2016- 80123-P). El cuestionario incluía, entre otras, tres subescalas para medir las dimensiones estereotípicas, así como una escala de calidad del contacto. Para comprobar el ajuste del modelo bifactorial se realizaron análisis factoriales confirmatorios en cada submuestra. Para la obtención de evidencias de validez se utilizó un modelo estructural en el que se incluyó la variable calidad de contacto. Los índices de ajuste relativo (CFI, RMSEA) muestran un ajuste muy bueno en los tres grupos. La correlación entre la calidad del contacto y el factor general (evaluación) es moderada/alta en los tres grupos. También es significativa la correlación con la dimensión de sociabilidad (moderada) en el caso de autóctonos que evaluaban a adolescentes de origen marroquí, y con moralidad (baja) y sociabilidad (moderada) en el caso de autóctonos que evalúan a los de origen ecuatoriano. Estos resultados, junto con un mayor ajuste teórico al proceso de evaluación intergrupala, muestran evidencias empíricas que apoyan el uso de este modelo. La importancia de la evaluación global (factor general) queda patente, siendo el que mejor explica la relación con la calidad del contacto en los tres grupos.

Estructura factorial del *Assessment of Children's Emotion Skills*. Ana Isabel Vergara Iraeta¹ y Natalia Alonso Alberca¹

¹Universidad del País Vasco UPV/EHU

El *Assessment of Children's Emotion Skills* es una herramienta dirigida a evaluar el conocimiento de las emociones en niños/as de entre 7 y 11 años. Esta prueba incluye las dimensiones de percepción emocional en rostros, percepción emocional en manifestaciones conductuales y la comprensión emocional, específicamente sobre las emociones básicas de alegría, tristeza, enfado y miedo. Consta de 56 estímulos donde se utilizan fotografías y descripciones de situaciones cotidianas, y se le pregunta al niño/a sobre las emociones vinculadas a las expresiones, conductas y situaciones presentadas. Existe una importante escasez de pruebas adaptadas a nuestro contexto dirigidas a evaluar el conocimiento emocional en estas edades, lo cual confiere a esta adaptación un gran interés. Tras la adaptación previa realizada por nuestro equipo de la prueba *Emotion Matching Task* para la evaluación del conocimiento emocional en niños/as de entre 3 y 6 años, resultaba necesario disponer de una prueba para evaluar dicho constructo en fases posteriores, y poder llevar a cabo estudios longitudinales. El objetivo de este trabajo es ofrecer evidencias acerca de las propiedades de la prueba ACES, analizando su estructura factorial a través del análisis factorial confirmatorio mediante Mplus, así como su capacidad predictiva de la conducta adaptativa-desadaptativa en la infancia. En este estudio han participado 250 niños/as entre 84 y 136 meses (54,8% niños). Los resultados obtenidos avalan las dimensiones planteadas por los autores de la prueba. Asimismo, se han encontrado relaciones entre el conocimiento emocional medido a través del ACES y dimensiones conductuales y emocionales como la agresividad o las habilidades sociales, avalando la validez de constructo. En este trabajo se presentan los análisis realizados, la necesidad de eliminar ítems con efecto techo o suelo, los diferentes modelos utilizados para analizar la estructura factorial y el modelo que finalmente se propone.

Propiedades psicométricas de la *Escala de Razones para Emigrar*: opiniones de jóvenes emigrantes y adultos no emigrantes. Raquel Gómez¹, Marcos Pascual-Soler², Héctor Monterde-i-Bort¹, Livia García³, Olaya Molina Palomero¹ y Dolores Frías Navarro¹

¹Universidad de Valencia, ²ESIC Business & Marketing School of Valencia, ³Universidad de La Laguna

En el presente trabajo se presenta el desarrollo y el análisis de las propiedades psicométricas de la "Escala de razones para emigrar". Los análisis se llevan a cabo en primer lugar con un análisis exploratorio con una muestra de jóvenes españoles emigrantes y posteriormente se describe la evolución de la escala RAZPE-2 con 56 ítems donde se amplían los ítems para formar un instrumento con al menos tres ítems por dimensión, utilizándose en este caso una muestra de adultos no emigrantes. La muestra total del estudio es de 1.436 participantes (435 jóvenes emigrantes españoles y 1001 adultos españoles no emigrantes). Los resultados señalan que los valores de alfa de Cronbach son aceptables (superiores a 0,70) en todas las subescalas. La escala RAZPE-2 está formada por 10 dimensiones: I: búsqueda de nuevas experiencias para la mejora personal, II: beneficios de la red social, III: huir del clima político, IV: búsqueda de formación académica, V: mejorar la capacitación profesional, VI: búsqueda de un mejor emplazamiento para vivir, VII: falta de trabajo en el país de origen, VIII: aumentar los ingresos económicos, IX: olvidar una experiencia de fracaso personal, X: influencia de los programas televisivos tipo "Españoles en el mundo". Los resultados de nuestro estudio aportan información sobre los motivos que se consideran más importantes para emigrar y se lleva a cabo un análisis donde comparan las respuestas de los jóvenes emigrantes y los adultos no emigrantes.

Utilidad de los ítems validados, y no incorporados, en las diferentes pruebas psicométricas. María Luisa Coca Hidalgo¹

¹Universidad de Huelva

De acuerdo con lo establecido en el art. 33 del Código Deontológico del Consejo General de Colegios Oficiales de Psicólogos, "Todo/a psicólogo/a, en el ejercicio de su profesión, procurará contribuir al progreso de la ciencia y de la profesión psicológica... comunicando su saber a estudiantes y otros profesionales según los usos científicos y/o a través de la docencia". En la práctica clínica, y con respecto a los diferentes desajustes y trastornos psicológicos, se hace necesario un banco de ítems, actividades, metáforas, así como de otros elementos contrastados, que son de utilidad para trabajar la psicoeducación, regulación emocional, reestructuración cognitiva, activación conductual, desensibilización, etc. Estos recursos van a ser interesantes, no solo para confirmar o refutar la sospecha clínica inicial, sino para reflexionar, analizar creencias, distorsiones cognitivas, etc., y van a estimular o permitir aflorar otros pensamientos, emociones y conductas más adaptativas. En este sentido, los ítems incluidos en las diferentes pruebas, válidamente contrastadas, no está permitido su uso, para evitar el aprendizaje de las respuestas, afectando, con ello, a las propiedades psicométricas del instrumento en la evaluación y diagnóstico. Así lo recoge el art. 19 del citado Código Deontológico. No obstante, durante la producción de versiones finales de pruebas psicométricas, el número global de ítems puede verse

reducido, por ser un instrumento muy extenso respecto al número de ítems, no mejorando ninguna de sus características psicométricas (ni empeorándola). Este estudio pretende dar valor a esos ítems que, no incorporados al instrumento final, ni en test adaptativos informatizados (TAIs), de ser recogidos en un banco de ítems para la práctica clínica, podrían aportar ahorro de tiempo y mayor precisión, mejorando los resultados con respecto a las diferentes demandas y necesidades de los usuarios, no siendo esperable que afecte al aprendizaje de los incluidos en las pruebas psicométricas.

