



FACULTAD DE CIENCIAS SOCIALES
ESCUELA DE PSICOLOGÍA

ANÁLISIS EMPÍRICO Y TEÓRICO DEL EFECTO DE REDACCIÓN
EN LA MEDICIÓN DE DIFERENCIAS INDIVIDUALES

Por:

FERNANDO PONCE CISTERNAS

Tesis presentada a la Escuela de Psicología de la Pontificia Universidad Católica de Chile, para optar al grado académico de Doctor en Psicología.

Profesor Guía:

Alvaro Vergés Gómez, PhD

Comisión informante:

Patricio Cumsille Eltit, PhD

David Torres Iribarra, PhD

Víctor Arias González, PhD

Marzo, 2022
Santiago, Chile

© 2022, Fernando Ponce Cisternas

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica que acredita al trabajo y a su autor.

Este trabajo fue patrocinado y financiado por la Agencia Nacional de Investigación y Desarrollo (ANID), a través del Programa de Formación de Capital Humano Avanzado, Beca de Doctorado Nacional (N.21161101) otorgado a Fernando P. Ponce.

TABLA DE CONTENIDOS

LISTA DE TABLAS	iv
LISTA DE FIGURAS	vi
LISTA DE ECUACIONES	viii
RESUMEN	ix
INTRODUCCIÓN	1
I. ANTECEDENTES TEÓRICOS Y EMPÍRICOS	3
1.1. Escalas de auto-reporte, redacción de ítems y efectos de redacción	3
1.2. Efectos de redacción: ¿Varianza efímera o estilo de respuesta?	4
1.3. Limitaciones del modelamiento bifactor del efecto de redacción	5
II. PLAN DE TRABAJO	7
2.1. Modelo de Investigación: Escala de Autoestima de Rosenberg (RSES)	8
III. MANUSCRITOS	10
3.1. Manuscrito 1: <i>Sobre el uso y abuso del factor de redacción en la RSES: Una revisión sistemática y estimación de índices derivados.</i>	11
3.2. Manuscrito 2: <i>Efectos de redacción en evaluación: Cuando el bosque no deja ver los árboles.</i>	99
3.3. Manuscrito 3: <i>La naturaleza efímera del efecto de redacción.</i>	153
IV. DISCUSIÓN GENERAL	242
BIBLIOGRAFÍA	246

LISTA DE TABLAS

MANUSCRITO 1

Tabla 1. Resumen de los registros seleccionados.	63
Tabla Suplementaria S1. Lista de verificación PRISMA.	76
Tabla Suplementaria S2. Lista de verificación PRISMA 2021 para resúmenes estructurados.	78
Tabla Suplementaria S3. Resumen indicadores derivados de los modelos bifactor recolectados.	80

MANUSCRITO 2

Tabla 1. Estadísticos descriptivos de los ítems de la RSES por muestra.	134
Tabla 2. Índices de ajuste para soluciones confirmatorias entre dos y seis perfiles.	135
Tabla 3. Probabilidad media de clasificación al perfil latente (ALPCP) y prevalencia por muestra.	136
Tabla 4. Índices de ajuste para soluciones confirmatorias por condición.	137
Tabla Suplementaria S1. Cargas factoriales estandarizadas y error estándar (en paréntesis) en la muestra 1 por condición.	142
Tabla Suplementaria S2. Cargas factoriales estandarizadas y error estándar (en paréntesis) en la muestra 2 por condición.	144
Tabla Suplementaria S3. Cargas factoriales estandarizadas y error estándar (en paréntesis) en ambas muestras por condición.	146
Tabla Suplementaria S4. Distribución de frecuencias para los perfiles identificados por condición muestral.	148
Tabla Suplementaria S5. Estimaciones de convergencia a partir del modelo Bifactor-FID por condición muestral, número de iteraciones y criterios de convergencia.	149
Tabla Suplementaria S6. Valores propios observados y simulados extraídos del análisis paralelo (Horn, 1965).	152
Tabla Suplementaria S7. Matriz de cargas rotadas y correlación entre factores.	152

MANUSCRITO 3

Tabla 1. Estadísticos descriptivos por muestra.	218
Tabla 2. Índices de ajuste de las soluciones de perfiles latentes para las escalas RSES y LOT-R.	219
Tabla 3. Probabilidad media de clasificación del perfil latente (PMC) para la pertenencia al perfil latente más probable (fila) por perfil identificado (columna) para las escalas RSES y LOT-R.	220
Tabla 4. Análisis Factorial Confirmatorio de las escalas RSES y LOT-R por condición.	221
Tabla 5. Índices de ajuste de los modelos multirasgo-multimétodo.	223
Tabla 6. Probabilidad media de clasificación (PCM), frecuencia y proporción de participantes (FPP) por patrón de perfil latente identificado.	224
Tabla 7. Estadísticos descriptivos por ola.	225
Tabla 8. Índices de ajuste de los modelos confirmatorios por ola y longitudinal.	226
Tabla 9. Resumen de los modelos de invarianza estimados.	227
Tabla 10. Índices de ajuste de las soluciones de perfiles latentes para la escala RSES por ola.	228
Tabla 11. Probabilidad media de clasificación del perfil latente (PMC) para la pertenencia al perfil latente más probable por perfil identificado para la escala RSES por ola.	229
Tabla 12. Índices de ajuste de las soluciones de transiciones latentes estimadas.	229
Tabla 13. Probabilidad media de clasificación de los patrones de tres olas de transición (PMC), probabilidades de transición entre pares de olas (τ), y frecuencia en cada patrón de perfil identificado.	230
Tabla 14. Índices de ajuste y varianza explicada por factor de los modelos longitudinales estimados por condición.	231
Tabla 15. Correlaciones estimadas entre los factores especificados por modelo (M8-M10) en las condiciones muestrales examinadas	232

LISTA DE FIGURAS

MANUSCRITO 1

Figura 1. Modelos bifactor examinados.	66
Figura 2. Diagrama de flujo del proceso de revisión, identificación y selección de los estudios.	67
Figura 3. Resumen evaluación del riesgo de sesgo.	68
Figura 4. Coeficientes omega (ω) y omega jerárquico (ω_h) de los factores general de autoestima (círculos grises) y específico de redacción asociado a ítems inversos (círculos blancos), estimados del modelo Bifactor-FI.	69
Figura 5. Coeficientes omega (ω) y omega jerárquico (ω_h) de los factores general de autoestima (círculos grises) y específicos de redacción asociado a ítems directos (círculos negros) e inversos (círculos blancos), estimados del modelo Bifactor-FID.	70
Figura 6. Valores del índice H por factor y modelo bifactor.	71
Figura 7. Valores de índice de determinación del factor (DF) por factor y modelo bifactor.	72
Figura 8. Panel izquierdo. Valores de varianza común explicada (ECV) por factor y modelo bifactor. Panel derecho. Relación entre los valores de varianza común explicada y sesgo de parámetro relativo promedio (ARPB) para los modelos Bifactor-FI (B-FI, círculos negros) y Bifactor-FID (B-FID, círculos blancos).	73
Figura 9. Modelos RIIFA y FMM asumiendo ítems inversos recodificados.	74
Figura Suplementaria S1. Representación tabular del riesgo de sesgo en estudios individuales.	79

MANUSCRITO 2

Figura 1. Representación conceptual de las fuentes de varianza asumiendo homogeneidad poblacional.	138
Figura 2. Medias estimadas para los ítems positivos y negativos de la RSES por perfiles identificados en ambas muestras.	139
Figura 3. Panel a. Distribución de los valores de distancia de inmodelabilidad (d_r^2) con los modelos de un factor y bifactor para los perfiles ordenados y el perfil $AE_{MÉTODO}$. Panel b. Distribución de la contribución individual a los valores de chi-cuadrado (IND_{CHI}) para los perfiles ordenados y el perfil $AE_{MÉTODO}$.	140

MANUSCRITO 3

Figura 1. Medias estimadas de los ítems positivos y negativos de la RSES y LOT-R por perfiles identificados en ambas muestras.	233
Figura 2. Modelos confirmatorios para las escalas RSES y LOT-R.	234
Figura 3. Modelos confirmatorios multirasgo-multimétodo.	235
Figura 4. Diagrama de dispersión de las puntuaciones factoriales contenidas en la asociación entre FD_{OP} y FD_{AE} en la condición MC.	236
Figura 5. Modelos confirmatorios longitudinales	237
Figura 6. Medias estimadas de los ítems positivos y negativos de la RSES por perfiles identificados en cada ola de medición por tipo de análisis empleado.	238
Figura 7. Diagrama de dispersión de las puntuaciones factoriales contenidas en la asociación entre FI_{T1} y FI_{T2} en la condición MC.	239
Figura 8. Distribución de porcentajes acumulados de las puntuaciones factoriales de los subgrupos por ola y condición muestral. Los valores entre paréntesis representan la puntuación factorial promedio y la proporción de participantes con puntuaciones entre -1 y +1.	240
Figura 9. Diagrama de dispersión de las puntuaciones factoriales contenidas en la asociación entre FI_{T1} y FI_{T2} en la condición MC-M.	241

LISTA DE ECUACIONES

MANUSCRITO 1

Ec.1 Coeficiente omega (ω)	19
Ec.2 Coeficiente omega jerárquico (ωh)	20
Ec.3 Coeficiente omega jerárquico por sub-escala ($\omega h s$)	20
Ec.4 Índice de determinación del factor (DF)	21
Ec.5 Coeficiente de replicabilidad (H_j)	22
Ec.6 Varianza común explicada (ECV)	23
Ec.7 Número de correlaciones por factor ($ncor_j$)	24
Ec.8 Porcentaje de correlaciones no contaminadas (PUC)	24
Ec.9 Sesgo de parámetro relativo promedio ($ARPB$)	33
Ec.10 Modelo factorial clásico	37
Ec.11 Modelo factorial de ítems con intercepto aleatorio	38

RESUMEN

Con el propósito de controlar potenciales estilos de respuesta, los investigadores a menudo utilizan ítems directos e inversos para evaluar atributos psicológicos, bajo la premisa de que son psicométricamente equivalentes. Sin embargo, diversos estudios han mostrado que los encuestados exhiben patrones de respuestas diferenciales dependiendo de la dirección de los ítems. Esto se conoce como efecto de redacción y los autores han reportado que conlleva consecuencias indeseadas sobre la calidad de los datos, alterando la estimación de las propiedades métricas de las escalas de evaluación. Esta tesis se centra, a través de la secuencia progresiva de tres estudios, en el análisis teórico y empírico de la aproximación teórica dominante del efecto de redacción que versa sobre la noción del estilo de respuesta como mecanismo desencadenante de este sesgo, con características de consistencia entre escalas y estabilidad temporal.

El primer manuscrito contiene los resultados de una revisión sistemática de estudios que emplearon el modelo bifactor con la escala de autoestima de Rosenberg (RSES) como estrategia para modelar el efecto de redacción, junto con la estimación de un conjunto de índices derivados con el propósito de examinar el valor de los factores de redacción para servir como predictores o criterios en modelamiento de ecuaciones estructurales (SEM), considerando el amplio uso de esta estrategia tanto para modelar el efecto de redacción como para utilizar sus factores específicos para examinar conclusiones teóricas sustantivas respecto de la naturaleza subyacente de este sesgo de método.

En el segundo manuscrito se propone una estrategia analítica alternativa, basada en el modelamiento de mixturas, para superar las limitaciones detectadas con uso del modelo bifactor. Utilizando análisis de perfiles latentes (LPA) con los datos de la primera ola del panel LISS (n=6762), se identificó un subgrupo de individuos con patrones de respuesta consistentes con la noción de efectos de redacción (n=585; Estudio 1). Posteriormente, se examinó el impacto de este subgrupo tanto a nivel colectivo (Estudio 2) como individual (Estudio 3) en la mejora del ajuste estadístico y factibilidad del modelo bifactor sobre el modelo de un factor de la RSES.

Finalmente, el tercer manuscrito empleó esta estrategia analítica para examinar la naturaleza sustantiva o efímera del efecto de redacción, contrastando la evidencia previa reportada con modelamiento estructural. Se replicaron y extendieron los hallazgos reportados con la RSES a la escala de optimismo disposicional (LOT-R; Estudios 1 y 2), confirmando la ubicuidad del efecto de redacción en escalas compuestas por ítems directos e inversos. Con los datos de ambas escalas (n=5953), se encontró que, en presencia de una asociación significativa entre los factores de redacción (Estudio 3a), se identificó una baja proporción de participantes que exhibieron respuestas asimétricas en ambas escalas (Estudio 3b). Análogamente, utilizando tres olas de los datos longitudinales del panel LISS (n=3712), se examinó la estabilidad temporal e invarianza longitudinal de este efecto (Estudio 4a). De nuevo, pese a encontrar consistencia temporal, hubo una baja proporción de participantes consistentemente identificados con respuestas asimétricas a lo largo del tiempo (Estudio 4b), reflejado en parámetros de transición inferiores en comparación con los restantes patrones de perfiles de interés.

Tomados en conjunto, los resultados de estos tres estudios ponen en cuestionamiento la utilidad del modelo bifactor como estrategia para modelar efectos de redacción, así como las conclusiones obtenidas a partir de las asociaciones estimadas con los factores de redacción, debido a su inestabilidad y escasa interpretabilidad. Se presenta el modelamiento de mixturas como estrategia alternativa tanto para la detección del efecto como la examinación de su consistencia y estabilidad. A partir de esto último, se propone la noción de una naturaleza efímera del efecto de redacción, asociada a mecanismos de naturaleza transitoria.

Palabras clave: Medición, efectos de método, efectos de redacción, estilo de respuesta, bifactor, homogeneidad/heterogeneidad poblacional, modelamiento de mixturas.

INTRODUCCIÓN

La evaluación de atributos psicológicos es uno de los ámbitos que más estudio y atención ha concitado durante el desarrollo de la disciplina, existiendo un particular interés sobre aquellos que permiten dar cuenta de diferencias inter-individuales y que muestran cierta estabilidad intra-individual (Cronbach, 1956; Sackett et al., 2017). Es decir, aquellas variables que permiten diferenciar a los sujetos en características relativamente estables como para hacer descripciones y predicciones de la conducta del individuo en el entorno (Geukes et al., 2017).

Entre las estrategias metodológicas para la evaluación de este tipo de variables, con excepción de la inteligencia, las escalas de auto-reporte han predominado como estrategia metodológica de recolección de información por sobre la observación de la conducta (Baumeister, Vohs, & Funder, 2007; Haeffel & Howard, 2010). En términos generales, el auto-reporte demanda que el individuo, con el propósito de expresar su percepción respecto a una afirmación, realice una introspección que le permita juzgar su conducta potencial en situaciones hipotéticas o apelar a la memoria sobre su conducta frente a situaciones vividas. Por lo tanto, es necesario considerar aquellos aspectos que influyan o interfieran en los procesos cognitivos asociados a la comprensión, memoria, establecimiento de juicios y selección de respuestas (Tourangeau, Rips, & Rasinski, 2000; Wilson & Dunn, 2004), así como aquellos factores inespecíficos de la instancia de evaluación, susceptibles de influir en el procesamiento de las afirmaciones de los participantes (Podsakoff et al., 2003; 2012).