Construcción de una escala de medida de evaluación de aplicaciones digitales (apps) para el desarrollo de la competencia artística. Feliciano Ordóñez Fernández^{1,2}, Martín Caeiro Rodríguez², Paula Fernández García³ y David Martín Espinosa³

¹Facultad Padre Oso, ²Universidad Internacional de La Rioja, ³Universidad de Oviedo

El estudio se ha desarrollado a partir de la necesidad detectada en la disciplina de arte en educación y constituye la primera fase del mismo. Se desarrolla la creación de una escala de evaluación de APPs teniendo en cuenta los contenidos formativos artísticos para las etapas formativas de Educación Primaria y Secundaria distribuidos entre dos dominios artísticos, y se realiza la evaluación de su fiabilidad, dimensionalidad y validez. En una primera fase se pidió a 4 expertos de diferentes disciplinas artísticas su contribución para construir un listado de posibles ítems de la escala. En relación a las APPs y a los procesos que desarrollan se identificaron diversos ítems relacionados con la competencia artística tanto en su formación perceptivo como expresiva. Con posterioridad, otros 3 expertos realizaron una nueva selección de los ítems. El conjunto de ítems preseleccionado fue administrado a una muestra de estudiantes con ánimo de evaluar la claridad de la redacción, la comprensión de los mismos y las dificultades en las respuestas. Se desecharon todos los ítems que mostraban problemas y otros fueron nuevamente redactados. Finalmente, otros dos expertos revisaron los ítems y se realizó un análisis de concordancia de atributos mediante el coeficiente kappa y el coeficiente de concordancia de Kendall. La escala ha sido administrada a una muestra accidental de alumnos de educación primaria y secundaria para evaluar provisionalmente su fiabilidad y dimensionalidad y la pertinencia de todos los ítems con ánimo de concluir con una escala de un número reducido de ítems.

Psychometric properties and Spanish adaptation of the *Readiness for InterProfessional Learning Scale (RIPLS)*. Tommy Norheim¹, Olga Greciano¹, Patricia Valencia¹, Eva Icaran¹, Juan José Beunza¹ y Gustavo González Cuevas¹

¹European University of Madrid

The Readiness for InterProfessional Learning Scale (RIPLS) is a widely used self-report instrument (Parsell y Bligh, 1999; 2005) to assess the readiness of students to engage interactively with students of other health professions in learning together. Our aims here were to adapt RIPLS to Spanish in a sample of health care undergraduate students as well as determine its psychometric properties. RIPLS was translated into Spanish using the translation and back-translation technique and administered to three hundred and twenty-seven health care undergraduate students at

European University of Madrid. Chronbach's alpha for the RIPLS showed favorable overall and specific reliabilities in terms of internal consistency (alpha > 0.8), except for the factor Roles & Responsibilities (alpha < 0.4). Moreover, confirmatory factor analysis indicated that the data fit a four-factor solution. All in all, the reliability and factorial validity properties obtained in this investigation support the conclusion that the Spanish adaptation of RIPLS may represent a useful tool for the measure of changes in interprofessional education. However, future research is needed to further evaluate and confirm the psychometric characteristics of this scale.

¿Eres un buen juez? Considerando la discriminación y severidad de expertos en la validación de contenido. Rodrigo Schames Kreitchmann¹, Pablo Nájera Álvarez¹, Susana Sanz Velasco¹ y Miguel A. Sorrel¹

¹Universidad Autónoma de Madrid

Introducción: dentro del contexto de construcciones de tests, la práctica más habitual para garantizar la validez de contenido se basa en la evaluación por parte de jueces expertos, que determinan la congruencia entre los ítems del cuestionario y el/los constructo/s que se pretende/n medir. A la hora de analizar las evaluaciones, tradicionalmente se han aplicado estadísticos que no tienen en cuenta que los jueces pueden diferir en su capacidad discriminativa y en su severidad. En el presente trabajo se propone la aplicación de un modelo paramétrico de las valoraciones de los jueces, en el que se permite que varíen en capacidad discriminativa y severidad. Se trata de un modelo de respuesta graduada (MRG) para valoraciones en escala ordinal, en el que se estiman los parámetros de escala y umbral para los jueces, y un parámetro de representatividad del constructo para cada ítem. Objetivo: comparar la calidad de los indicadores de representatividad del constructo ofrecidos por el MPVJ y por los indicadores tradicionales (i.e., V de Aiken e índice de validez de contenido a partir de escala dicotomizada). Procedimiento: se aplica una escala graduada a 10 jueces expertos en la que indica la representatividad de 140 ítems de personalidad con respecto al constructo que se pretende medir. Posteriormente, los ítems son aplicados a una muestra de estudiantes, y se estima un MRG. Se obtienen las correlaciones entre los parámetros de discriminación obtenidos con el MRG y los parámetros de representatividad obtenidos con el MPVJ y los indicadores tradicionales. Resultados esperados: al tener en cuenta la variabilidad de los jueces, se espera que el MPVJ modele mejor las valoraciones que los indicadores tradicionales, resultando en correlaciones mayores con los parámetros de discriminación estimados con el MRG.

¿Existen estimadores de asimetría y curtosis adecuados a distribuciones no normales? Roser Bono¹, Jaume Arnau¹, Rafael Alarcón² y María J. Blanca²

¹Universidad de Barcelona y ²Universidad de Málaga

En la mayoría de estudios, los índices de asimetría y curtosis se estiman en base al tercer y cuarto momento de una distribución. Sin embargo, sus valores suelen estar sesgados cuando las distribuciones son no normales, es decir, puede darse una desviación entre los valores teóricos y los empíricos de las medidas de asimetría y curtosis. A fin de reducir este sesgo, Hogg propuso esti-

madores alternativos que pueden ser una buena elección en una amplia variedad de distribuciones. El objetivo de este estudio de simulación fue comparar la precisión de los estimadores convencionales con los de Hogg, considerando las distribuciones continuas más frecuentes en ciencias sociales, de la educación y de la salud (gamma, lognormal y exponencial). Los resultados indicaron que los estimadores de curtosis de Hogg presentan menor sesgo que el estimador convencional, mientras que para la asimetría el estimador con menor sesgo es el convencional. Por último, cabe resaltar que los estimadores de curtosis propuestos por Hogg no se ven afectados por el tamaño de la muestra. Agradecimientos: este trabajo ha sido subvencionado por el Ministerio de Ciencia, Innovación y Universidades de España, proyecto PSI2016-78737-P (AEI/FEDER, UE).