La dirección con que las afirmaciones son redactadas es uno de los factores cuyo impacto sobre las respuestas de los individuos fue tempranamente descrito (Rugg, 1941), y cuyas consecuencias sobre la calidad de la medida han sido ampliamente descritas (Greenberger et al., 2003; Huang et al., 2015; Wood et al., 2017), pero cuya naturaleza subyacente y mecanismos involucrados todavía está en discusión.

En el marco de la discusión teórica y metodológica sobre este tema, la presente tesis pretende abordar la evidencia empírica y las propuestas teóricas planteadas en relación al efecto de la redacción de los ítems en el proceso de respuesta a un test y sus

consecuencias para la estimación de modelos de medida y la asociación con variables criterio, aportando a esta discusión desde tres objetivos centrales: (a) evaluar el valor del modelo bifactor para examinar el efecto de redacción, examinando las propiedades del factor general y los factores específicos de redacción, (b) proponer una estrategia analítica alternativa, basada en la identificación de un grupo de individuos, para examinar el efecto de redacción, y (c) emplear esta estrategia para evaluar la consistencia entre atributos y la estabilidad temporal del efecto de redacción. La tesis comprendió la ejecución de tres estudios secuenciales, presentados en tres manuscritos, los que en su conjunto abarcaron los objetivos propuestos.

Este documento se organiza en cuatro secciones. La primera sección presenta los antecedentes teóricos que sostienen la problematización y justificación de los estudios presentados en esta tesis. La segunda sección muestra la descripción del plan de trabajo que organiza los tres estudios ejecutados y el modelo de investigación propuesto. La tercera sección incluye los tres manuscritos elaborados. Finalmente, en la cuarta sección se presenta la discusión general de esta tesis, incluyendo limitaciones y proyecciones para futuras investigaciones.

I. ANTECEDENTES TEÓRICOS

1.1. Escalas de auto-reporte, redacción de ítems y efectos de redacción

En el marco de la construcción de escalas de auto-reporte para medir atributos psicológicos, diversos autores han sugerido incorporar ítems directos e inversos para medir el constructo de interés con el propósito de contrarrestar formas estereotipadas de respuesta (Likert, 1932; Nunnally, 1978; Paulhus, 1991), actuando como “obstáculos cognitivos” que promueven que los encuestados, al momento de responder, se involucren con un procesamiento cognitivo más controlado (Podsakoff et al., 2003). Adicionalmente, la idea de incorporar ítems inversos descansa en el supuesto de que un atributo puede ser conceptualizado como un conjunto de observables (Cureton, 1951, citado en Markus & Borsboom, 2013), donde ítems directos e inversos son indicadores psicométricamente equivalentes del constructo de interés y cuya única diferencia reside en la dirección de la afirmación.

Sin embargo, pese al uso masivo del balanceo de ítems directos e inversos en la construcción y aplicación de instrumentos psicológicos, la noción de equivalencia fue tempranamente resistida por algunos autores (*e.g.*, Schriesheim et al., 1991; Spector et al., 1997), quienes dieron cuenta de respuestas diferenciales entre ambos subconjuntos de ítems (también descrito como efecto de redacción). Esto conlleva la emergencia de varianzas sistemáticas de error asociada al método de redacción de los ítems, ajena a la varianza del atributo, produciendo alteraciones que inducen: (a) relaciones espurias con otras variables criterio (Billiet & McClendon, 2000; Huang et al., 2015; McPherson & Mohr, 2005), (b) alteraciones en la fiabilidad de la escala (Roszkowski & Soven, 2010; Wood et al., 2017), o desajustes en la dimensionalidad debido a la emergencia de factores asociados al conjunto específico de ítems (Greenberger et al., 2003; Woods, 2006), por mencionar algunos ejemplos. Diversos autores han reportado, empleando distintas escalas compuestas por ítems directos e inversos, la presencia de este efecto y sus respectivas consecuencias (Biderman et al., 2011; Dueber et al., 2021; Kam, 2018).

Estos hallazgos llevaron a los investigadores a utilizar el modelo confirmatorio bifactor (ver Figura 1, p.66) dada su ventaja comparativa por sobre otros modelos para dar

cuenta de los efectos de método, reflejado en la mejora exhibida en el ajuste estadístico respecto de otros modelos competidores (Maydeu-Olivares & Coffman, 2006; Widaman, 2012), y, especialmente, en el valor informativo de los factores específicos de redacción para examinar si el efecto de redacción refleja un estilo de respuesta estable (como un rasgo de personalidad) o representa varianza efímera que requiere ser estadísticamente controlada (DiStefano & Motl, 2006; Marsh, 1996; Marsh et al., 2010).

1.2. Efectos de redacción: ¿Varianza efímera o estilo de respuesta?

Otra cuestión importante que se ha planteado se refiere a la definición y significado subyacente del efecto del método. Existen al menos dos concepciones opuestas sobre su naturaleza. Por un lado, algunos autores han señalado que los factores asociados a la redacción de los ítems constituyen *artefactos metodológicos* (ver Cronbach, 1946), cuya varianza sistemática de error debe ser incorporada en el modelo de medición para minimizar su efecto sobre la varianza asociada al rasgo de interés (Marsh, 1996; Weijters et al., 2013).

Por ejemplo, Weijters et al (2013) propusieron este efecto como resultado de la acción conjunta de tres mecanismos (aquiescencia, descuido y sesgo de confirmación) en interacción con la configuración y distribución de los ítems, proponiendo que las características de la escala ejercen un impacto diferencial sobre estos mecanismos. Por un lado, la dirección del primer ítem del instrumento (*i.e.*, directo o inverso) tiene un efecto sobre el sesgo de confirmación, mientras que la distribución de los ítems (masiva o distribuida) tendría un efecto sobre el descuido y el sesgo de confirmación, y un efecto nulo sobre la aquiescencia.

Sin embargo, tanto los resultados reportados por estos autores como hallazgos posteriores (Cabooter, et al., 2016; 2017) sólo respaldaron parcialmente el funcionamiento diferencial de los mecanismos propuestos por Weijters et al. (2013). En este sentido, el modelo basado en características del instrumento carece de suficiente evidencia empírica, tanto en su formulación original como en investigaciones posteriores.

Por otra lado, un extenso cuerpo de evidencia respalda la noción del efecto de redacción como expresión de un *estilo de respuesta* asociado a características individuales

o atributos potencialmente medibles (Jackson & Messick, 1958), que opera de manera consistente al momento de evaluar distintos atributos y es estable a lo largo del tiempo (Bentler, Jackson, & Messick, 1971; Tomas & Oliver, 1999).

Utilizando los factores de redacción del modelo bifactor, una serie de estudios dieron cuenta de la validez del efecto de redacción al asociarlo con atributos como la ansiedad física social (Motl & Conroy, 2000; Motl et al., 2000), miedo al fracaso (Conroy, 2001), auto-motivación (Motl et al., 2003), autovaloraciones centrales (Arias & Arias, 2017; Gu et al., 2015), autoestima (DiStefano y Motl, 2006; Lindwall et al., 2012; Quilty et al., 2006) u optimismo disposicional (Alessandri et al., 2011), por mencionar algunos. Similarmente, otros autores reportaron evidencia respaldando la noción del efecto de redacción como estilo de respuesta, tras encontrar asociaciones significativas entre factores de redacción de distintas escalas (DiStefano & Motl, 2006; Horan, DiStefano, & Motl, 2003), así como entre factores de redacción a lo largo del tiempo (Marsh et al., 2010; Motl & DiStefano, 2002), exhibiendo, en todos los casos, excelentes índices de ajuste.

Sobre la base de este tipo de hallazgos, existe un amplio consenso en cuanto al valor del modelo bifactor como herramienta para modelar la multidimensionalidad generada por el efecto de redacción y, como veremos más adelante, del factor específico de redacción para dar cuenta de las diferencias individuales asociadas a este efecto y, por extensión, de las relaciones nomológicas estimadas con este factor. Precisamente, como lo reflejan los estudios citados, el efecto de redacción como un estilo de respuesta ha sido la noción que ha predominado durante las últimas dos décadas.

1.3. Limitaciones del modelamiento bifactor del efecto de redacción

Los hallazgos que respaldan la noción del estilo de respuesta no están exentos de una serie de problemas o limitaciones ligadas al modelamiento e interpretabilidad del modelo bifactor, que han sido ignorados o directamente omitidos, pero que recientemente han sido tomados en cuenta en el contexto de la aplicabilidad del modelo bifactor en la investigación clínica (Bonifay et al., 2017; Watts et al., 2019).

El primer problema refiere a la propensión al ajuste identificada en el modelo bifactor (Bonifay & Cai, 2017) debido a su tendencia para acomodar patrones de respuesta

implausibles y potencialmente inválidos (Reise et al., 2016), dada la forma funcional derivada de su especificación. Esto induce a la estimación de parámetros potencialmente inestables, generando factores con interpretabilidad cuestionable (Bonifay, 2015; Bonifay et al., 2017; Watts et al., 2019). Lo cual pone en duda tanto la replicabilidad de las asociaciones estimadas de los factores de redacción del modelo bifactor con variables criterio como la validez misma de los factores de redacción extraídos. Esto fue examinado en la revisión sistemática reportada en el primer manuscrito el cual, siguiendo el primer objetivo propuesto, examinó las propiedades del factor general y específicos de redacción, utilizando una serie de indicadores complementarios calculados a partir de las cargas factoriales estimadas del modelo bifactor (Rodríguez et al., 2016a,b).

El segundo problema asociado al uso del modelo bifactor es conceptual y refiere a su pertinencia para representar adecuadamente el efecto de redacción. Esto surge a propósito de los hallazgos de diversos investigadores que sugieren que sólo un porcentaje relativamente bajo de participantes exhibiría efectos de redacción (Reise et al., 2016; Huang et al., 2015; Woods, 2006), lo cual lleva a cuestionar si es posible asumir *homogeneidad poblacional* (Muthén, 1989) como para imponer un único set de parámetros bajo la especificación bifactor como consecuencia del patrón de respuesta de una porción significativamente menor de participantes. Desafortunadamente, no es posible implementar modelamiento multi-grupo (Jöreskog, 1971; Sörbom, 1974) ya que este grupo de participantes con respuestas sesgadas es desconocido.

II. PLAN DE TRABAJO

Esta tesis tuvo tres objetivos principales: (a) evaluar el valor del modelo bifactor para examinar el efecto de redacción, examinando las propiedades del factor general y los factores específicos de redacción, (b) proponer una estrategia analítica alternativa, basada en la identificación de un grupo de individuos, para examinar el efecto de redacción, y (c) emplear esta estrategia para evaluar la consistencia entre atributos y la estabilidad temporal del efecto de redacción.

Siguiendo la estructura propuesta de tres objetivos, este trabajo se realizó en tres fases. La primera fase estuvo dirigida a llevar a cabo una revisión sistemática, siguiendo las directrices del protocolo PRISMA para revisiones sistemáticas y meta-análisis (Page et al., 2021), de los modelos bifactor reportados de la escala de autoestima de Rosenberg (1965) y archivos de datos disponibles para poder estimar dicho modelo. A continuación, se estimaron diversos indicadores derivados del modelo bifactor para examinar el grado en que los factores de rasgo y redacción son psicométricamente interpretables y la adecuación de los factores de redacción para examinar la naturaleza subyacente del efecto de redacción.

Posteriormente, con una idea sobre la pertinencia del modelo bifactor para esta clase de investigaciones, la segunda fase se orientó a proponer una estrategia analítica alternativa, basada en la identificación de individuos con respuestas diferenciales en función de la dirección de los ítems. Utilizando los datos de la escala de autoestima de Rosenberg de la primera ola del panel LISS (n=6762), se utilizó análisis de perfiles latentes para evaluar la presencia de un grupo de participantes con el patrón de respuesta de interés, junto con una posterior evaluación de su impacto sobre las diferencias en el ajuste estadístico a favor del modelo bifactor. Esto permitiría obtener, utilizando una técnica basada en heterogeneidad poblacional, una solución de perfiles que permitiera la detección de un grupo de individuos que exhibiera patrones de respuestas y un impacto sobre modelos factoriales consistentes con los propuestos en la literatura.

Finalmente, utilizando la noción de heterogeneidad poblacional y sus estrategias analíticas derivadas, la tercera fase buscó examinar la naturaleza subyacente (efímera o

sustantiva) del efecto de redacción. En primer lugar, se examinó separadamente la presencia del subgrupo identificado con respuestas asimétricas y sus consecuencias sobre el modelo de medida, con el propósito de replicar la solución reportada previamente con la escala de autoestima y extender este hallazgo a la escala de optimismo disposicional, utilizando los datos de la quinta ola del panel LISS (n=5927). Posteriormente, utilizando ambas escalas, se estimó la presencia de participantes con respuestas diferenciales conjuntas en ambas escalas para evaluar el grado con que los participantes exhiben consistentemente el efecto de redacción. Finalmente, utilizando tres olas de los datos longitudinales de la escala de autoestima del panel LISS (n=3712), se estimó la presencia de participantes que exhibieron sistemáticamente el patrón de respuesta diferencial a lo largo del tiempo para evaluar el grado de estabilidad temporal del efecto de redacción.

2.1. Modelo de Investigación: Escala de Autoestima de Rosenberg (RSES)

Con el propósito de abordar los aspectos teóricos y empíricos del efecto de redacción en la evaluación de variables individuales con dispositivos balanceados de auto-reporte, se utilizó la Escala de Autoestima de Rosenberg (1965) como modelo de investigación.

Entre las razones que justifican esta decisión se puede mencionar que es una de las escalas de auto-reporte en la que consistentemente ha sido corroborada la presencia de efectos de redacción a partir de la estimación de la variable latente de redacción en numerosas investigaciones (Alessandri et al., 2013; 2015; Corwyn, 2000; DiStefano & Motl, 2006; 2009a,b; Horan et al., 2003; Hyland et al., 2014; Marsh, 1996; Marsh et al., 2010; Michaelides et al., 2016a,b; Motl & DiStefano, 2002), la que ha sido utilizada para evaluar su relación con otras variables criterio (DiStefano y Motl, 2006; 2009a; Lindwall et al., 2012; Michaelides et al., 2016a,b; Quilty et al., 2006; Tomas et al., 2013).

En segundo lugar, es la única escala cuyo factor de redacción acumula evidencia, empleando modelamiento estructural con un mismo set de datos, de consistencia del efecto de redacción entre atributos y estabilidad temporal (Horan et al., 2003; DiStefano y Motl, 2006), siendo esto una oportunidad para la evaluación conjunta de ambas piezas de evidencia utilizando estrategias alternativas con un mismo set de datos.

Finalmente, en términos de la investigación aplicada, esta escala es una de las más frecuentemente aplicadas para evaluar una de las variables individuales más extensamente utilizadas en el estudio del individuo, su conducta y sus relaciones sociales (Donnellan et al., 2013), así como su desempeño en ámbitos de la vida cotidiana (Baumeister et al., 2003). A partir de esto, los resultados que se pueden obtener en relación a la naturaleza sustantiva del sesgo en la evaluación de este atributo pueden significar contribuciones relevantes a la interpretación de la evidencia obtenida con esta escala.

III. MANUSCRITOS

Esta Tesis comprende la ejecución de diez estudios secuenciales, presentados en tres manuscritos, los que en su conjunto buscaron cumplir con los objetivos propuestos.

En el primer manuscrito se reporta el proceso de búsqueda, selección y revisión de una serie de estudios y archivos de datos, junto con la estimación de indicadores para evaluar la adecuación de la solución bifactor y los factores específicos para examinar el efecto de redacción. El material suplementario contiene las listas de verificación del protocolo PRISMA, la evaluación del grado de riesgo por registro evaluado, y el resumen de los indicadores estimados en cada modelo bifactor evaluado.

El segundo manuscrito, publicado en la revista *Multivariate Behavioral Research*, reporta, en tres estudios, la evaluación y selección de la solución de perfiles latentes, junto con la evaluación del impacto colectivo e individual de los participantes identificados con respuestas asimétricas de acuerdo a la dirección de los ítems. El material suplementario muestra las cargas factoriales estandarizadas de los modelos factoriales estimados en el Estudio 2, junto con la examinación exhaustiva de la falta de convergencia identificada.

Finalmente, el tercer manuscrito presenta los resultados de los seis estudios llevados a cabo para identificar y examinar el impacto de los individuos con respuestas asimétricas en las escalas de autoestima y optimismo disposicional, junto con evaluar la naturaleza sustantiva o efímera del efecto de redacción, a través de la examinación de la evidencia tanto de consistencia entre escalas como estabilidad temporal.

3.1. MANUSCRITO 1:

Sobre el uso y abuso del factor de redacción en la RSES: Una revisión sistemática, análisis de datos secundarios y estimación de índices derivados

Manuscrito presentado para revisión en Psychological Bulletin

Fernando P. Ponce¹, Fabiola Gómez¹, Víctor B. Arias² & Patricio Cumsille¹

¹ Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile.

² Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Facultad de Psicología, Universidad de Salamanca.

Resumen

Basados en el ajuste estadístico obtenido, diversos estudios han propuesto la estructura bifactor como modelo óptimo para representar el efecto de redacción observado con escalas compuestas de ítems directos e inversos. Esta configuración permite separar la varianza del atributo y la atribuible al método de medición, y utilizar esta última para examinar la naturaleza sustantiva o efímera del efecto de redacción. Sin embargo, la idoneidad de estos factores de método para este uso es incierta debido al riesgo de sobreajuste y la falta de estabilidad e interpretabilidad de los parámetros estimados, aspectos pobremente examinados en investigaciones previas.

Para abordar esto, examinamos el grado en que los factores de atributo y redacción son psicométricamente interpretables y adecuadamente definidos para servir como predictores o criterios en modelos de ecuaciones estructurales, utilizando como ejemplo la Escala de Autoestima de Rosenberg. Para cumplir con este objetivo, llevamos a cabo una revisión sistemática utilizando cuatro bases de datos (Web of Science, Scopus, PsycArticles, y Google Scholar), lo que dio como resultado 71 registros que reportaron las cargas factoriales del modelo bifactor o dejaron disponibles sus datos (60 estudios y 11 archivos de datos). Los resultados sugieren una mayor predominancia del factor general en todos los indicadores asociados a la varianza común explicada, confiabilidad, estabilidad y replicabilidad del constructo; mientras que los factores de método exhibieron índices que indican baja replicabilidad y consistencia.

Estos hallazgos respaldan la idea de que a la base de la Escala de Autoestima de Rosenberg hay una estructura esencialmente unidimensional, y que los factores de método no pueden ser considerados como criterios apropiados para examinar la naturaleza subyacente del efecto de redacción. (Registro OSF: <https://osf.io/96gvn/>).

Palabras clave: Efectos de redacción, efectos de método, modelo bifactor, escala de autoestima de Rosenberg, coeficiente omega, replicabilidad de constructo, varianza común explicada.

El efecto de redacción es un sesgo de respuesta comúnmente observado en escalas compuestas por ítems directos e inversos (Podsakoff et al., 2003; 2012). Este efecto ha sido descrito como una tendencia a responder diferencialmente a ambos subsets de ítems, produciendo alteraciones sustanciales en la calidad de los datos que pueden inducir (a) relaciones espurias entre variables que, de otro modo, no estarían correlacionadas (Huang et al., 2015), (b) variaciones en la consistencia interna de las escalas (Wood et al., 2017), (c) el desajuste en modelos unidimensionales o la aparición de factores de método en escalas balanceadas (Greenberger et al., 2003), solo por mencionar algunos ejemplos. Este tipo de consecuencias contradicen la idea original de que la inclusión de ítems negativamente redactados ayuda a detectar o eliminar estilos de respuesta (Dalal & Carter, 2014).

Distintos autores han propuesto el modelo bifactor como estrategia analítica para hacer frente a este problema, mediante la incorporación de factores asociados a la dirección con que los ítems son redactados. Esta estrategia se justifica, fundamentalmente, en el buen ajuste estadístico obtenido con el modelo bifactor, en comparación al obtenido con los modelos originalmente propuestos (*e.g.*, un factor general del rasgo de interés). Más importante aún, utilizando los factores de redacción como entidades representativas de las diferencias individuales en el efecto de redacción, varios investigadores han empleado el modelo bifactor para examinar la consistencia del efecto de redacción en diferentes constructos, su estabilidad temporal y su asociación con variables criterio, sugiriendo la naturaleza sustantiva de este efecto de método.

Sin embargo, recientes estudios han cuestionado la utilidad del modelo bifactor, dada su propensión a acomodar patrones de respuesta implausibles (Reise et al., 2016) y, en consecuencia, distorsionar la magnitud de los índices de ajuste (Bonifay & Cai, 2017; Markon, 2019) así como la tendencia de este modelo a enmascarar estructuras de datos previamente simuladas para representar modelos de segundo orden (Morgan et al., 2015; Murray & Johnson, 2013) o a contener heterogeneidad no observada (Raykov et al., 2019). Por lo tanto, la solución bifactor podría generar estimaciones de parámetros inestables,

que no se replican entre muestras o en distintas mediciones de la misma muestra a lo largo del tiempo.

Considerando que un buen ajuste obtenido de una solución bifactor no reflejará necesariamente una buena calidad del modelo (Rodríguez, Reise, & Haviland, 2016a), su idoneidad para representar y examinar las propiedades del efecto de redacción debe verificarse, utilizando índices complementarios que permitan evaluar la calidad del modelo, en términos de fiabilidad, reproductibilidad y aplicabilidad en escenarios de modelamiento estructural. Sin embargo, la mayoría de los estudios que han empleado el modelo bifactor para examinar efectos de redacción no han reportado ni interpretado en detalle este tipo de indicadores, por lo que la interpretabilidad de los factores extraídos y las asociaciones observadas es insuficiente.

En el presente estudio, seleccionamos un conjunto de trabajos que emplearon el modelo bifactor para examinar la dimensionalidad de la escala de autoestima de Rosenberg (1965) y calculamos diferentes índices alternativos para evaluar la idoneidad del modelo para dar cuenta de los efectos de redacción.

Preguntas opuestas para evaluar el mismo atributo

Cuando se construyen instrumentos destinados a evaluar atributos psicológicos se suelen utilizar ítems que dan cuenta de la presencia y ausencia de dicho atributo. Por ejemplo, una escala que mide autoestima incluye *ítems directos* (e.g., “Tengo una actitud positiva hacia mí mismo/a”) cuyo grado de acuerdo refleja un mayor nivel del atributo y, al mismo tiempo, incorporar *ítems inversos* (e.g., “A veces me siento verdaderamente inútil”) cuyo acuerdo expresa un menor nivel del atributo. Esta estrategia se basa en el supuesto de que un atributo puede conceptualizarse como un conjunto de observables (Cureton, 1951, citado en Markus y Borsboom, 2013), donde tanto los ítems directos como los inversos son indicadores psicométricamente equivalentes del constructo de interés, cuya única diferencia es la dirección con que la afirmación es redactada.

Sin embargo, los individuos pueden exhibir patrones de respuesta diferencial a los ítems directos e inversos (Kamoen et al., 2013; Quilty et al., 2006), lo que, potencialmente, produce varianza sistemática asociada al método (*i.e.*, efecto de redacción). Un extenso

cuerpo de evidencia ha revelado que el efecto de redacción tiene un impacto sustantivo en la calidad de los datos, produciendo alteraciones en las estimaciones de consistencia interna (Schriesheim et al., 1991; Wood et al., 2017), la estructura de cargas factoriales (Savalei and Falk, 2014), o el desajuste estadístico por multidimensionalidad asociada a la emergencia de varianza sistemática de error (Greenberger et al., 2003; Ponce et al., 2021), entre otras consecuencias.

Las variaciones en el ajuste estadístico observadas en escalas compuestas por ítems directos e inversos han promovido un extenso debate acerca de la dimensionalidad subyacente de los atributos que miden. Por ejemplo, en el caso de la escala de autoestima de Rosenberg (RSES; Rosenberg, 1965), algunos investigadores han sugerido que la escala refleja una estructura bidimensional, con dos facetas referidas a autoestima positiva y negativa (Ang et al., 2006; Owens, 1994). Sin embargo, otros indican que ambas dimensiones se correlacionan de manera equivalente con variables criterio, sugiriendo que se trata de una estructura unidimensional y que ambos factores emergen a partir de efectos de método (Carmines & Zeller, 1979; Huang & Dong, 2012), producidos por la respuesta inconsistente entre ítems directos e inversos (Schmitt & Allik, 2005). A partir de este debate, se han planteado dos preguntas: (a) ¿Cuál estructura factorial representa mejor la dimensionalidad subyacente? (Corwyn, 2000; Lindwall et al., 2012; Quilty et al., 2006; Tomás & Oliver, 1999) y, (b) el efecto de redacción ¿representa un estilo de respuesta estable (como un rasgo de personalidad) o varianza efímera que requiere ser controlada estadísticamente? (DiStefano & Motl, 2006; Marsh, 1996; Marsh et al., 2010).

Para abordar estas inquietudes, el modelo bifactor ha sido extensamente utilizado tanto para representar la separación de las fuentes de varianza asociadas al atributo y al método, como para utilizar estas fuentes de varianza de método y examinar su asociación con variables criterio.

Modelo bifactor para examinar efectos de redacción: Aplicaciones y limitaciones

El modelo bifactor (Holzinger & Swineford, 1937; Reise, 2012) ha sido propuesto como estrategia analítica para separar la fuente de varianza asociada al atributo de la asociada al método de medición (Horan et al., 2003). El modelo bifactor incluye un factor general

asociado al atributo (*e.g.*, autoestima; AE en la Figura 1), que representa la varianza compartida por todos los ítems, y un conjunto de factores específicos cuya varianza es compartida por un grupo de ítems que comparten contenido similar o, en este caso, un mismo tipo de redacción (*e.g.*, Factor Directo [FD], para ítems directos, y Factor Inverso [FI] para ítems inversos; *vid.*, Figura 1). De esta manera, la varianza de cada ítem se parcela en una parte común o compartida por todos los ítems (*i.e.*, varianza del atributo medido) y una parte compartida por sólo algunos ítems (*i.e.*, varianza específica o del método de medición). A partir de esta noción, FD y FI han sido tratados como entidades que representan los efectos de redacción. En el caso de la RSES, los modelos bifactor más frecuentemente reportados son los que contienen ambos factores de redacción o sólo el factor de redacción asociado a ítems inversos (en adelante, Bifactor-FID y -FI, respectivamente).

[Insertar Figura 1 aproximadamente aquí]

Como parte de la especificación del modelo, las correlaciones entre los factores de rasgo y redacción son restringidas a cero (Reise et al., 2007). Esta ortogonalidad permite separar la varianza compartida por todos los ítems de aquella asociada a grupos específicos, permitiendo examinar la partición de la varianza total cuando se asume que la escala contiene fuentes de varianza común tanto a todos los ítems como a un grupo específico de ellos (Simms, et al., 2008); este modelo permite además controlar la multidimensionalidad generada por la presencia de varianza común no sustantiva, en un modelo que puede ser considerado como “esencialmente unidimensional” (Raykov & Pohl, 2013). Esto es relevante en la medida que no es realista la idea de que un ítem destinado a evaluar un atributo está libre de especificidades que puedan producir factores vagamente definidos (Cattell, 1978; Marsh et al., 2004; Morin et al., 2015).

Con respecto a la inspección de la naturaleza sustantiva o efímera del efecto de redacción, los factores de método del modelo bifactor han sido empleados para estimar su relación con diversas variables criterio (Lance et al., 2002). Siguiendo esta estrategia, algunos autores han propuesto que los efectos de redacción son un estilo de respuesta tras reportar evidencia de estacionalidad y estabilidad temporal del efecto de redacción (Gana et al., 2013; Marsh et al., 2010; Motl & DiStefano, 2002), así como de validez convergente entre escalas (DiStefano & Motl, 2006; Horan et al., 2003) y de criterio con variables asociadas a

diferencias individuales (DiStefano & Motl, 2009a; Michaelides, Zenger et al., 2016; Quilty et al., 2006).

Tomando en consideración estos antecedentes, el modelo bifactor parece adecuado para representar la multidimensionalidad que emerge de la presencia de respuestas asimétricas que son consistentes con la noción de efectos de redacción, permitiendo agrupar la varianza sistemática atribuible a la dirección de los ítems y utilizar los factores de método extraídos para examinar su estabilidad a lo largo del tiempo, o su relación con factores de redacción de otras escalas o con otros atributos, exhibiendo, en todos los casos, excelentes índices de ajuste. En un reciente meta-análisis, Gnambis et al. (2018) emplearon modelos de ecuaciones estructurales meta-analíticos (MASEM; Cheung, 2014; Gnambis & Staufienbiel, 2016) y CFA con la matriz de correlaciones combinada de 113 muestras independientes, encontrando que la RSES exhibió multidimensionalidad asociada a la redacción de los ítems y que el modelo Bifactor-FID fue el que mostró el mejor ajuste.