Decisiones importantes que debe tomar el investigador para el análisis de los tiempos de respuesta: una aproximación empírica con la librería de R lmer. Raquel Colorado Cajas¹, Lorena Alicia Martín-Arnal¹, José A. León¹, Ricardo Olmos Albacete¹ y Beatriz Gil Gómez de Liaño^{1,2}

¹Universidad Autónoma de Madrid, ²Harvard University

El tiempo de respuesta ha sido una variable de estudio enormemente utilizada durante más de 100 años para comprender los procesos mentales. Inspirándonos en estudios recientes son muchos los aspectos metodológicos que quedan aún por resolver, sobre todo para que el investigador aplicado sepa cómo abordar de la mejor forma posible los análisis sobre este tipo de variables dependientes. Entre ellos destacan: (1) las distintas propuestas sobre las distribuciones de probabilidad que mejor modelan los tiempos de respuesta (ex gaussian, gamma, shifted Wald, etc.), (2) cómo transformar los datos no linealmente (p.ej., la inversa de la VD o el logaritmo neperiano) para aproximarlos a la distribución gaussiana o (3), los interrogantes y dudas que suscitan a los metodólogos estadísticos como la media y la desviación típica como buenos representantes para resumir las distribuciones tan asimétricas como el tiempo de respuesta. El estudio que se presenta analiza dos cuestiones sobre datos reales de movimientos oculares donde la VD es el tiempo de respuesta: (1) cómo afecta una transformación no lineal (en este caso la logarítmica) a las estimaciones y conclusiones de los efectos fijos analizados en el experimento y (2) qué ocurre en las estimaciones de los efectos fijos y de los errores típicos al analizar los datos bajo la óptica de los modelos mixtos con distribuciones no gaussianas (p. ej., ex gaussian y gamma) con las nuevas facilidades que dispone para ello la librería de R lmer. Los resultados muestran que las estimaciones quedan afectadas de forma distinta dependiendo de la decisión de análisis que se adopte. Por lo tanto, las implicaciones metodológicas son importantes para los investigadores puesto que las decisiones sobre los aspectos metodológicos aquí estudiados pueden afectar seriamente a las conclusiones del estudio.

Temas de interés de alumnos de la licenciatura en Administración de la Universidad Nacional Autónoma de México que participan en certámenes académicos de investigación en recursos humanos. Alma Lucero Sosa Blancas¹ y Miguel Alejandro Villavicencio Carranza¹

¹Universidad Nacional Autónoma de México

La Facultad de Contaduría y Administración de la Universidad Nacional Autónoma de México (UNAM), desde hace 14 años celebra anualmente el "Certamen académico de recursos humanos", el cual tiene como objetivo "promover en los alumnos el desarrollo del talento en las competencias de investigación e incrementar la capacidad de aplicación de conocimientos para la solución de problemáticas específicas que coadyuven a la efectividad organizacional particularmente en el área de recursos humanos". Para ello, los alumnos son convocados a realizar una investigación de tipo: (a) documental, (b) experimental o (c) cuasi-experimental. Con la finalidad de conocer cuáles son los temas en los cuales los alumnos han mostrado mayor interés al realizar investigación, se tomaron como referencia los certámenes académicos de las ediciones XII, XIII y XIV, los cuales tuvieron lugar en los años 2017, 2018 y 2019 respectivamente. En esta primera etapa se realizó un análisis de contenido, agrupando los trabajos por categorías de acuerdo con las asignaturas que pertenecen al área de recursos humanos y que se imparten en el plan de estudios de la licenciatura en Administración. El resultado indica que el 70% de los trabajos de investigación presentados corresponden a: reclutamiento y selección de personal (14%), tendencias actuales en administración de recursos humanos (10%), planeación e integración de los recursos humanos (9%), desarrollo y calidad de vida para los recursos humanos (9%), sistemas de información para recursos humanos (7%), seguridad e higiene en las organizaciones (7%), administración de la remuneración (6%) y otros temas (9%). El 30% se encuentra distribuido entre otros, en: capacitación de personal, coaching, comunicación y motivación, competencias laborales, administración de recursos humanos en el ámbito internacional y sociología de la organización. Estos datos permiten identificar interés que muestran los alumnos al realizar investigación para participar en el Certamen Académico.

Una aplicación del funcionamiento diferencial del ítem para analizar procesos de meta-comprensión en distintos tipos de texto. José Ángel Martínez Huertas¹, José David Moreno¹, Ricardo Olmos Albacete¹, Inmaculada Escudero¹ y José A. León¹

¹Universidad Autónoma de Madrid

En el campo de la metacognición, las autoevaluaciones sobre el desempeño en tareas de comprensión se conocen como meta-comprensión. Estos procesos tienen una gran repercusión en la comprensión lectora y son muy dependientes del tipo de texto que se lee. El objetivo de este trabajo fue estudiar las diferencias de comprensión en función de las habilidades metacognitivas en distintos tipos de texto en lectores con edades sensibles a los cambios en el desarrollo de estos procesos. 823 estudiantes de educación primaria, con un rango de edad comprendido entre los 9 y los 11 años, respondieron a la prueba de evaluación de comprensión lectora ECOMPLEC.Pri sobre textos narrativos, expositivos y discontinuos. Se llevó a cabo una clasificación de las habilidades de metacompreensión de los participantes en base al ajuste entre su rendimiento en cada texto y su autoevaluación sobre su desempeño en esa misma tarea de comprensión lectora. Así, se dividió a los participantes en dos grupos distintos según un criterio arbitrario: uno en el que se daba un buen ajuste entre el desempeño percibido y el desempeño observado, y otro en el que el ajuste era menor (baja metacompreensión). Se llevó a cabo un

análisis del funcionamiento diferencial del ítem (DIF) entre los grupos de alta y baja metacompreensión utilizando el método de regresión logística generalizada para comprobar qué ítems mostraban diferencias. Los resultados mostraron notables diferencias entre el grupo de alta y baja metacompreensión en el texto narrativo frente a los otros dos tipos de texto, especialmente en aquellos ítems relacionados con el modelo mental del lector. Estos resultados indican que el DIF puede ser una herramienta sensible para estudiar diferencias en procesos cognitivos de alto nivel, tales como la metacompreensión.

Desarrollo y validación de un cuestionario sobre el uso de la tecnología. María del Pilar Egea Romero¹, Ángela Socastro Gómez² y Ana Jiménez Perianes¹

¹Universidad San Pablo CEU (Madrid), ²Universidad Complutense de Madrid

Es un hecho que hoy en día hay un aumento del número de tecnologías en nuestro entorno así como de los trastornos que se asocian con el uso de las mismas. Este tipo de avances tiene muchas ventajas para nuestro desarrollo personal y profesional, pero si se desvirtúa su uso podemos enfrentarnos a un grave problema de salud pública. Los principales problemas derivados de este mal uso de la tecnología se relacionan con patologías derivadas del tecnoestrés como son la tecnoansiedad y la tecnoadicción. Estas patologías se consideran polos opuestos y dependiendo de la época se ha centrado más el estudio en una u otra. A pesar de la existencia de diferentes cuestionarios para la evaluación de cada una de estas patologías, no existe ningún instrumento que mida en términos similares ambas, ni que contenga ítems tanto de la esfera cognitiva como afectiva y conductual de la persona. Por lo tanto, el objetivo de este estudio es desarrollar y validar un instrumento de evaluación sobre la tecnoansiedad y la tecnoadicción con dichas características. Para ello, inicialmente se planteó un cuestionario de 59 ítems que se aplicó vía online a 215 personas de entre 14 y 65 años. Los resultados del análisis de fiabilidad señalaron valores de alfa de Cronbach superiores a 0,80 en ambas escalas. También se realizaron análisis sobre la validez de constructo y convergente. Por último, los resultados finales indicaron que el cuestionario se componía de 24 ítems para cada una de las escalas y la estructura factorial señaló 6 factores para cada escala, relacionados con las principales manifestaciones de las patologías.

Metodología adaptativa para la evaluación docente en enseñanza universitaria. Ignacio Carou¹, Alicia Riso¹ y Manuel Peralbo¹

¹Universidad de A Coruña

La intención de una institución educativa de analizar y mejorar la práctica docente se manifiesta en la implantación de sistemas de evaluación. Mientras que se han dado pasos importantes en el diseño y sistematización de herramientas y fuentes de datos para realizar estas evaluaciones, no se han desarrollado cambios metodológicos considerables en el análisis de la información contenida en los datos acumulados. Este hecho tiene especial relevancia cuando se trata de la evaluación docente en la Educación Superior. Los datos procedentes de las evaluaciones y auto-

evaluaciones docentes, y de las propias administraciones, pueden alcanzar un volumen considerable y difícilmente gestionable por métodos convencionales, por lo que suelen someterse a análisis meramente descriptivos. Un área que ha mostrado un gran potencial en la búsqueda de información útil y novedosa es la minería de datos. En este ámbito, se han propuesto diferentes metodologías siendo la más ampliamente utilizada CRISP-DM. Esta metodología tiene una estructura de seis fases con implicaciones bidireccionales entre algunas de ellas, lo que implica que la realización de ciertas fases conlleva la revisión de otras anteriores. Las fases que la componen son: análisis del problema, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación. En este trabajo se propone la adaptación de la metodología CRISP-DM, planteada principalmente para entornos empresariales y bienes tangibles, al proceso de evaluación de la calidad docente, en la que los datos hacen referencia a entornos educativos y productos intangibles. Las ventajas que presenta esta adaptación son varias: es una metodología probada, sistematizada y con un ciclo de vida evolutivo que asegura el análisis continuo y adaptativo; puede revelar información útil e inaccesible por otros medios, que permita optimizar la toma de decisiones para mejorar la práctica docente; y asegura el cambio de un paradigma descriptivo a uno predictivo.

Gamificación en la enseñanza de “Métodos y diseños de investigación”. María del Pilar Egea Romero¹

¹Universidad San Pablo CEU (Madrid)

En los últimos años se está haciendo especial hincapié en la innovación docente en las universidades, integrando metodologías activas y participativas en las que los estudiantes tengan una postura más dinámica en relación con el proceso de aprendizaje. La gamificación, que supone la adaptación del juego en el ámbito educativo, es una de las estrategias que se están utilizando para motivar al alumnado y mejorar su aprendizaje y grado de satisfacción con el proceso. Se ha realizado una investigación para estudiar el efecto que, sobre el rendimiento académico, tiene la gamificación en el proceso de enseñanza de la asignatura “Métodos y diseños de investigación” en alumnos de primer curso del grado en psicología. Se ha empleado un diseño cuasiexperimental de cohortes de retardo temporal (Baltes et al., 1988) con dos cohortes: una compuesta por alumnos de dos cursos en los que no se usó la gamificación ($n = 115$) y la segunda ($n = 110$) compuesta por alumnos de otros dos cursos en los que sí se empleó esta herramienta, consistente en la realización de crucigramas, puzzles y jeroglíficos, entre otros juegos. Igualmente se recogió el grado de satisfacción que la segunda cohorte tenía con respecto al uso de la gamificación en la asignatura.

Relación entre factores familiares y adicción a internet en adolescentes europeos: una revisión meta-analítica. Anna Círaso Calí¹, Carla Quesada-Pallarés¹, María Concepción San Luis² y Javier Ibias²

¹Universitat Autònoma de Barcelona, ²Universidad Nacional de Educación a Distancia

La adicción a internet, por su creciente prevalencia y por las implicaciones que tiene en términos de dificultades sociales, psicoló-

gicas y académicas, se ha convertido en un tema emergente a nivel social y de salud mental, sobre todo entre adolescentes. A pesar de los progresos de la investigación en este ámbito, la posibilidad de establecer las causas o factores de riesgo todavía es controvertida. En la literatura reciente, los factores familiares están bastante explorados en sus relaciones con la adicción a internet; sin embargo, dadas las diferentes facetas de los aspectos familiares que podrían constituir factores de riesgo o protectores, emerge la necesidad de llevar a cabo una revisión sistemática de las investigaciones empíricas realizadas hasta la fecha; para esclarecer la covariación entre estas variables, e intentar explicar la eventual heterogeneidad entre resultados. Además, los esfuerzos de investigación empírica se han centrado en los últimos años en estudiar muestras extraídas de poblaciones asiáticas, por lo que es especialmente relevante focalizarse en las pocas investigaciones llevadas a cabo con poblaciones europeas. Con este propósito, el estudio que se presenta analiza las relaciones entre la adicción a internet, con diferentes definiciones operativas, y factores familiares como los estilos parentales, el funcionamiento familiar o la comunicación, entre la población adolescente europea. Utilizando PsycINFO y Google Scholar, se recopilaron 104 referencias; una vez aplicados los criterios de inclusión, entraron en el análisis 9 estudios independientes, con método de encuesta, realizados entre 2010 y 2019; cubriendo una muestra total de 7.920 adolescentes y 28 tamaños de efecto. El análisis, que todavía está en curso, facilitará la comprensión de las relaciones entre diferentes factores familiares y la adicción a internet, posiblemente mediadas por variables como el género, la edad o los diferentes instrumentos de medida.