Sin embargo, la selección e interpretabilidad del modelo bifactor a partir del ajuste estadístico obtenido ha sido extensamente cuestionada dada su propensión a acomodar todo tipo de patrones de respuesta (*i.e.*, sobreajuste; Bonifay & Cai, 2017; Ponce et al., 2021; Reise et al., 2016), incluyendo patrones de respuesta implausibles y potencialmente inválidos (Reise et al., 2016), lo cual induce a la estimación de parámetros inestables y factores potencialmente no interpretables (Bonifay, 2015; Bonifay et al., 2017; Watts et al., 2019). Esto último es de particular interés a partir del uso extendido de los factores de redacción como criterio para examinar el efecto de redacción como estilo de respuesta.

En vista de lo anterior, el apropiado uso e interpretación del modelo bifactor, así como de los factores de redacción, no sólo depende del ajuste estadístico obtenido sino también del grado en que las variables latentes del modelo exhiben una adecuada representación de los indicadores que contienen, reflejado en la magnitud de sus cargas factoriales (Reise et al., 2013). A partir de la ortogonalidad entre las variables latentes del modelo bifactor, es posible evaluar las características de cada factor independientemente, utilizando diversos índices psicométricos relevantes para evaluar medidas psicológicas (Rodríguez et al., 2016b).

Índices derivados del modelo bifactor

Los índices aquí examinados proporcionan información acerca de la calidad del modelo en relación a dos objetivos fundamentales (Rodríguez et al., 2016a). El primero, y de particular valor para la investigación en evaluación, consiste en determinar el grado de precisión con que las puntuaciones totales y de subescala reflejan los dominios previstos, junto con examinar si los puntajes de las subescalas proporcionan información única por encima y más allá del puntaje total. El segundo objetivo, de interés en la investigación teórica, comprende determinar el grado en que los ítems reflejan adecuadamente las variables latentes consideradas en el modelo bifactor y la idoneidad del modelo en la estimación de asociaciones entre constructos en el contexto del modelamiento de ecuaciones estructurales.

En el presente estudio, examinamos diversos índices psicométricos basados en la revisión de la literatura sobre el modelo bifactor de la RSES que complementan el análisis de la calidad del modelo a partir de los objetivos anteriormente descritos: coeficientes de confiabilidad omega (ω , ω_h , ω_{hs} ; McDonald, 1999), varianza común explicada (*ECV*; Ten Berge & Sočan, 2004), porcentaje de correlaciones no contaminadas (*PUC*; Bonifay et al., 2015; Reise et al., 2013), determinación del factor (*DF*; Grice, 2001), y replicabilidad del constructo (*H*; Hancock & Mueller, 2001). Estos indicadores han demostrado un adecuado funcionamiento con modelos bifactor cuyos factores específicos contienen igual número de ítems (Liu et al., 2022).

Coefficiente omega. Para evaluar la interpretabilidad de una puntuación total o de subescala en condiciones de multidimensionalidad generada por la presencia de ítems directos e inversos, se requieren índices que estimen el grado en que la varianza de las puntuaciones totales o de subescala es atribuible a la varianza asociada a un solo factor (ya sea del atributo o de método). En el caso del modelo bifactor, los autores han sugerido estimar esto empleando el coeficiente omega, omega jerárquico y omega jerárquico de la subescala (Canivez, 2016; Reise, Moore, & Haviland, 2010).

El coeficiente omega (ω) es una estimación de la confiabilidad del puntaje total, basada en el modelo factorial estimado (McDonald, 1999; Revelle & Zinbarg, 2009;

Zinbarg et al., 2005). El valor de omega representa la proporción de varianza del puntaje total atribuible a todas las fuentes de varianza común incluidas en el modelo factorial (Revelle & Zinbarg, 2009).

Para el modelo Bifactor-FID de la RSES, el factor de autoestima (AE) representa el factor general del modelo y los ítems directos e inversos cargan en los factores de redacción FI y FD, respectivamente. A partir de esta especificación, la expresión del coeficiente omega del puntaje global es la siguiente:

$$\omega = \frac{(\sum \lambda_{i,AE})^2 + (\sum \lambda_{i,FI})^2 + (\sum \lambda_{i,FD})^2}{(\sum \lambda_{i,AE})^2 + (\sum \lambda_{i,FI})^2 + (\sum \lambda_{i,FD})^2 + \sum(1 - h_i^2)} \quad (1)$$

Donde $\lambda_{i,j}$ representa la carga factorial estandarizada del ítem i en el factor j . De este modo, el numerador incluye todas las fuentes de varianza común asociadas a los factores especificados en el modelo, y el denominador incluye estos mismos componentes junto con la varianza única (*i.e.*, $\sum(1 - h_i^2)$). En el modelo Bifactor-FI, que incluye un factor común y un factor de método para los ítems inversos, el término $(\sum \lambda_{i,FD})^2$ sería igual a cero dado que no hay varianza común especificada para el conjunto de ítems directos.

Independiente del modelo bifactor empleado, en la estimación del coeficiente omega se incluyen conjuntamente la varianza del factor general y de los factores de redacción para calcular un estimador de la confiabilidad. Para clarificar la contribución relativa de cada una de las fuentes de varianza común, el coeficiente omega jerárquico (ω_h) y el omega jerárquico por subescala (ω_{hs}) permiten capturar separadamente la varianza confiable atribuible al factor general y los factores específicos (McDonald, 1999; Reise, 2012; Zinbarg et al., 2005), respectivamente.

El coeficiente omega jerárquico (ω_h) estima la proporción de varianza del puntaje total que puede atribuirse al factor general, por lo que la variabilidad en las puntuaciones asociada a los factores específicos se trata como error de medición (Rodríguez, Reise, & Haviland, 2016b; Zinbarg et al., 2005; 2006). Este coeficiente se calcula dividiendo el cuadrado de la suma de las cargas factoriales del factor general por la varianza del puntaje total del modelo estimado:

$$\omega h = \frac{(\sum \lambda_{i,AE})^2}{(\sum \lambda_{i,AE})^2 + (\sum \lambda_{i,FI})^2 + (\sum \lambda_{i,FD})^2 + \sum(1 - h_i^2)} \quad (2)$$

Tal como en la ecuación anterior, $\lambda_{i,j}$ representa la carga factorial estandarizada del ítem i en el factor j y $(1 - h_i^2)$ representa la varianza de error. El valor del omega jerárquico refleja el porcentaje de varianza del puntaje total atribuible a las diferencias individuales en el factor general (Rodríguez, Reise, & Haviland, 2016b). Tal como en la Ecuación 1, en el modelo Bifactor-FI, el término $(\sum \lambda_{FD})^2$ sería igual a cero. Altos valores indican una mayor preponderancia del factor general como fuente de varianza sistemática y una menor influencia de la multidimensionalidad capturada por los factores específicos.

Del mismo modo, es posible estimar la influencia del factor general y los factores específicos sobre los puntajes asociados al subconjunto de ítems redactados positiva o negativamente. El coeficiente omega jerárquico de la subescala (ω_{hs}) es un índice que refleja la confiabilidad de la puntuación de una subescala después de controlar la varianza debida al factor general (Reise, Bonifay, & Haviland, 2013), permitiendo examinar la influencia relativa de los factores de redacción. Por ejemplo, en el caso de la subescala asociada a los ítems inversos (FI), la expresión del coeficiente omega jerárquico de la subescala FI es la siguiente:

$$\omega_{hs_{FI}} = \frac{(\sum \lambda_{i,FI})^2}{(\sum \lambda_{i,AE})^2 + (\sum \lambda_{i,FI})^2 + \sum(1 - h_i^2)} \quad (3)$$

De nuevo, $\lambda_{i,j}$ representa la carga factorial estandarizada del ítem i en el factor j y $(1 - h_i^2)$ representa la varianza de error. Por su parte, para la estimación del coeficiente omega jerárquico de la subescala asociada a los ítems directos (FD), $\omega_{hs_{FD}}$, el término $(\sum \lambda_{FI})^2$ debe ser reemplazado por $(\sum \lambda_{FD})^2$. Bajos valores de ω_{hs} indican que gran parte de la varianza confiable de las puntuaciones de la subescala puede atribuirse a variaciones en el rasgo representado por el factor general, y no a lo que es exclusivo de los factores específicos. Por esto es que, en ocasiones, es más informativo examinar el coeficiente omega jerárquico de la subescala (ω_{hs}) que el coeficiente omega de la subescala (ω_s ; no descrito aquí).

Determinación del factor y replicabilidad del constructo. Los coeficientes omega revisados anteriormente, permiten comprender la contribución relativa de las fuentes de varianza común (de factor general y los factores de redacción), y definir el valor informativo de la puntuación ponderada total y por subescala. Otros aspectos de interés para los investigadores que permiten examinar la calidad de un modelo bifactor bien ajustado, tienen que ver con: (a) la idoneidad de estimar puntuaciones factoriales para utilizarlas en análisis complementarios¹, y (b) la adecuación de especificar tanto el factor general como los factores de redacción como modelo de medición para examinar relaciones estructurales sustantivas con otras variables². Para abordar ambas posibilidades, los índices de determinación del factor (*DF*; Grice, 2001) y de replicabilidad del constructo (*H*; Hancock, 2001; Hancock y Mueller, 2001) permiten ayudar en la toma de decisiones sobre qué estrategia analítica emplear en el contexto SEM y, en última instancia, avanzar en la comprensión de la calidad del modelo de medición.

El grado de determinación de un factor está dada por la consistencia con que las puntuaciones factoriales de una variable latente (de rasgo o de método) representan adecuadamente las diferencias individuales en dicho factor. Grice (2001) señaló que, incluso en una solución factorial bien ajustada, es posible obtener, en una condición de indeterminación, un número infinito de puntuaciones factoriales consistentes con un mismo set de cargas factoriales. Esto es problemático en la medida que cualquier conjunto de puntuaciones factoriales podría producir resultados distintos e incluso contradictorios.

Para calcular el índice de determinación (*DF*) para los factores de rasgo y método, utilizamos la fórmula descrita por Beauducel (2011), en la que la matriz de correlación reproducida por el modelo se deriva de la solución de cargas estandarizadas reportada:

$$DF = \text{diag}(\Phi\Lambda'\Sigma^{-1}\Lambda\Phi)^{1/2} \quad (4)$$

¹ Hasta donde sabemos, no hay investigaciones reportadas que hayan empleado puntuaciones factoriales derivadas de los factores de redacción como variables para análisis subsecuentes. A pesar de esto, describiremos el uso e interpretación del índice de determinación del factor como herramienta que permite examinar la calidad del modelo bifactor.

² Esta estrategia ha sido extensamente empleada para examinar la naturaleza subyacente del efecto de redacción (ver por ejemplo, DiStefano & Motl, 2006; 2009a; Horan et al., 2003; Motl & DiStefano, 2002).

Donde Φ es una matriz $m \times m$ de intercorrelaciones entre los factores y m es el número de factores especificados (3 y 2 para los modelos Bifactor-FID y -FI de la RSES, respectivamente). En el caso de ambos modelos bifactor, esta matriz siempre contiene 1s en la diagonal principal y 0s en las celdas restantes. El término Λ es una matriz $k \times m$ de cargas factoriales estandarizadas donde k es el número de ítems (10 para la RSES), y Σ es una matriz $k \times k$ que contiene la matriz de correlaciones reproducida por el modelo (para una descripción detallada del cómputo de este índice, *vid.*, Beauducel, 2011). Los valores posibles de DF fluctúan entre 0 y 1, donde valores cercanos a 1 reflejan un mayor grado de determinación y que las diferentes estimaciones de puntuaciones factoriales posibles, compatibles con la estructura identificada, están altamente correlacionadas entre sí (Guttman, 1955).

Por su parte, la replicabilidad del constructo (originalmente descrita como confiabilidad del constructo; Hancock, 2001; Hancock & Mueller, 2001) permite examinar el grado en que una variable latente es adecuadamente representada por un conjunto de indicadores y, por ende, replicable entre estudios y útil para modelamiento estructural (Ferrando & Lorenzo-Seva, 2018). En el caso de los factores j del atributo o de redacción, la expresión del coeficiente H es la siguiente:

$$H_j = \frac{1}{1 + \left(\frac{1}{\sum (\lambda_{i,j}^2 / (1 - \lambda_{i,j}^2))} \right)} \quad (5)$$

El término central de la ecuación expresa la suma de las razones entre la proporción de varianza del ítem i explicada y la no explicada por parte del factor j . De este modo, un elevado número de ítems o altas cargas factoriales producirán un denominador ligeramente mayor a 1 y, con ello, un índice H cercano a 1. De hecho, los valores del índice H fluctúan entre 0 y 1, donde valores cercanos a 1 indican una variable latente bien definida con una mayor probabilidad de que sea estable entre estudios (Hancock & Mueller, 2001).

Este indicador puede ser particularmente informativo para los factores de redacción, que contienen un menor número de ítems que el factor de rasgo (en el caso de

la RSES, 5 ítems) y que por lo general arrojan bajas cargas factoriales. En este sentido, la expectativa es que estas variables latentes exhiban índices H inferiores y que, en el contexto de modelamiento estructural (como los ejemplos que mencionamos anteriormente), sus estructuras y asociaciones sean inestables y, por lo tanto, no replicables Y no interpretables.

Varianza común explicada y porcentaje de correlaciones no contaminadas.

Incluso con un modelo bifactor bien ajustado, para examinar el grado en que la solución multidimensional refleja una adecuada calidad del modelo, es necesario considerar la contribución relativa de cada fuente de varianza común (expresada a través de las cargas factoriales; Rodríguez et al., 2016b). Si el factor general es fuerte y los factores específicos están mal definidos (exhibiendo bajas cargas factoriales), un modelo de medición multidimensional en el marco de modelamiento estructural puede exhibir déficits en la replicabilidad de los resultados reportados de las asociaciones de los factores específicos. Dado lo anterior, algunos autores han propuesto distintas estrategias, como por ejemplo examinar la posibilidad de ignorar los factores específicos (especificando un modelo de un factor), en la medida en que no alteren sustantivamente las cargas factoriales del factor general (Donnellan, Ackerman, & Brecheen, 2016).

Una forma sencilla para examinar la contribución relativa de las fuentes de varianza común, tanto del factor de rasgo como de redacción, es calculando el índice de varianza común explicada (ECV; Sijtsma, 2009; Ten Berge & Sočan, 2004). Este índice permite evaluar el grado en que un modelo bifactor bien ajustado refleja una solución “esencialmente unidimensional”, estimando la proporción de varianza capturada por el factor general sobre todas las fuentes de varianza común explicada (bajo el supuesto de un modelo ortogonal). En el modelo Bifactor-FID de la RSES, la varianza común explicada por el factor de autoestima es:

$$ECV = \frac{\sum \lambda_{i,AE}^2}{\sum \lambda_{i,AE}^2 + \sum \lambda_{i,FI}^2 + \sum \lambda_{i,FD}^2} \quad (6)$$

Por su parte, en el modelo bifactor FI, el término $\sum \lambda_{i,FD}^2$ sería igual a cero dado que no hay varianza común especificada para el conjunto de ítems directos. Los valores del

índice *ECV* fluctúan entre 0 y 1 donde altos valores de *ECV* indican un mayor dominio del factor general y un bajo riesgo de sesgo entre las cargas factoriales del factor general estimadas entre un modelo bifactor y un modelo de un factor (una solución esencialmente unidimensional).