DÉPORVIDA. Programa de autoconocimiento en jóvenes futbolistas de un club profesional en España. Constantino Arce Fernández¹ y David Tomé Lourido²

¹Universidad de Santiago de Compostela, ²Universidad de Vigo

El presente trabajo refleja los resultados de un programa de intervención realizado en jóvenes jugadores de un club de fútbol profesional. El objetivo del programa fue mejorar, a través del fomento del autoconocimiento, los niveles de bienestar percibido, motivación intrínseca y satisfacción con el rendimiento. Para ello, se siguió una metodología cuasi-experimental pretest-postest con grupo control. Los jugadores que desearon participar en el programa formaron parte del grupo experimental, mientras que los que decidieron no hacerlo, formaron parte del grupo control ($n = 18$). Por tanto, la variable independiente (programa de autoconocimiento DÉPORVIDA) presentó dos niveles: participación en el programa o no participación en el programa. Las variables dependientes (bienestar percibido, autorregulación, necesidades psicológicas básicas y satisfacción con el rendimiento) se evaluaron a través de test estandarizados y adaptados al español. Las distintas actividades del programa se relacionaron con el modelo VIA de fortalezas del carácter (Peterson y Seligman, 2004) y se llevaron a cabo durante siete semanas consecutivas. Todas ellas estuvieron guiadas por personal del club con formación superior en educación, pedagogía y psicología. El trato con los participantes siguió las recomendaciones de la Declaración de Helsinki (1964). Los resultados del pretest indicaron que inicialmente no existían diferencias significativas en las variables dependientes entre los dos grupos de participantes. En cuanto a la evaluación postest, los

participantes del grupo experimental presentaron puntuaciones superiores en satisfacción con el rendimiento de la temporada ($t(23,23) = 2,42$; $p = 0,024$) y mayor percepción de locus de control relacionado con la autonomía en el deporte ($t(24,16) = 2,83$; $p = 0,009$). Por otra parte, al finalizar el programa, los participantes del grupo control indicaron experimentar un mayor porcentaje de tiempo infeliz que los del grupo experimental ($t(26) = 2,08$; $p = 0,048$). En definitiva, el programa DÉPORVIDA mejoró los niveles de satisfacción con el rendimiento y de locus de control en el deporte.

La distribución *Shifted Wald* para el análisis de los tiempos de respuesta. Una aproximación empírica con R. Lorena Alicia Martín Arnal¹, José A. León¹ y Ricardo Olmos Albacete¹

¹Universidad Autónoma de Madrid

Recientemente han surgido publicaciones novedosas sobre la mejor forma de abordar los análisis estadísticos cuando se trabaja con tiempos de respuesta. Una distribución de probabilidad que goza de popularidad es la *Shifted Wald* cuya metodología se explica en el estudio de Anders, Xavier y van Maanen (2016). El objetivo de este trabajo es presentar un ejemplo de uso de la metodología SW en un diseño experimental y comparar los resultados en relación al modelo clásico de ANOVA que compara medias poblacionales. Mientras que en distribuciones tan asimétricas como las que se derivan de los tiempos de respuesta, la media y la desviación típica son parámetros poco informativos (no resumen correctamente la distribución de la variable y están sobreestimados), la distribución SW caracteriza los datos atendiendo a tres parámetros: (1) θ o parámetro de comienzo (onset); (2) α o parámetro de tendencia central y de desviación en torno a la moda y (3) γ o parámetro que refleja el área de la cola derecha de la distribución. El diseño experimental que se analizó incluía una variable dependiente de tipo tiempo de respuesta a unos cómics ($n = 20$) dependiendo a dos tipos de viñetas (*peak* vs. *release*) y de la edad del sujeto (niños y adultos) ($n = 63$ participantes). Los resultados con la metodología SW fueron satisfactorios en cuanto al ajuste de los datos a la distribución SW. A su vez, se muestra que las conclusiones sobre los efectos fijos utilizando un contraste sobre medias (clásico) frente al contraste sobre los parámetros SW difieren. Las implicaciones de este estudio son importantes dentro de la psicología experimental. La metodología SW puede informar de aspectos más relevantes que pasan desapercibidos cuando se comparan medias.

Una aproximación a los modelos cuánticos de similitud. Alejandro Martínez Mingo¹, Guillermo Jorge Botana² y Ricardo Olmos Albacete¹

¹Universidad Autónoma de Madrid, ²Universidad Nacional de Educación a Distancia

En este poster se realizará un acercamiento a la aplicación de la teoría cuántica para comprender juicios humanos basados en la similitud entre palabras. Se llevará a cabo una breve descripción del modelo cuántico de similitud propuesto por Pothos, Bussey y Trueblood en 2013, y se ofrecerán resultados de un estudio preliminar sobre juicios de similitud de palabras proyectadas sobre un espacio semántico latente asumiendo los supuestos del modelo cuántico.

Nuevas propuestas para detectar automáticamente conceptos en respuesta abierta: estudio comparativo de dos métodos de evaluación automática (inbuilt rubric vs. Partial contents similarity). José Ángel Martínez Huertas¹, Ricardo Olmos Albacete¹ y José A. León¹

¹Universidad Autónoma de Madrid

Tradicionalmente, en evaluación automática de textos se han empleado métodos provenientes de modelos computacionales que arrojan una puntuación global de la calidad de la respuesta del estudiante. Sin embargo, esto no posibilita dar un feedback concreto a los estudiantes, sobre todo cuando es necesario concretar qué contenidos importantes han conseguido cubrir bien en su respuesta y cuáles necesitan claramente más atención y desarrollo. Por tanto, un reto ha sido crear métodos automáticos que detecten sutilmente contenidos concretos en la respuesta construida emitida por los estudiantes. En este estudio se compararon dos métodos que tienen el propósito de detectar contenidos concretos en respuesta construida: *partial contents similarity* e *inbuilt rubric*. El primer método representa un procedimiento clásico y utilizado en tutores virtuales (p. ej., *summary street*) y el segundo un método innovador que ha funcionado con éxito en aplicaciones más recientes. Método: en este estudio participaron 257 estudiantes y se utilizaron tres textos instruccionales diseñados para la tarea. Cada estudiante resumió un texto que a su vez fue evaluado por varios jueces expertos atendiendo a una rúbrica previamente discutida. La idea fue utilizar las puntuaciones de los jueces en cada contenido fundamental que debía contener la respuesta del estudiante como criterio externo con el que poder comparar ambos métodos automáticos. El modelo computacional utilizado fue *latent semantic analysis*. Resultados: los resultados mostraron que el método *partial contents similarity* tenía mucha más colinealidad y menos poder discriminante que el método *inbuilt rubric* en todos los textos. Este último fue mucho más sensible a los contenidos concretos establecidos en la rúbrica. Conclusión: el método moderno *inbuilt rubric* se postula como un método interesante para detectar contenidos sutiles y formar parte de tutores virtuales con el que poder medir y evaluar automáticamente con más calidad la respuesta construida.