Sin embargo, como lo señalan Bonifay et al. (2015), la relación entre la magnitud del índice *ECV* y el grado de sesgo en la estimación de las cargas factoriales entre ambos modelos está moderada por la estructura general de datos, en términos de la proporción de correlaciones afectadas por la presencia conjunta de los factores generales y específicos (referida como porcentaje de correlaciones no contaminadas [*PUC*]; Bonifay et al., 2015; Reise et al., 2013).

Para comprender este indicador y su influencia sobre el riesgo de sesgo en la estimación de parámetros, consideremos la RSES como ejemplo. Para calcular el valor de *PUC*, es necesario identificar, en primer lugar, el número de correlaciones (*ncor*) contenidas en cada factor *j* especificado en el modelo bifactor:

$$ncor_j = \frac{k_j(k_j - 1)}{2} \quad (7)$$

Donde *k* es el número de ítems especificados para el factor *j* (10 para el factor de autoestima y 5 para cada factor de redacción). Esto implica un total de 45 correlaciones entre los ítems de la RSES. Dentro de cada factor de redacción, el número de correlaciones (10 en cada factor) representan asociaciones “contaminadas” por multidimensionalidad atribuida tanto al factor general como al respectivo factor de redacción. A partir de estos valores es posible calcular el índice *PUC*:

$$PUC = \frac{ncor_{AE} - (ncor_{FI} + ncor_{FD})}{ncor_{AE}} \quad (8)$$

En el caso del modelo Bifactor-FI, el valor de *PUC* es $(45 - [10 + 0])/45 = 0,778$ y en el modelo Bifactor-FID es igual a $(45 - [10 + 10])/45 = 0,556$. Ambos valores reflejan que la mayoría de las correlaciones informan estrictamente sobre el factor general de autoestima, que representa el atributo objetivo para el cual el instrumento fue diseñado para evaluar. Bajo estas condiciones, es esperable que haya pequeñas diferencias entre las

cargas factoriales entre un modelo unidimensional y el factor general en un modelo bifactor. Reise et al. (2013) y Bonifay et al. (2015) mostraron que con mayores valores de *PUC* observados, la magnitud del índice *ECV* se vuelve menos importante para determinar, cuando se especifica un modelo unidimensional a datos multidimensionales con una estructura bifactor, el grado de sesgo tanto de las cargas factoriales del factor del atributo así como su relación con variables criterio.

El presente estudio

Los estudios sobre los efectos de redacción utilizando el modelo bifactor con la RSES han abarcado una amplia variedad de muestras, edades y países. La estructura bifactor propuesta parece ser robusta y exhibe excelentes índices de ajuste. Los trabajos de revisión y meta-análisis publicados recientemente (Gnambs et al., 2018; Jak & Cheung, 2020) han confirmado el modelo Bifactor-FID como la solución factorial con mejor ajuste estadístico. Sin embargo, estos estudios se han enfocado en agrupar estudios para identificar el mejor modelo de medición sin examinar en detalle las características de las variables latentes asociadas a los efectos de redacción. La interpretación adecuada del modelo puede realizarse a través de una serie de índices que permiten evaluar los méritos del modelo bifactor estimado, así como de cada factor independientemente.

Por lo tanto, el presente estudio tuvo como propósito examinar, a través de la revisión de estudios previos y la estimación de indicadores, el grado en que los factores de atributo y método de la RSES son psicométricamente interpretables, y el grado en que los factores de atributo y método de la RSES están lo suficientemente bien definidos para servir en el contexto del modelamiento de ecuaciones estructurales como predictores o criterios. Utilizamos los modelos bifactor estimados con la escala de autoestima de Rosenberg (1965) para inspeccionar el valor de este tipo de especificación para controlar la presencia de efectos de redacción y examinar su naturaleza sustantiva o artefactual. Para ello, identificamos 60 estudios reportados previamente y 11 archivos de datos públicos que emplearon la RSES y calculamos una serie de índices estadísticos que permiten examinar: (a) la calidad de las puntuaciones totales y por subescala (coeficientes omega), (b) la idoneidad, en el contexto de modelamiento estructural, de los factores de rasgo y

método en función de los ítems que contienen (utilizando los índices *DF* y *H*), y (c) el grado de dependencia del conjunto de datos a la estructura bifactor y el grado de unidimensionalidad subyacente que justifique, con el menor margen de sesgo, la posibilidad de estimar un modelo unidimensional (a través de los índices *ECV* y *PUC*). Se espera que el hacerlo proporcione información adicional sobre las fortalezas y debilidades del modelo bifactor para abordar diferentes preguntas y aplicaciones de investigación.

Método

Este estudio contó con la aprobación del CEC de la Pontificia Universidad Católica de Chile (ID 180903005) y fue llevada a cabo siguiendo los lineamientos propuestos en el protocolo PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis; Page et al., 2021). La lista de verificación PRISMA puede encontrarse en el material suplementario (ver tablas S1 y S2). El protocolo fue también registrado en el Open Science Framework en octubre 2021 (<https://osf.io/96gvn>).

Estrategia de búsqueda y fuentes de información

La búsqueda de estudios elegibles fue llevada a cabo por dos autores (FPP y FG) e involucró múltiples pasos para descartar registros duplicados, estudios no relacionados o estudios que cayeron en los criterios de exclusión (ver Figura 2).

[Insertar Figura 2 aproximadamente aquí]

En primer lugar, se llevó a cabo una búsqueda de estudios utilizando Web of Science (Clarivate Analytics), Scopus (Elsevier), PsycArticles (American Psychological Association) y Google Scholar (Google LLC), bajo los siguientes criterios de búsqueda: a) [ALL (Rosenberg AND self-esteem) AND ALL (bifactor)], b) [ALL (Rosenberg AND self-esteem) AND ALL ("wording effects")], y c) [ALL (Rosenberg AND self-esteem) AND ALL ("method effects")]. La fecha límite para la búsqueda fue el 10 de octubre de 2021 y no se aplicó ninguna restricción de idioma. En segundo lugar, con el propósito de incrementar la validez externa de los resultados, también buscamos en repositorios de archivos de datos públicos (Inter-university Consortium for Political and Social Research [ICPSR], Harvard Dataverse, UK Data Service) utilizando como criterio de búsqueda el

término clave “Rosenberg self-esteem” o el contenido de los ítems (*e.g.*, “I feel that I have a number of good qualities”).

Criterios de elegibilidad

Se incluyeron estudios que examinaron los efectos de redacción de la escala RSES empleando modelamiento bifactor (Bifactor-FI, Bifactor-FID o ambos) o aquellos que compartieron sus datos de la escala. No aplicamos restricciones asociadas al idioma o estado del reporte (*i.e.*, manuscritos no publicados eran elegibles para la inclusión).

Por otra parte, fueron excluidos aquellos estudios que: (1) No reportaron las cargas factoriales estandarizadas o las cargas reportadas contienen errores (*e.g.*, cargas estandarizadas cuyos suma de cuadrados por ítem exceden 1; Syropoulou et al., 2021), (2) no emplearon los 10 ítems de la RSES (*e.g.*, Motl & DiStefano, 2002), (3) No estimaron un modelo bifactor ortogonal (*e.g.*, Quilty et al., 2006), o (4) emplearon un tamaño muestral inferior a 200.

Selección de registros y recolección de datos

El proceso de selección de los registros y recolección de datos fue llevado a cabo por dos de los investigadores (FPP y FG). No se emplearon herramientas de automatización en las etapas de revisión, selección y recolección de información.

En primer lugar, los artículos duplicados fueron manualmente removidos empleando identificadores permanentes (*i.e.*, DOI, URI, el título del registro) como criterio de duplicidad. Luego de remover duplicidades, se revisaron los títulos, resúmenes y las tablas de resultados presentadas en cada uno de los registros identificados para evaluar su elegibilidad. A continuación, se examinó el contenido completo de los artículos seleccionados para determinar su inclusión o exclusión. En el caso de las base de datos, luego de descartar aquellas cuyos datos no se encontraban accesibles, se revisaron los libros de códigos para identificar el número de ítems de la RSES aplicados, el número de categorías empleadas y el tamaño muestral recolectado para determinar su inclusión o exclusión.

Una vez identificados los registros que cumplieron con los criterios de inclusión (ver Figura 2), se recolectaron las cargas factoriales estandarizadas reportadas

considerando todas las posibles condiciones multigrupo examinadas en cada registro (*e.g.*, sexo, ola). Por su parte, con los set de datos identificados (11 archivos de datos públicos), se estimaron tres modelos CFA (un factor, Bifactor-FI y Bifactor-FID) y se recolectaron las cargas factoriales estandarizadas (ver Material Suplementario para una descripción detallada del procedimiento de estimación utilizado). Junto con ello, se recopiló la información descriptiva sobre las características de la muestra (número de participantes, país, edad, sexo), la versión de la RSES aplicada (idioma, distribución de ítems, opciones de respuesta) y el procedimiento analítico aplicado (método de estimación, índices de ajuste examinados).

Riesgo de sesgo de los estudios incluidos

Se evaluó el riesgo de sesgo en los estudios incluidos para identificar potenciales debilidades del diseño de los estudios incluidos que pudiera influir en los resultados obtenidos. Para cada estudio, un investigador (FPP) revisó y evaluó: (a) las estrategias de selección y muestreo empleadas, (b) el número de categorías utilizadas, la distribución multivariada de datos y el método de estimación empleado (ver, Rhemtulla, Brosseau-Liard, & Savalei, 2012), (c) el grado de consistencia interna de la RSES reportado en el estudio, y (d) el ajuste estadístico obtenido de los modelos bifactor estimados. En cada uno de estos criterios examinados, se concluyó el grado en que el sesgo pudiera alterar los resultados (bajo riesgo, riesgo de sesgo poco claro, y alto riesgo; ver Higgins et al., 2011).

Medidas y síntesis de datos

A partir de las cargas factoriales recolectadas en cada uno de los estudios identificados o bases de datos examinadas, calculamos los índices derivados del modelo bifactor utilizando el paquete de R *BifactorIndicesCalculator* (Dueber, 2021).

Para examinar la contribución relativa de los factores de rasgo y redacción a la varianza común tanto en el conjunto total de ítems como en los subsets de ítems redactados positiva o negativamente, evaluamos la confiabilidad basada en el modelo mediante el cálculo de los coeficientes omega (ω , ω_s) y omega jerárquico (ω_h , ω_{hs}) de los factores especificados en los modelos Bifactor-FI y -FID. A medida que el valor de ω_h se aproxima a 1, el factor general de rasgo controla la varianza asociada a la puntuación total del

instrumento (reflejada en el valor de ω) y, por lo tanto, la influencia de la multidimensionalidad expresada a través de los factores específicos será menor. Por su parte, cuanto mayor sea el valor de ω_{hs} mayor será la influencia de los factores específicos sobre la varianza asociada al subconjunto específico de ítems. Se ha sugerido que valores de omega jerárquico mayores a 0,70 son adecuados para una interpretación sustantiva del factor de rasgo o de redacción (Canivez, 2015; Rodríguez et al., 2016a).

Con respecto a la utilidad de los factores especificados en los modelos bifactor examinados, tanto como puntuaciones factoriales estables o como variables latentes replicables, calculamos el índice de determinación (DF) y el índice H , respectivamente. En el caso de los valores de DF , los autores recomiendan considerar las puntuaciones factoriales como aceptables solo si el índice de determinación del factor es igual o mayor a 0,90 (Gorsuch, 1983; Grice, 2001; Rodríguez et al., 2016a). Por su parte, cuando el valor de H es bajo, la variable latente no está bien definida por sus indicadores y, por lo tanto, se espera que varíe entre estudios, debilitando su calidad como predictor/criterio (Arias et al., 2018). Hancock y Mueller (2001) sugirieron que un valor de H mayor a 0,70 implica una baja probabilidad de fluctuación de la magnitud de las cargas factoriales y una mayor estabilidad tanto del factor como de su relación con otras variables criterio.

Finalmente, calculamos la varianza común explicada (ECV) para examinar, en interacción con los valores de PUC de los modelos Bifactor-FI y Bifactor-FID, el grado en que ambos modelos estimados reflejan una solución esencialmente unidimensional. Valores de ECV cercanos a 1 sugieren que, si estimamos un modelo unidimensional, las cargas factoriales serán similares a las estimadas con el factor general del modelo bifactor (Arias, Ponce, & Núñez, 2018; Rodríguez et al., 2016a), así como la asociación con otras variables criterio en un modelo SEM (Bonifay, Reise, Scheines, & Meijer, 2015; Reise, Scheines et al., 2013). Diversos autores han propuesto puntos de corte para interpretar los valores de ECV del factor general como indicativos de una solución “esencialmente unidimensional” (e.g., Quinn, 2014; Rodríguez et al., 2016b; Stucky et al., 2014). Aquí, optamos por, siguiendo a Bonifay et al. (2015), examinar los valores de ECV en interacción con los de PUC , empleando los valores del sesgo de parámetro relativo

promedio como criterio para analizar el grado en que los parámetros obtenidos con una solución multidimensional varían sustantivamente al forzarlos a una estructura unidimensional.

Resultados

Estudios incluidos

El proceso de revisión identificó 3576 artículos (1832 luego de remover registros duplicados o no recuperados) que fueron examinados para elegibilidad. Muchos registros fueron excluidos porque simplemente empleaban la RSES como medida de validez de criterio de otra escala, dejando 134 artículos de texto completo potencialmente elegibles. De ellos, 74 fueron descartados por las razones listadas en la Figura 2, dejando 60 estudios para su inclusión en el análisis final. Por su parte, se identificaron 95 archivos de datos públicos (55 luego de descartar registros no recuperados). De estos registros, 44 fueron descartados por emplear un menor número de ítems o insuficiente tamaño muestral, dejando 11 registros que fueron considerados para el análisis. Las características de los 71 estudios incluidos, sus subgrupos y los modelos estimados se resumen en la Tabla 1.

[Insertar Tabla 1 aproximadamente aquí]

Riesgos de sesgo

Los resultados de la evaluación de la calidad de los estudios se muestran en la Figura 3. El riesgo de sesgo en los resultados como producto de la estrategia de muestreo fue el que exhibió una mayor proporción de estudios con niveles altos o inciertos de sesgo (48,21%) entre los criterios considerados en la presente revisión. Esto se debe al uso extendido, en este tipo de estudios, de muestras por conveniencia o por disponibilidad (*e.g.*, estudiantes universitarios). Por su parte, los criterios asociados a la calidad de los datos y el método de estimación, así como el grado de consistencia interna reportado arrojaron una proporción de estudios con bajo riesgo de sesgo cercana al 90%, mientras que la proporción de estudios con bajo nivel de riesgo del criterio asociado al ajuste estadístico de los modelos bifactor fue superior al 70%. La Figura Suplementaria 1 muestra un desglose del nivel de riesgo por estudios y criterios de sesgo. Una lista

completa de los estudios con las cargas factoriales extraídas o estimadas, junto con los valores de los indicadores calculados puede ser encontrada en la Tabla Suplementaria 3.

Calidad de las puntuaciones totales y por subescala

La Figura 4 muestra la distribución conjunta de los valores de omega y omega jerárquico de los factores asociados al atributo (AE) y el método (FI) del modelo Bifactor-FI. En el caso del factor general, los valores de omega (Me=0,889, Mdn=0,895, DE=0,037) y omega jerárquico (Me=0,793, Mdn=0,805, DE=0,052) sugieren una predominancia del factor general en la proporción de varianza asociada al puntaje total. Consistente con esto, la razón entre los valores de omega y omega jerárquico (ω/ω_h ; Me=0,891, Mdn=0,894, DE=0,037) indican que, en promedio, cerca del 90% de la varianza confiable en las puntuaciones compuestas de la RSES es atribuible al factor general. Del mismo modo, en el caso del factor de redacción, los valores de omega jerárquico (Me=0,313, Mdn=0,312, DE=0,093) y la razón ω/ω_h (Me=0,382, Mdn=0,383, DE=0,193) confirman una mayor influencia del factor general, incluso dentro del subconjunto específico de ítems redactados negativamente.

[Insertar Figura 4 aproximadamente aquí]

Esta preponderancia del factor general de autoestima sobre las puntuaciones globales y por subescala fue confirmada con los valores de omega y omega jerárquico calculados en los modelos Bifactor-FID (Figura 5). Pese a la inclusión de un factor específico adicional para el subconjunto de ítems directos (FD), los valores de omega jerárquico (Me=0,781, Mdn=0,789, DE=0,029) y la razón ω/ω_h (Me=0,865, Mdn=0,871, DE=0,053) del factor general fueron similares a los obtenidos empleando el modelo Bifactor-FI, mientras que los valores de omega jerárquico tanto del factor FI (Me=0,286, Mdn=0,300, DE=0,113) y FD (Me=0,104, Mdn=0,083, DE=0,092) fueron notoriamente inferiores.

[Insertar Figura 5 aproximadamente aquí]

Idoneidad de los factores de atributo y redacción en el contexto SEM

Con respecto al grado en que los factores de rasgo y de método son adecuadamente representados por sus respectivos indicadores, la Figura 6 muestra la distribución de los

valores del índice H para ambos factores, calculados con los modelos Bifactor-FI y -FID examinados. En el caso del factor general, los valores del índice H tanto en el modelo Bifactor-FI (Me=0,880, Mdn=0,882, DE=0,038) como FID (Me=0,884, Mdn=0,885, DE=0,036) fueron superiores a los observados con los factores específicos, siendo en todos los casos mayores al punto de corte (0,7; Hancock & Mueller, 2001). Esto conlleva una baja probabilidad de variación en la magnitud de las cargas factoriales, así como una alta estabilidad del factor y, por ende, con buena replicabilidad entre estudios y útil para modelamiento estructural. Por su parte, los valores de los factores de redacción de los modelos Bifactor-FI (FI; Me=0,586, Mdn=0,586, DE=0,081) y Bifactor-FID (FI, Me=0,564, Mdn=0,572, DE=0,098; FD, Me=0,421, Mdn=0,427, DE=0,117) fueron considerablemente inferiores, exhibiendo una mayor dispersión, y una alta proporción de valores menores al punto de corte. Esto implica que tanto los factores de redacción como la relación de estos con otras variables criterio son mayormente inestables, lo cual dificulta su adecuado uso e interpretabilidad como variables criterio asociadas al efecto de redacción.

[Insertar Figura 6 aproximadamente aquí]

Del mismo modo, los valores de índice de determinación obtenidos con ambos modelos (Figura 7) confirman la tendencia de que los factores específicos de redacción del modelo Bifactor-FI (FI, Me=0,800, Mdn=0,802, DE=0,061) y Bifactor-FID (FI, Me=0,794, Mdn=0,791, DE=0,068; FD, Me=0,731, Mdn=0,722, DE=0,093) se encuentran pobremente representados por sus respectivos indicadores. De hecho, menos del 5% de los modelos examinados arrojaron valores de DF iguales o superiores al punto de corte (0,9; Gorsuch, 1983; Grice, 2001; Rodríguez et al., 2016a). Esto conlleva a que los factores específicos producen puntuaciones factoriales susceptibles de arrojar resultados inestables y no interpretables (incluso contradictorios). Por el contrario, el factor general exhibió excelentes valores con el modelo Bifactor-FI (Me=0,929, Mdn=0,932, DE=0,023) y con el modelo Bifactor-FID (Me=0,930, Mdn=0,932, DE=0,025).

[Insertar Figura 7 aproximadamente aquí]

Unidimensionalidad y sesgo de los parámetros

El panel izquierdo de la Figura 8 muestra la distribución de los valores de varianza común explicada por el factor general y los factores de redacción de los modelos examinados. Para el modelo Bifactor-FI, la proporción de varianza común explicada por el factor general (Me=0,793, Mdn=0,797, DE=0,049) fue notablemente superior en comparación a la explicada por el factor de redacción (Me=0,207, Mdn=0,203, DE=0,049). En efecto, menos del 4% de los modelos examinados exhibieron valores de *ECV* inferiores a 0,7. Esta tendencia también se encontró al comparar la proporción de varianza asociada al factor general del modelo Bifactor-FID (Me=0,716, Mdn=0,713, DE=0,059) y la de ambos factores de redacción (FI, Me=0,177, Mdn=0,179, DE=0,051; FD, Me=0,108, Mdn=0,101, DE=0,042). En ambos casos, los resultados indican una predominancia del factor general de autoestima en la proporción de varianza explicada, sugiriendo que el modelo en lugar de ser estrictamente unidimensional o multidimensional podría ser considerado como “esencialmente unidimensional”.

[Insertar Figura 8 aproximadamente aquí]

Para examinar en detalle esta posibilidad, calculamos el sesgo de parámetro relativo promedio (*ARPB*; Reise et al., 2013) como índice general de la diferencia entre las cargas factoriales estimadas en el factor general de los modelos bifactor y las cargas factoriales de un modelo unidimensional³. Este coeficiente se calcula promediando los valores absolutos de la diferencia por ítem entre la carga factorial del modelo unidimensional y bifactor dividida por la carga factorial del factor general del modelo bifactor:

$$ARPB = \frac{\sum \left| \frac{\lambda_{i,U} - \lambda_{i,B}}{\lambda_{i,B}} \right|}{k} \quad (9)$$

Donde $\lambda_{i,U}$ es la carga factorial del ítem i en el factor del modelo unidimensional, $\lambda_{i,B}$, es la carga factorial del ítem i en el factor general del modelo bifactor, y k es el número de ítems. Bajos valores de *ARPB* (<15%; Muthén et al., 1987; Reise et al., 2013) indican

³ Sólo fueron considerados aquellos estudios que reportaron las cargas factoriales estandarizadas del modelo de un-factor (110 modelos Bifactor-FI y 102 modelos Bifactor-FID; ver Tabla TS3).

un bajo nivel de sesgo de los parámetros si los ítems son forzados a una estructura unidimensional (el factor general del modelo bifactor y el factor único de la solución unidimensional representan variables latentes aproximadamente equivalentes). Por lo tanto, se puede utilizar el modelo unidimensional en contextos SEM sin riesgo de sesgar sustancialmente los coeficientes de estructura (Bonifay et al., 2015; Rodríguez et al., 2016a).

El panel derecho de la Figura 8 muestra los valores de *ARPB* calculados con los modelos Bifactor-FID y -FI. En ambos casos, el 70% de los casos examinados arrojaron un nivel marginal o nulo de riesgo de sesgo confirmando tanto la idea del modelo esencialmente unidimensional como la potencial intrascendencia del efecto de redacción, modelado a través de factores específicos en escenarios de modelamiento estructural.

Discusión

Los modelos bifactor proporcionan una forma conceptual útil para examinar la presencia de múltiples fuentes de varianza en una escala de evaluación, especialmente cuando se sospecha la presencia de varianza sistemática atribuible al método de medición que impide la medición apropiada del constructo de interés. Complementado revisiones previas sobre la estructura factorial óptima de la RSES (Gnambs et al., 2018), que concluyeron que el modelo bifactor exhibía el mejor ajuste, la presente revisión sistemática ilustró el valor de una serie de indicadores derivados del marco analítico factorial del modelo bifactor para proporcionar una adecuada examinación de las propiedades psicométricas de la medida y de la calidad del modelo, empleando la RSES como modelo de investigación.

Esto último es crucial considerando el hecho de que, si bien es evidente que los factores de redacción son factores residualizados (*i.e.*, representan la covarianza que permanece después de dar cuenta del factor general ortogonal), toda la evidencia previamente reportada acerca de la naturaleza sustantiva o efímera del efecto de redacción ha sido propuesta a través de la estimación de redes nomológicas para examinar la validez relacionada a criterio (Cronbach & Mehl, 1955), utilizando el factor de redacción como

una entidad distintiva del efecto de redacción (DiStefano & Motl, 2006, 2009a,b; Horan et al., 2003; Quilty et al., 2006).

Finalmente, la baja contribución de los factores de redacción para dar cuenta de la multidimensionalidad en el modelo fue confirmada tras encontrar una mínima variación o sesgo de los parámetros observados con el modelo de un factor respecto de los obtenidos con el modelo bifactor lo cual, junto con la magnitud de varianza común explicada, respalda la idea de una solución esencialmente unidimensional donde las variables latentes subyacentes a los factores de redacción no son sustantivamente relevantes. Estos resultados son consistentes con aquellos que reportan una ausencia de diferencias entre las puntuaciones estimadas empleando una aproximación unidimensional y modelos bifactor, así como la relación del factor de rasgo, extraído bajo ambas estrategias, y diversas variables criterio (Donnellan, Ackerman, & Brecheen, 2016; Michaelides, Koutsogiorgi, & Panayiotou, 2017).

Factores de redacción como expresión del efecto de redacción

Frente a esta situación, la inquietud más relevante refiere al grado con que los factores específicos de redacción reflejan el efecto de redacción. Nuestros hallazgos nos llevan a concluir que la baja replicabilidad y estabilidad de los parámetros estimados en cada una de las soluciones factoriales examinadas indican una baja probabilidad de que estos factores reflejen adecuadamente diferencias individuales en la expresión de este tipo de sesgos.

Esto no es trivial ya que las correlaciones de estos factores residualizados han sido interpretadas unívocamente como evidencia de asociación con variables criterio, estabilidad del efecto entre escalas, o a lo largo de mediciones sucesivas, llevando a diversos autores a proponer el efecto de redacción como un estilo de respuesta (Horan et al., 2003). Sin embargo, las asociaciones reportadas del factor de redacción con variables criterio u otros factores de redacción muestran tamaños de efecto débiles (*e.g.*, DiStefano & Motl, 2006; 2009a; Donnellan et al., 2016; Lindwall et al., 2012; Quilty et al., 2006) y exhiben inconsistencias entre escalas (*e.g.*, DiStefano & Motl, 2006; Rauch et al., 2007). Kam (2018), para explicar esta variabilidad en los resultados, propuso que el efecto de

redacción sería “específico de la escala”, sugiriendo que la relación encontrada de un factor de redacción con una variable criterio dada “se aplica solo a una medida particular involucrada en una investigación” (p.574). No obstante, esta noción no explica asociaciones inconsistentes reportadas utilizando el factor de redacción de una misma escala (e.g., Alessandri et al., 2015; Donnellan et al., 2016; Lindwall et al., 2012).

Por el contrario, basados en nuestros hallazgos, creemos que los factores de redacción representan variables latentes pobremente definidas, extraídas de una solución factorial propensa al sobreajuste estadístico (Bonifay & Cai, 2017) y, con ello, no representan un dominio sustantivo ni reflejan diferencias individuales en la expresión del efecto de redacción. A partir de esta definición, las asociaciones del factor de redacción con variables latentes criterio pueden interpretarse directamente como espurias y que, como lo sugirieron DiStefano y Motl (2006), siguen la dirección del factor de rasgo⁴. Al mismo tiempo, la inconsistencia en los hallazgos reportados son el resultado de la inestabilidad del factor (reflejado en los bajos valores de los índices de determinación y replicabilidad observados).

Por otra parte, Dueber y Toland (2021) recientemente publicaron un estudio de simulación para examinar la relación de los índices estimados asociados a las propiedades de los factores específicos (ω_s , ω_{hs} , ECV) y el grado en que éstos aportan información adicional sustancial por sobre el factor general (expresado en valores de $VAR \geq 1.1$; ver Feinberg & Jurich, 2017, citados en Dueber & Toland, 2021), con el propósito de establecer condiciones de idoneidad para interpretar las subpuntuaciones derivadas del modelo bifactor. Sin embargo, los autores proponen esta solución en escenarios donde los factores específicos representan factores sustantivos del atributo de interés y se espera que proporcionen más información. Los factores de redacción, por su parte, corresponden a factores espurios desde un punto de vista sustantivo, cuyos valores de ECV son tan bajos

⁴ DiStefano y Motl (2006) sugirieron esto para explicar la inesperada asociación negativa entre los factores de método de la escala de autoestima de Rosenberg y la de ansiedad física social ($r = -0.37$; ver Tabla 3, p.456). Para ilustrar este punto, el lector podría examinar la Tabla 3, publicada por Kam (2018), que contiene tanto las correlaciones entre los factores de rasgos como entre los factores de redacción. La correlación entre ambos subsets fue de 0,50.

(en nuestro caso $Mdn \leq 0.20$; ver Figura 8) que, de acuerdo a los resultados reportados por Dueber y Toland (2021), en un escenario de pocos factores específicos (Bifactor-FI y Bifactor-FID) se requerirían altos valores de ω_s para considerar una interpretabilidad sustantiva de sus puntuaciones.

Estrategias de modelamiento alternativas

Si expresamos la respuesta de un participante p a un ítem i como y_{ip} , el modelo factorial clásico unidimensional puede ser escrito como:

$$y_{ip} = \mu_i + \lambda_i \eta_p + \varepsilon_{ip} \quad (10)$$

Donde μ_i es el intercepto del ítem para el ítem i , λ_i denota el vector de la carga factorial del ítem i , η_p es el vector del participante sobre el factor común, y ε_{ip} es el término de error para el participante p en el ítem i . En esta ecuación tanto el factor común como el término de error son variables aleatorias que varían entre participantes, mientras que el intercepto y la carga factorial son coeficientes fijos que no cambian su valor para estimar la función de respuesta de los individuos.