Índices del grado de experiencia sensorial para 5.500 palabras en español: datos normativos y aplicaciones para la investigación. Antonio Manuel Díez Álamo¹, Emiliano Díez¹, Dominika Z. Wojcik¹, María Á. Alonso² y Ángel Fernández¹

¹Universidad de Salamanca, ²Universidad de La Laguna

El índice de experiencia sensorial es una variable léxica de carácter subjetivo recientemente desarrollada que refleja la medida en que una palabra evoca una experiencia sensorial o perceptiva, constituyendo una medida integradora de la activación sensorial que aglutina todas las modalidades sensoriales (Juhász y Yap, 2013; Juhász, Yap, Dicke, Taylor, y Gullick, 2011). En el estudio normativo que presentamos (en prensa) se recopilaron índices SER para 5.500 palabras en español, conformando el mayor conjunto de normas SER en español existente hasta la fecha. Los análisis de fiabilidad y validez demostraron que este estudio proporciona estimaciones adecuadas de los SER. Además, con el objetivo de analizar la naturaleza e implicaciones de esta nueva variable se realizaron análisis correlacionales y de regresión, que permitieron explorar sus relaciones con otras variables. Los resul-

tados mostraron que los SER correlacionaron significativamente con diversas variables psicolingüísticas y predijeron una cantidad significativa de varianza en los tiempos de decisión léxica de las palabras cuando se controlaron otras variables. La disponibilidad de índices SER, desde un punto de vista metodológico, será de utilidad para el diseño de estudios sobre representación conceptual, memoria semántica, procesamiento del lenguaje afectivo, sinestesia o bilingüismo, algunos de ellos ya en marcha en nuestro laboratorio. Asimismo, los materiales y la información sobre la relación entre el SER y otras variables pueden ser útiles para el desarrollo de nuevas líneas de investigación en psicología y neurociencia, y especialmente relevantes para las propuestas teóricas recientes, que plantean que la representación del conocimiento conceptual en el cerebro implica la existencia de un centro de actividad o hub semántico responsable de la integración de todas las fuentes de información específicas de modalidad (Paterson, Nestor, y Rogers, 2007).

Cuestionario de personalidad múltiple en QB64: PRF.BAS. Pablo García Medina¹ y Ernesto Juan Darías Morales¹

¹Universidad de La Laguna

Se expone PRF.BAS v. 1.0, un cuestionario informatizado que apresa cuatro factores. Los contenidos del cuestionario PRF (prueba de respuesta a la frustración) evalúan las tendencias a dar respuestas de: (a) agresividad, (b) autopenalización, (c) alternativa solucionadora y (d) frustración. Este último factor aparece como desagrado en el formato escrito. Presentamos a partir del original de papel una versión en programa de ordenador. La idea es facilitar que el cuestionario PRF de García Medina pueda contestarse de modo informatizado. Las respuestas son de modalidad de elección múltiple en 30 triadas de cada una de las cuales el evaluado elegirá sólo una. Debiendo contestar a una escala de desagrado en orden a la intensidad que el mismo evaluado dará (de 0 a 4). El programa está escrito y compilado en QB64, corriendo también para otras versiones de BASIC (todos ellos de dominio público en Internet y pueden utilizarse en Google (GW-BASIC, a 3.23, BASIC v 1.2, VBDOS v 1.00, QBASIC v 1.1, QuickBasic 4.5 y QBX v 7.1) bajándolos gratuitamente. Se ejecuta MS-DOS (v. 5.0 a 6.11), MS-DOS y diferentes Windows, Linux, Android, Mac OSX y en cualquier sistema operativo con VirtualBox, Virtual PC o el emulador DOSBOX. Esta versión de cuestionario con 30 ítems y una opción escalar ocupa poca memoria RAM. Y denominado PRF.EXE, una vez compilado, se mueve muy ágilmente. Tenemos previsto hacer este programa en Java, C, C++, y/o Visual Basic Net. Se pasará por Europa para saber si tiene la misma estructura factorial.

Medición de las DEAL (dificultades específicas del aprendizaje en la lectura): desde la infancia a la edad adulta. Ana Sánchez Dorado¹, David Cabello González¹ y José Luis Ramos Sánchez²

¹GiuntiEOS PSYCHOMETRICS, ²Universidad de Extremadura

En su última edición, el manual de los trastornos psicológicos publicado por la American Psychology Association enumera una serie de trastornos no recogidos en sus anteriores ediciones proponiendo además una serie de nuevas agrupaciones para ciertos grupos de trastornos que anteriormente tenían un concepto propio. Entre ellos está el trastorno específico del aprendizaje, carac-

terizado por: (A) dificultad en el aprendizaje y en la utilización de las aptitudes académicas, evidenciado por la presencia al menos uno de los siguientes síntomas que han persistido por menos durante 6 meses, a pesar de intervenciones dirigidas a estas dificultades: (a) lectura de palabras imprecisa o lenta y con esfuerzo, (b) dificultad para comprender el significado de lo que lee, (c) dificultades ortográficas, (d) dificultades con la expresión escrita, (e) dificultades para dominar el sentido numérico, los datos numéricos o el cálculo, (f) dificultades con el razonamiento matemático. (B) Las aptitudes académicas afectadas están sustancialmente y en grado cuantificable por debajo de lo esperado de la edad cronológica del individuo, e interfieren significativamente con el rendimiento académico o laboral... (C) Las dificultades de aprendizaje comienzan en la edad escolar pero pueden no manifestarse totalmente hasta que las demandas de las aptitudes académicas afectadas superan las capacidades limitadas del individuo. (D) las dificultades de aprendizaje no se explican mejor por discapacidades intelectuales... Es por ello, que debe realizarse una evaluación completa de estos trastornos. Uno de ellos, es el trastorno de dificultades en la lectura precisión en la lectura de palabras, velocidad o fluidez de la lectura o comprensión de la lectura 315.00 (F81.0). Para ello se deben utilizar herramientas desde la infancia a la adultez, es por ello que GiuntiEOS Psychometrics y Giunti Psychometrics poseen herramientas, punteras en el mercado.

Perfil metabólico en personas mayores: relación con las trayectorias sintomáticas depresivas. Yolanda Sánchez¹, Alejandro de la Torre Luque¹, Javier de la Fuente¹, Pilar López García¹ y José Luis Ayuso Mateos¹

¹Centro de Investigación Biomédica en Salud Mental (CIBERSAM), Universidad Autónoma de Madrid

Se ha documentado ampliamente la relación entre desregulación metabólica y la depresión mayor, especialmente en la edad avanzada. Sin embargo, no existe tanta evidencia que relacione desarrollos sintomáticos con el paso del tiempo y riesgo metabólico. Este estudio pretendía analizar la relación entre curso de sintomatología depresiva y perfil metabólico en adultos mayores. Se recogió datos sobre diagnóstico médico y psiquiátrico en una muestra de 1.536 participantes ingleses mayores de 50 años (56,58% mujeres; M = 61,73 años, DT = 7,64). Los participantes mostraban tres cursos de síntomas depresivos: no síntomas, síntomas subclínicos, curso de síntomas crónicos. Asimismo, se les tomó muestras de sangre para explorar el nivel de indicadores plasmáticos. Se usaron modelos multi-causa multi-indicador para estudiar el perfil metabólico de estas personas en función de cinco indicadores diagnósticos y plasmáticos de riesgo metabólico, controlando por covariables sociodemográficas y clínicas. Se estudió la influencia del curso sintomatológico mediante invarianza de medida multigrupo. Como resultados, el modelo metabólico ajustó bien a los datos ($\chi^2 = 70,73$, gl = 51; RMSEA = 0,023, CFI = 0,96), pero el que mostraba restricciones multigrupo estrictas tuvo mejor ajuste (deltaCFI = 0,003, deltaRMSEA = -0,002). Los resultados apuntan a que los individuos con trayectoria crónica de síntomas, pero también aquellos con trayectoria subclínica, mostraban mayor riesgo metabólico que los de trayectoria sin síntomas. Este estudio pone de manifiesto la relevancia de cursos crónicos de alta (o moderada) sintomatología depresiva y su potencial influencia en el desarrollo de enfermedades metabólicas posteriormente en la vida.