En un modelo así, se espera que la respuesta de dos individuos con el mismo nivel de rasgo latente sea igual. Sin embargo, pueden exhibir diferencias debido al término de error individual. Este término de error incluye el error de medición y las diferencias individuales sistemáticas en factores no modelados. La noción central del modelamiento del efecto de redacción bajo la perspectiva bifactor reside en incorporar factores comunes adicionales asociados a diferencias sistemáticas individuales en el set de ítems directos, inversos o ambos simultáneamente. Sin embargo, considerando la baja magnitud de las cargas factoriales de los factores específicos encontrada en esta revisión, parece evidente que su contribución para la interpretación de las puntuaciones esperadas es baja y, más importante, su adecuada interpretabilidad es cuestionable. Diversos autores han señalado esta preocupación (Bonifay et al, 2017; Geiser et al., 2008; Maydeu-Olivares & Coffman, 2006). Esta falta de interpretabilidad ha llevado a diversos autores a proponer estrategias alternativas para modelar el efecto de redacción.

El análisis factorial de ítems con intercepto aleatorio (en inglés RIIFA; Billiet & McClendon, 2000; Maydeu-Olivares & Coffman, 2006), introduce un factor común con

cargas factoriales fijas para representar un método que es común entre los ítems, y libera los interceptos para permitir las diferencias individuales en el uso de la escala de respuesta (panel izquierdo Figura 9). El modelo puede ser escrito como:

$$y_{ip} = \mu_i + \lambda_i \eta_p + (\varepsilon_{ip} + \zeta_p) \quad (11)$$

Donde el nuevo término, ζ_p , es una variable aleatoria que varía entre los encuestados pero que es común para todos los ítems. En otras palabras, en el ejemplo anterior, para ambos individuos con el mismo nivel de rasgo cuyas puntuaciones pueden diferir debido a efectos asociados al uso de la escala de respuesta, dicho efecto es capturado por el intercepto aleatorio. Este modelo ha mostrado ser una aproximación prometedora para modelar la varianza de método atribuida a efectos de redacción (Savalei & Falk, 2014).

Alternativamente, bajo la idea de los efectos de redacción como resultado de la presencia de un subconjunto específico de individuos, diversos autores han examinado el efecto de redacción empleando modelamiento de mixturas (Arias et al., 2020; Ponce et al., 2021; Steinmann, Strietholt, & Braeken, 2021). Estos autores, utilizando modelamiento de mixtura de factores (en inglés FMM; Lubke & Muthén, 2005; Clark et al., 2013), distinguen entre clases de respondedores con respuestas consistentes e inconsistentes a los ítems directos e inversos y aplican especificaciones para cada clase⁵. Por ejemplo, el modelo propuesto por Arias et al. (2020) es un modelo FMM restringido, basado en el modelo RIIFA, en el que se especifica un patrón de cargas factoriales para reflejar el patrón simétrico y asimétrico de respuestas a ítems directos e inversos en las clases consistente e inconsistente, respectivamente (ver panel derecho Figura 9). A su vez, los interceptos son fijados para que sean iguales en ambas clases ($v_{k=1,i} = v_{k=2,i}$) para enfocar las diferencias entre las clases a la dirección de las cargas del ítem, y las medias de los factores se fijan en cero para ambas clases ($\alpha_{k=1} = \alpha_{k=2} = 0$), lo que indica que los efectos de la redacción no están relacionados con los niveles de los rasgos en el factor sustantivo.

⁵ Para el lector interesado, García et al. (2021) provee una descripción y comparación de las especificaciones de los modelos FMM restringidos de Arias et al (2020) y Steinmann, Strietholt, & Braeken (2021).

Estas técnicas han mostrado detectar exitosamente encuestados con respuestas inconsistentes y mejorar, tras remover a los sujetos identificados, las propiedades métricas de numerosas escalas de personalidad y habilidades cognitivas (Arias et al., 2020; García et al., 2021; Steinmann, Strietholt, & Braeken, 2021; Steinman, Sánchez, et al., 2021).

Limitaciones y direcciones para investigaciones futuras

Aunque esta revisión sistemática proporcionó información valiosa, se deben reconocer algunas limitaciones. Una limitación general de la presente revisión sistemática, al igual que trabajos similares, se refiere a la calidad general de los estudios incluidos y potenciales sesgos. En primer lugar, aunque nuestra estrategia intentó ser lo más exhaustiva e insesgada posible, incorporando investigaciones no publicadas, investigación con datos abiertos de la RSES (independiente de su objetivo de investigación) o repositorios de archivos de datos, no contempló otro tipo de literatura gris. Sin embargo, no existe consenso sobre el impacto de la literatura gris en los hallazgos de revisiones sistemáticas (McAuley, et al., 2000; Schmucker et al., 2017).

En segundo lugar, pese a nuestros esfuerzos por construir una revisión lo más amplia y representativa posible, la mayoría de los estudios revisados utilizaron participantes de muestras no-clínicas, y de sociedades WEIRD (del inglés, Western, educated, industrialized, rich and democratic; Henrich et al., 2010a). Desafortunadamente, esto refleja una limitación común en diversas áreas de investigación en las ciencias de la conducta. Aunque Schmitt y Allik (2005), luego de recolectar datos de la RSES en 53 países, mostraron diferencias generalizadas entre los ítems directos e inversos, también encontraron interdependencia cultural al exhibir este efecto. Es por ello que sugerimos tomar nuestros resultados con cautela y alentamos la investigación con muestras no-WEIRD para contrarrestar esta tendencia y evitar la sobregeneralización de ciertas afirmaciones sobre la psicología y el comportamiento humano (Henrich et al. 2010b).

En tercer lugar, debido a los criterios de exclusión empleados, muchos de los estudios más citados que examinan la naturaleza subyacente del efecto de redacción de la RSES (Alessandri et al., 2013; 2015; Horan et al., 2003; Lindwall et al., 2012; Marsh et al., 2010; Motl & DiStefano; Quilty et al., 2006) fueron descartados debido a que los autores utilizaron versiones reducidas de la escala o incorporaron especificaciones al

modelo ortogonal (*e.g.*, factores de redacción correlacionados). Esto último no es trivial considerando que algunos de estos arreglos introducen dificultades adicionales para la adecuada interpretación de los factores específicos.

Finalmente, la inspiración inicial para llevar a cabo esta revisión fue examinar el valor de los factores de redacción de la escala de Rosenberg, considerando el uso que se les ha dado tanto a la escala como a estos factores específicos para investigar diversos fenómenos asociados al proceso de respuesta a ítems y efectos de método. Sin embargo, consideramos que la investigación en esta área requiere algo más que “el modelo más libre de errores de interpretación posible” o “la técnica con la mejor sensibilidad y especificidad para detectar individuos que responden inconsistentemente a ítems inversos”, sino que también se necesita proponer estrategias de investigación que permitan ahondar en los procesos de respuesta desde un punto de vista de los procesos cognitivos subyacentes (Tourangeau et al., 2000; Tourangeau, 2017), utilizando procedimientos experimentales o cuasi-experimentales para examinar modificaciones en las condiciones de aplicación de encuestas, características de los participantes o del instrumento. En este sentido, la neurociencia cognitiva-conductual desempeñará progresivamente un papel igual de importante que la estadística.

Información adicional

Disponibilidad de datos. Todos los datos generados o analizados en este estudio se incluyen en el artículo publicado y en el material suplementario. El archivo de datos se encuentra disponible en el Open Science Framework (<https://osf.io/j3e9z/>).

Financiamiento. Este trabajo fue apoyado por la Agencia Nacional de Investigación y Desarrollo (ANID, N.21161101) otorgado a Fernando P. Ponce

Rol de los patrocinadores. Ninguno de los patrocinadores de esta investigación tuvo algún rol en el diseño y ejecución del estudio.

Agradecimientos: En este artículo utilizamos los datos del Longitudinal Internet Studies for the Social Sciences panel administrado por el CentERdata, la National Longitudinal Survey of Youth (NLSY79) y National Longitudinal Survey of Youth for Childs and Young adults (NLSY79 Child/YA) administrados por la US Bureau of Labor Statistics (US Department of Labor), the Eugene-Springfield Community Sample from the Oregon Research Institute, the Aging and Autism Study from the Institute for Autism Research, the National Survey of Wales administrados por el Welsh Assembly Government, la German Ageing Survey del Deutsches Zentrum für Altersfragen, el Longitudinal Study of Generations administrado por el US National Archive of Computerized Data on Aging, y el Family and Children Study (FACS) manejado por el UK Department for Work and Pensions.

Referencias

- Arias, V.B., Garrido, L.E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, 52, 2489-2505.
<https://doi.org/10.3758/s13428-020-01401-8>
- Arias, V.B., Ponce, F.P., Núñez, D.E. (2018). Bifactor models of Attention-Deficit/Hyperactivity Disorder (ADHD): An evaluation of three necessary but underused psychometric indexes. *Assessment*, 25(7), 885-897.
<https://doi.org/10.1177/1073191116679260>
- *Alessandri, G., Cenciotti, R., Laguna, M., Różycka-Tran, J., & Vecchione, M. (2017). Individual-level and culture-level self-esteem: A test of construct isomorphism. *Journal of Cross-Cultural Psychology*, 48(9) 1328-1341.
<https://doi.org/10.1177/0022022117722633>
- Alessandri, G., Vecchione, M., Eisenberg, N., & Laguna, M. (2015). On the factor structure of the Rosenberg (1965) General Self-Esteem Scale. *Psychological Assessment*, 27(2), 621-536. <https://doi.org/10.1037/pas0000073>
- Alessandri, G., Vecchione, M., Donnellan, B.M., & Tisak, J. (2013). An application of the LC-LSTM framework to the self-esteem instability case. *Psychometrika*, 78(4), 769-792. <https://doi.org/10.1007/s11336-013-9326-4>
- *Bákker, R., Sheferaw, E.D., Stekelenburg, J., Yigzaw, T., de Kroon, M.L.A. (2020). Development and use of a scale to assess gender differences in appraisal of mistreatment during childbirth among Ethiopian midwifery students. *PLoS ONE*, 15(1), e0227958. <https://doi.org/10.1371/journal.pone.0227958>
- *Barnett, R.C. (2010). Longitudinal Study of Dual-Earner Couples, 1989-1992 (Harvard Dataverse; Version V5) [Data set]. <https://doi.org/10.7910/DVN/YWFANY>
- Beauducel, A. (2011). Indeterminacy of factor score estimates in slightly misspecified confirmatory models. *Journal of Modern Applied Statistical Methods*, 10, 583-598.
<https://doi.org/10.22237/jmasm/1320120900>

- Billiet, J.B., & McClendon, M.J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7(4), 608-628. https://doi.org/10.1207/S15328007SEM0704_5
- Bonifay, W. (2015). *An integrative framework of model evaluation* (unpublished doctoral dissertation). University of California, Los Angeles. <https://escholarship.org/uc/item/5j04v50q>
- Bonifay, W., & Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behavioral Research*, 52(4), 465-484. <https://doi.org/10.1080/00273171.2017.1309262>
- Bonifay, W., Lane, S.P., & Reise, S.P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184-186. <https://doi.org/10.1177/2167702616657069>
- Bonifay, W.E., Reise, S.P., Scheines, R., & Meijer, R.R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. *Structural Equation Modeling*, 22(4), 504-516. <https://doi.org/10.1080/10705511.2014.938596>
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80(4), 796-846. <https://doi.org/10.1111/j.1467-6494.2011.00749.x>
- *Bueno, A., Lima, S., Arias, P., Peña, E., Aguilar, M., & Cabrera, M. (2020). Estructura factorial, invarianza y propiedades psicométricas de la escala de autoestima de Rosenberg en el contexto Ecuatoriano [Factor structure, invariance and psychometric properties of the Rosenberg self-esteem scale in the Ecuadorian context]. *Revista Iberoamericana de Diagnóstico y Evaluación*, 56, 87-100. <https://doi.org/10.21865/RIDEP56.3.01>
- Canivez, G.L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for multidimensionality and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements*. Germany: Hogrefe.

- Carmines, E.G., & Zeller, R.A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Cattell, R.B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum Press. <http://dx.doi.org/10.1007/978-1-4684-2262-7>
- *CentERdata. (2008). Longitudinal Internet studies for the social sciences [Computer file]. Tilburg, The Netherlands: Tilburg University [Distributor]. Retrieved from <http://www.lissdata.nl>
- Cheung, M.W.L. (2014). Fixed-and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods*, 46, 29-40. <https://doi.org/10.3758/s13428-013-0361-y>
- Clark, S.L., Muthén, B., Kaprio, J., D’Onofrio, B.M., Viken, R., & Rose, R.J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling*, 20(4), 681-703. <https://doi.org/10.1080/10705511.2013.824786>
- *Corker, K.S., Donnellan, M.B., Kim, S.Y., Schwartz, S.J., & Zamboanga, B.L. (2017). College student samples are not always equivalent: The magnitude of personality differences across colleges and universities. *Journal of Personality*, 85(2), 123-135. <https://doi.org/10.1111/jopy.12224>
- Corwyn, R.F. (2000). The factor structure of global self-esteem among adolescents and adults. *Journal of Research in Personality*, 34(4), 357-379. <https://doi.org/10.1006/jrpe.2000.2291>
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302. <http://doi.org/10.1037/h0040957>
- *Csajbók, Z., Havlíček, J., Demetrovics, Z., Berkics, M. (2019). Self-perceived mate value is poorly predicted by demographic variables. *Evolutionary Psychology*, 17(1), 1-14. <https://doi.org/10.1177/1474704919829037>
- Dalal, D.K., & Carter, N.T. (2014). Negatively worded items negatively impact survey research. In C.E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends*. New York, NY: Routledge.