Trayectorias heterogéneas de síntomas de ansiedad en la adolescencia: un estudio basado en modelos de crecimiento mixto latente. Alejandro de la Torre Luque¹, Aina Fiol Veny², María Balle² y Xavier Bornas²

¹Centro de Investigación Biomédica en Salud Mental (CIBERSAM), Universidad Autónoma de Madrid, ²Universidad de las Islas Baleares

La adolescencia constituye una etapa vital de profundo cambio hacia el desarrollo del individuo adulto. En el plano de la salud mental, determinados factores biológicos (sexo, edad) y ambientales (p. ej., acontecimientos traumáticos) pueden repercutir en el curso natural psicopatológico. La adolescencia temprana constituye un periodo de especial vulnerabilidad para el desarrollo de problemas de ansiedad. Los modelos mixtos de clase latente (MMCL) pueden contribuir en la identificación de trayectorias heterogéneas de desarrollo sintomatológico, dada su flexibilidad al no partir de la existencia de un curso unitario. Este estudio pretendía estudiar la existencia de trayectorias diferenciales de síntomas de ansiedad en una muestra de 825 adolescentes (55,60% chicas; M = 13,01 años al inicio del estudio, DT = 0,56) que fue seguida durante 18 meses y preguntada por síntomas de ansiedad cada seis meses. Se monitorizó sintomatología ansiosa según los subtipos de trastornos. Como resultados, se identificaron entre dos y cuatro trayectorias diferenciales de síntomas de ansiedad en fobia social, ansiedad generalizada y ansiedad pánico. En todos los casos, se delimitó una trayectoria mayoritaria con baja sintomatología y al menos una clase de riesgo (clase con crecimiento elevado de síntomas o con síntomas crónicos). Ser chica y vivir con hermanos se relacionaron con trayectorias de riesgo, especialmente en el caso de la fobia social. Como conclusión, este estudio refleja la utilidad del uso de metodología centrada en la persona (MMCL) en la determinación de perfiles heterogéneos de desarrollo y en la búsqueda de alternativas personalizadas de abordaje preventivo y terapéutico.

SESIÓN PREMIO AEMCCO 2019*

*En los trabajos incluidos en esta sesión se han omitido los coautores; en cada trabajo aparece solo la persona que opta al premio.

Validación y desarrollo de una escala de liderazgo de servicio: análisis de las evidencias de validez. David Santos¹

¹Universidad Autónoma de Madrid

El liderazgo de servicio se refiere a la inclinación natural por servir a otros como una máxima prioridad. En esta investigación presentamos los resultados de 4 estudios cuyo objetivo fue desarrollar y validar una versión abreviada de la escala "Perfil de liderazgo de servicio". En el estudio 1, diversos expertos evaluaron y redujeron la escala original a 14 ítems, denominando a esta última versión "Escala de liderazgo de servicio - versión abreviada" (ELSVA). Se estudió también la estructura de la escala mediante un análisis factorial exploratorio en una muestra de estudiantes ($n = 1.001$), obteniendo dos soluciones posibles: un modelo unifactorial o un modelo de dos factores correlacionados. En el estudio 2 se utilizó una segunda muestra ($n = 456$) para contrastar ambas soluciones a través de un análisis factorial confirmatorio, por el que se propuso un modelo unifactorial (χ^2 /gl = 3.4, CFI = 0,91, TLI = 0,90,

RMSEA = 0,07, SRMR = 0,06). En el estudio 3, se puso a prueba la validez predictiva de la escala en una tercera muestra ($n = 202$) utilizando la autoeficacia como criterio. Finalmente, en el estudio 4 se puso a prueba la validez discriminante del instrumento en una nueva muestra ($n = 110$), mostrando que el estilo de liderazgo de servicio predice las actitudes hacia el emprendimiento social más allá que cualquier otro tipo de liderazgo (i.e., transformacional, transaccional o laissez-faire). Los análisis de fiabilidad de los cuatro estudios demostraron que la versión española de 14 ítems tiene una consistencia interna elevada en las cuatro muestras. En conjunto, los resultados de estos cuatro estudios avalan el uso de la ELSVA en futuras investigaciones.

Procedimientos de detección del falseamiento en test: alternativas al Z-test de Guo y Drasgow. Susana Sanz Velasco¹

¹Universidad Autónoma de Madrid

Introducción: las pruebas de selección en línea sin vigilancia son susceptibles a intentos de falseamiento para obtener una puntuación superior. Para detectarlo, se han propuesto procedimientos que comparan la puntuación UIT con la obtenida en una versión breve del mismo test (VT). Objetivo: detectar de forma precisa a las personas que falsean en un test adaptativo informatizado de rendimiento óptimo. Se compara uno de los procedimientos más utilizados, el de Guo y Drasgow (2010), con el de Finkelman et al. (2010) y se utilizan EAP, MAP y WLE como procedimientos de estimación, pues la selección de ítems puede variar. Método: se realizaron tres simulaciones (una por cada procedimiento de estimación) con 25.000 simulados honestos y 300.000 falseadores en cada una, con valores en el rasgo de -3 a 3 a intervalos de 0,25. El UIT contaba con 30 ítems; el VT, con diez. Los simulados honestos contestaban en ambas fases con su nivel real de rasgo; los falseadores, con su nivel real en el VT y con uno mayor en el UIT. Después, se realizaban los contrastes de hipótesis correspondientes, y se obtenían las tasas de error Tipo I y la potencia. Posteriormente, también se tenía en cuenta la tasa de error Tipo I para la comparación de los resultados de potencia, usando un procedimiento basado en la igualdad con el valor nominal real y otro de bootstrapping. Resultados: las tasas de error Tipo I estaban más cerca del valor nominal cuando se utilizaba el estadístico de Guo y Drasgow en combinación con los procedimientos bayesianos. La potencia es mejor con el estadístico de Guo y Drasgow-MAP, pero cuando se utiliza el procedimiento de igualdad del nominal, Finkelman-WLE se posiciona como mejor procedimiento, mientras que con bootstrapping sigue siendo el procedimiento de Guo y Drasgow-MAP.

Estimación EAP completamente bayesiana para contextos de falseamiento. Rodrigo Schames Kreitchmann¹

¹Universidad Autónoma de Madrid

Introducción: dadas las importantes consecuencias de la medición en contextos de selección de personal, es común que los evaluados falseen sus respuestas intentando maximizar sus posibilidades de ser escogidos. En estos contextos, se evidencia en la literatura una mejora significativa del ajuste de modelos de TRI al incluir un rasgo latente de "candidato ideal" o de "falseamiento" aparte de los rasgos sustantivos. Estos hallazgos plantean un desafío a la medición, dado que sugieren que el proceso de res-

puesta en las situaciones de selección difiere del proceso llevado a cabo en los estudios de validación de las escalas, donde no se espera que los evaluados falseen sus respuestas. Por otra parte, el proceso de respuesta podrá variar de una situación de selección a otra, dado que la imagen de "candidato ideal" dependerá del puesto de trabajo al que se aspira. Objetivos: considerando la inviabilidad de realizar estudios de validación para cada proceso de selección, se investiga la aplicación de un método de estimación *expected-a-posteriori* completamente bayesiano teniendo en cuenta la incertidumbre sobre los parámetros estructurales. Método: a través de simulación, se han generado respuestas de 1.000 sujetos a 30 ítems siguiendo: (1) un modelo unidimensional de respuesta graduada (proceso de respuesta honesto) y (2) un modelo bidimensional compensatorio de respuesta graduada (proceso de respuesta con falseamiento) con mayor tendencia a elegir categorías más altas. Se investiga la estimación de parámetros de los sujetos a partir de los datos con falseamiento utilizando: (1) estimación EAP tradicional empleando parámetros estructurales calibrados en condición honesta – emulando lo que ocurre al puntuar sujetos usando los parámetros estimados en un estudio de validación– y (2) estimación EAP completamente bayesiana. Resultados esperados: se espera que el método EAP completamente bayesiano ofrezca mejores estimadores que el método EAP tradicional basado en parámetros de la calibración honesta.

Does modeling wording effects help recover uncontaminated person scores? A systematic evaluation with random intercept item factor analysis. María Dolores Nieto¹

¹Universidad Autónoma de Madrid

A concern regarding the use of self-report measures is the potential influence of wording effects on examinees responses. The item wording (or keying) effect consists of logically inconsistent answers to positively and negatively worded items that tap into similar (but polarly opposite) content. Previous research has amply demonstrated that ignoring wording effect variance may have detrimental effects on the psychometric properties of questionnaires. The random intercept item factor analysis (RIIFA) model has proved useful in controlling for wording effects satisfactorily as evidenced by the improvements in model fit in comparison to models that only contain substantive factors. However, little is known regarding the capability of the RIIFA in recovering the uncontaminated person scores, which in the end are the focus of research in the social and behavioral sciences. To address this issue, the current study analyzed for the first time the performance of the RIIFA approach across three types of wording effects proposed in the literature: Carelessness, item verification difficulty, and acquiescence. In the context of unidimensional substantive models, four independent variables were manipulated using Monte Carlo methods: type of wording effect, amount of wording effect, sample size, and test length. The results corroborated previous findings by showing that the models that included the RIIFA wording factor were consistently able to account for the variance in the data, attaining almost perfect fit regardless of the type and amount of wording effect. Conversely, the models without the RIIFA factor produced increasingly poorer fit with greater amounts of wording effect. Surprisingly, however, the RIIFA model did not enable to better estimate the uncontaminated person scores for any type of wording effect in comparison to the substantive unidimensional

model. These apparently paradoxical findings are explained in light of the properties of the factor models examined, and practical guidelines are offered for applied researchers.

Mejoras en la validación empírica de la matriz-Q. Pablo Nájera Álvarez¹

¹Universidad Autónoma de Madrid

Los modelos de diagnóstico cognitivo son modelos estadísticos multidimensionales de rasgos latentes categóricos, llamados habitualmente atributos. Estos modelos requieren una matriz-Q, que determina la relación entre los ítems y los atributos. Normalmente, la especificación de la matriz-Q se hace por parte de expertos en la materia de conocimiento. Éste es un proceso subjetivo que puede llevar a errores de especificación. Estos errores pueden afectar negativamente a la estimación de los atributos dominados por los evaluados. Por estas razones, en los últimos años se han desarrollado algunos métodos de validación empírica de la matriz-Q. El método basado en el índice de discriminación general (GDI, por sus siglas en inglés general discrimination index) posee algunas ventajas destacables, como su flexibilidad para ser aplicado a distintos MDCs. Sin embargo, este método asume que la matriz-Q original está correctamente especificada, lo cual no será cierto en la mayoría de los contextos aplicados. Para solucionar este inconveniente, el presente estudio investiga la aplicación iterativa del método GDI donde solamente un ítem es modificado en cada paso, y el punto de corte requerido por el método se actualiza en función de las nuevas estimaciones de los parámetros. El rendimiento del nuevo procedimiento es evaluado mediante un estudio de simulación. Los resultados muestran que el rendimiento del método GDI mejora sustancialmente al aplicar el procedimiento iterativo a nivel de ítem y un punto de corte apropiado. Las mejoras con respecto al método GDI original son mayores a medida que el porcentaje de errores de especificación en la matriz-Q original es mayor. El procedimiento se ilustra con datos reales de un test de resta de fracciones.

Prediciendo la emocionalidad de las palabras a partir de modelos computacionales del lenguaje y machine learning. José Ángel Martínez Huertas¹

¹Universidad Autónoma de Madrid

La producción científica que durante los últimos años ha incorporado técnicas de aprendizaje supervisado ha crecido exponencialmente y está suponiendo una revolución para la ciencia cognitiva. De este modo, el progreso del modelado computacional está generando cambios significativos en el estudio de los procesos cognitivos y, especialmente, en el campo de la psicolingüística. El objetivo de este estudio es aunar dos campos: la lingüística computacional y las técnicas de machine learning dentro de una aplicación concreta: la computación afectiva. Se estudió la predicción de respuestas emocionales a partir de un modelo computacional del lenguaje. En una muestra de 11,357 palabras se entrenaron distintos modelos de redes neuronales para predecir la emocionalidad de dichas palabras a partir de sus representaciones vectoriales en un espacio semántico. Así, se comparó el rendimiento de modelos de redes neuronales univariados y multivariados en una misma tarea de aprendizaje supervisado para predecir diferentes emociones (felicidad, asco, miedo...). Los resultados obtenidos mostraron que los modelos de redes neuronales son capaces de predecir adecuadamente las distintas emociones. Además, no se encontraron grandes diferencias entre los modelos univariados y multivariados. También se analizó la especificidad de las predicciones de los modelos de redes neuronales encontrando una baja relación entre las predicciones emocionales de las palabras y las predicciones emocionales de sus vecinos semánticos. Las implicaciones de este estudio son relevantes desde un punto de vista teórico, puesto que se muestra cómo los modelos puramente simbólicos (computacionales) y, aparentemente, sin ningún contacto explícito con la emocionalidad, explican una buena parte de la variabilidad de los juicios humanos emocionales. Y también desde el punto de vista práctico puesto que tienen mucha utilidad en la generación de bases de datos normativas sobre la emocionalidad de las palabras que podrán ser útiles para futuras investigaciones.