- *Demetrovics, Z., Király, O., Koronczi, B., Griffiths, M.D., Nagygyörgy, K., Elekes, Z., Tamás, D., Kun, B., Kökönyei, G., & Urbán, R. (2016). Psychometric properties of the Problematic Internet Use Questionnaire short-form (PIUQ-SF-6) in a nationally representative sample of adolescents. *PLoS ONE*, *11*(8), e0159409. <https://doi.org/10.1371/journal.pone.0159409>
- *Deutsches Zentrum für Altersfragen (2021). Forschungsdatenzentrum des DZA (FDZ-DZA). Retrieved from <https://www.dza.de/forschung/fdz>
- *DiStefano, C., & Motl, R.W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464. https://doi.org/10.1207/s15328007sem1303_6
- DiStefano, C., & Motl, R.W. (2009a). Personality correlates of method effects due to negatively worded items on the Rosenberg self-esteem scale. *Personality and Individual Differences*, *46*(3), 309-313. <https://doi.org/10.1016/j.paid.2008.10.020>
- *DiStefano, C., & Motl, R.W. (2009b). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling*, *16*, 134-146. <https://doi.org/10.1080/10705510802565403>
- *Donnellan, M.B., Ackerman, R.A., & Brecheen, C. (2016). Extending structural analyses of the Rosenberg self-esteem scale to consider criterion-related validity: Can composite self-esteem scores be good enough? *Journal of Personality Assessment*, *98*(2), 169-177. <http://dx.doi.org/10.1080/00223891.2015.1058268>
- *Donnellan, M.B. & Rakhshani, A. (2020). How does the number of response options impact the psychometric properties of the Rosenberg self-esteem scale? <https://doi.org/10.31234/osf.io/fnywz>
- Dueber, D. M. (2019). BifactorIndicesCalculator: Bifactor indices calculator. (R package version 3.1) [Computer software]. <https://cran.r-project.org/web/packages/subscore/index.htm>
- Dueber, D.M., & Toland, M.D. (2021). A bifactor approach to subscore assessment. *Psychological Methods*. <https://doi.org/10.1037/met0000459>

- *Dueber, D.M., Toland, M.D., Lingat, J.E., Love, A.M.A., Qiu, C., Wu, R., & Brown, A.V. (2021). To reverse item orientation or not to reverse item orientation, that is the question. *Assessment*. <https://doi.org/10.1177/10731911211017635>
- Feinberg, R. A., & Jurich, D. P. (2017). Guidelines for interpreting and reporting subscores. *Educational Measurement: Issues and Practice*, 36(1), 5-13. <https://doi.org/10.1111/emip.12142>
- Ferrando, P.J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of for solutions and factor scores estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, 78(5), 762-780. <https://doi.org/10.1177.001316441771930>
- Gagné, P., & Hancock, G.R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, 41(1), 65-83. https://doi.org/10.1207/s15327906mbr4101_5
- *Gana, K., Alaphilippe, D., & Bailly, N. (2005). Factorial structure of the French version of the Rosenberg self-esteem scale among the elderly. *International Journal of Testing*, 5(2), 169-176. https://doi.org/10.1207/s15327574ijt0502_5
- *Gana, K., Saada, Y., Bailly, N., Joulain, M., Hervé, M., & Alaphilippe, D. (2013). Longitudinal factorial invariance of the Rosenberg self-esteem scale: Determining the nature of method effects due to item wording. *Journal of Research in Personality*, 47(4), 406-416. <https://doi.org/10.1016/j.jrp.2013.03.011>
- *García, Z.E., Guerra, K., Garrido, L.E., Cantisano, L.M., Moretti, L., Cano, A., Arias, V.B., & Medrano, L.A. (2021). Using constrained factor mixture analysis to validate mixed-worded psychological scales: The case of the Rosenberg self-esteem scale in the Dominican Republic. *Frontiers in Psychology*, 12, 636693. <https://doi.org/10.3389/fpsyg.2021.636693>
- Geiser, C., Eid, M., & Nussbeck, F.W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods*, 13(1), 49-57. <https://doi.org/10.1037/1082-989X.13.2.171>

- Gnambs, T., & Staufenbiel, T. (2016). Parameter accuracy in meta-analyses of factor structures. *Research Synthesis Methods*, 7, 168-186.
<https://doi.org/10.1002/jrsm.1190>
- *Gnambs, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg self-esteem scale: A cross-cultural meta-analysis. *Zeitschrift für Psychologie*, 226(1), 14-29. <https://doi.org/10.1027/2151-2604/a000317>
- *Gnambs, T., & Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg self-esteem scale. *Assessment*, 27(2), 404-418.
<https://doi.org/10.1177/1073191117746503>
- *Goldberg, L.R. (2018). Eugene-Springfield Community Sample: Personal Attribute Survey (PAS) (Harvard Dataverse; Version 1) [Data set].
<https://doi.org/10.7910/DVN/QYKXUE>
- *Golec de Zavala, A., Federico, C. M., Sedikides, C., Guerra, R., Lantos, D., Mroziński, B., Cypryńska, M., & Baran, T. (2020). Low self-esteem predicts out-group derogation via collective narcissism, but this relationship is obscured by in-group satisfaction. *Journal of Personality and Social Psychology*, 119(3), 741-764.
<http://dx.doi.org/10.1037/pspp0000260>
- *Gómez, M., Espada, J.P., Morales, A., Marchal, L., Soler, F., Vallejo, P. (2016). Adaptation, validation, reliability and factorial equivalence of the Rosenberg self-esteem scale in Colombian and Spanish population. *The Spanish Journal of Psychology*, 19, e66. <https://doi.org/10.1017/sjp.2016.67>
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S.P. (2003). Item-wording and the dimensionality of the Rosenberg self-esteem scale: Do they matter? *Personality and Individual Differences*, 35(6), 1241-1254. [https://doi.org/10.1016/S0191-8869\(02\)00331-8](https://doi.org/10.1016/S0191-8869(02)00331-8)
- Grice, J.W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430-450. <https://doi.org/10.1037/1082-989X.6.4.430>
- Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

- Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, 8, 65-81. <https://doi.org/10.1111/j.2044-8317.1955.tb00321.x>
- Hancock, G.R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between groups hypothesis testing of means on a single latent construct. *Psychometrika*, 66, 373-388. <http://dx.doi.org/10.1007/BF02294440>
- Hancock, G.R., & Mueller, R.O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future*. Lincolnwood, IL: Scientific Software International.
- Henrich, J., Heine, S.J., & Norenzayan, A. (2010a). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83. <https://doi.org/10.1017/S0140525X0999152X>
- Henrich, J., Heine, S.J., & Norenzayan, A. (2010b). Most people are not WEIRD. *Nature*, 466, 29. <https://doi.org/10.1038/466029a>
- Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savović, J., Schulz, K.F., Weeks, L., & Sterne, J.A.C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343. <https://doi.org/10.1136/bmj.d5928>
- *Holden, C.J., Zeigler-Hill, V., Shackelford, T.K., & Welling, L.L.M. (2018). The impact of relationship-contingent self-esteem on mate retention and reactions to threat. *Personal Relationships*, 25(4), 611-630. <https://doi.org/10.1111/per.12255>
- Holzinger, K.J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41-54. <http://dx.doi.org/10.1007/BF02287965>
- Horan, P.M., DiStefano, C., & Motl, R.W. (2003). Wording effects in self-esteem: Methodological artifact or response style. *Structural Equation Modeling*, 10(3), 435-455. https://doi.org/10.1207/S15328007SEM1003_6

- Huang, C., & Dong, N. (2012). Factor structures of the Rosenberg Self-Esteem Scale: A meta-analysis of pattern matrices. *European Journal of Psychological Assessment*, 28(2), 132-138. <https://doi.org/10.1027/1015-5759/a000101>
- Huang, J.L., Liu, M., & Bowling, N.A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *The Journal of Applied Psychology*, 100(3), 828-845. <https://doi.org/10.1037/a0038510>
- *Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245919882903>
- *Hyland, P., Boduszek, D., Dhingra, K., Shevlin, M., Egan, A. (2014). A bifactor approach to modelling the Rosenberg Self Esteem Scale. *Personality and Individual Differences*, 66, 188-192. <https://doi.org/10.1016/j.paid.2014.03.034>
- *Jabłońska, M.R., & Zajdel, R. (2020). Artificial neural networks for predicting social comparison effects among female Instagram users. *PLoS ONE*, 15(2), e0229354. <https://doi.org/10.1371/journal.pone.0229354>
- Jak, S., & Cheung, M.W.-L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods*, 25(4), 430-455. <https://dx.doi.org/10.1037/met0000245>
- Kam, C.C.S. (2018). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociological Methods & Research*, 47(3), 574-597. <https://doi.org/10.1177/0049124115626177>
- *Kam, C.C.S. (2021). Bifactor model is not the best-fitting model for self-esteem: Investigation with a novel technique. *Assessment*, 28(7), 1799-1812. <https://doi.org/10.1177/1073191120949916>
- Kamoen, N., Holleman, B., & van den Bergh, H. (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, 7(3), 181-189. <https://doi.org/10.18148/srm/2013.v7i3.5034>
- *Kocalevent, R.-D., Zenger, M., Heinen, I., Dwinger, S., Decker, O., & Brähler, E. (2015). Resilience in the general population: Standardization of the Resilience Scale

- (RS-11). *PLoS ONE*, *10*(11), e0140322.
<https://doi.org/10.1371/journal.pone.0140322>
- Lance, C.E., Noble, C.L., & Scullen, S.E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, *7*(2), 228-244. <https://doi.org/10.1037/1082-989X.7.2.228>
- Li, T., Higgins, J.P.T., & Deeks, J.J. (2019). Collecting data. In: J.P.T., Higgins, J., Thomas, J., Chandler, M., Cumpston, T., Li, M.J. Page, & V.A. Welch (Eds). *Cochrane handbook for systematic reviews of interventions*. (2nd ed., pp. 109-142). Wiley Cochrane Series.
- *Lifshin, U., Horner, D.E., Helm, P.J., Solomon, S., & Greenberg, J. (2021). Self-esteem and immortality: Evidence regarding the terror management hypothesis that high self-esteem is associated with a stronger sense of symbolic immortality. *Personality and Individual Differences*, *175*, 1107121.
<https://doi.org/10.1016/j.paid.2021.110712>
- Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, *94*(2), 196-204.
<https://doi.org/10.1080/00223891.2011.645936>
- Liu, X., Cao, P., Lai, X., Wen, J., & Yang, Y. (2022). Assessing essential unidimensionality of scales and structural coefficient bias. *Educational and Psychological Measurement*. <https://doi.org/10.1177/00131644221075580>
- *Lo Coco, G., Salerno, L., Ingoglia, S., Tasca, G.A. (2021). Self-esteem and binge eating: Do patients with binge eating disorder endorse more negatively worded items of the Rosenberg self-esteem scale? *Journal of Clinical Psychology*, *77*(3), 818-836.
<https://doi.org/10.1002/jclp.23065>
- *Lodi-Smith, J., Rodgers, J., Luna, M. V. M., Long, C., Khan, S., & Lundy, Z. (2021). Aging and Autism Study. Retrieved from <https://osf.io/g9c3e/>

- Lubke, G.H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 21–39. <https://doi.org/10.1037/1082-989X.10.1.21>
- McAuley, L., Pham, B., Tugwell, P., & Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *The Lancet*, 356, 1228-1231. [https://doi.org/10.1016/S0140-6736\(00\)02786-0](https://doi.org/10.1016/S0140-6736(00)02786-0)
- *McKay, M.T., Boduszek, D., & Harvey, S.A. (2014). The Rosenberg self-esteem scale: A bifactor answer to a two-factor question? *Journal of Personality Assessment*, 96(6), 654-660. <http://dx.doi.org/10.1080/00223891.2014.923436>
- Markon, K.E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretation. *Annual Review of Clinical Psychology*, 15, 51-69. <https://doi.org/10.1146/annurev-clinpsy-050718-095522>
- Markus, K.A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.
- Marsh, H.W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, 70(4), 810-819. <https://doi.org/10.1037/0022-3514.70.4.810>
- Marsh, H.W., Morin, A.J.S., Parker, P.D., & Kaur, G. (2014). Exploratory structural equation modelling: An integration of the best features of exploratory and confirmatory factor analyses. *Annual Review of Clinical Psychology*, 10, 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Marsh, H.W., Scalas, L.F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg self-esteem scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment*, 22(2), 366-388. <https://doi.org/10.1037/a0019225>
- *Martin, J., Nuñez, J.L., Navarro, J.G., Grijalvo, F. (2007). The Rosenberg self-esteem scale: Translation and validation in university students. *The Spanish Journal of Psychology*, 10(2), 458-467. <https://doi.org/10.1017/S1138741600006727>

- *Masselink, M., van Roekel, E., Hankin, B.L., Keijsers, L., Lodder, G.M.A., Vanhalst, J., Verhagen, M., Young, J.F., & Oldehinkel, A.J. (2018). The longitudinal association between self-esteem and depressive symptoms in adolescents: Separating between-person effects from within-person effects. *European Journal of Personality*, 32(6), 653-671. <https://doi.org/10.1002/per.2179>
- Maydeu-Olivares, A., & Coffman, D.L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344-362. <https://doi.org/10.1037/1082-989X.11.4.344>
- McDonald, R.P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.
- *Medeiro, S.A., Pinho, L.H., & Evangelho, J.A. (2021). Análise bifator da rosenberg self-esteem scale na obesidade mórbida [Bifactor analysis of the rosenberg self-esteem scale in morbid obesity]. *Psico, Porto Alegre*, 52(1), 1-11. <http://dx.doi.org/10.15448/1980-8623.2021.1.35760>
- *Michaelides, M.P., Koutsogiorgi, C., & Panayiotou, G. (2016). Method effects on an adaptation of the Rosenberg self-esteem scale in Greek and the role of personality traits. *Journal of Personality Assessment*, 98(2), 178-188. <http://dx.doi.org/10.1080/00223891.2015.1089248>
- Michaelides, M.P., Koutsogiorgi, C., & Panayiotou, G. (2017). Method/group factors: Inconsequential but meaningful—A comment on Donnellan, Ackerman, and Brecheen (2016). *Journal of Personality Assessment*, 99(3), 334-335. <https://doi.org/10.1080/00223891.2016.1233560>
- *Michaelides, M.P., Zenger, M., Koutsogiorgi, C., Brähler, E., Stöbel-Richter, Y., & Berth, H. (2016). Personality correlates and gender invariance of wording effects in the German version of the Rosenberg self-esteem scale. *Personality and Individual Differences*, 97, 13-18. <https://doi.org/10.1016/j.paid.2016.03.011>
- Morgan, G.B., Hodge, K.J., Wells, K.E., & Watkins, M.W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, 3, 2-20. <https://doi.org/10.3390/jintelligence3010002>

- Morin, A.J.S., Arens, A.K., & Marsh, H.W. (2015). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, 23(1), 116-139. <https://doi.org/10.1080/10705511.2014.961800>
- Motl, R.W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*, 9(4), 562-578. https://doi.org/10.1207/S15328007SEM0904_6
- Murray, A.L., & Johnson, W. (2013). The limitations of model fit in comparing the bifactor versus higher-order models of human cognitive ability structure. *Intelligence*, 41(5), 407-422. <https://doi.org/10.1016/j.intell.2013.06.004>
- Muthén, B.O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52, 431-462. <http://dx.doi.org/10.1007/BF02294365>
- *National Centre for Social Research (2011). Families and Children Study 1999-2008. (UK Data Service. SN: 4427) [Data set]. <http://doi.org/10.5255/UKDASN44271>
- *Nguyen, P.L. (2020). Identity processes and well-being: Exploring different types of identity exploration. <https://doi.org/10.31234/osf.io/jz5md>
- *Orsini, C.A., Binnie, V.I., & Tricio, J.A. (2018). Motivational profiles and their relationships with basic psychological needs, academic performance, study strategies, self-esteem, and vitality in dental students in Chile. *Journal of Educational Evaluation for Health Professions*, 15, 1-6. <https://doi.org/10.3352/jeehp.2018.15.11>
- *Ortuño, V.E. & Cabana, A. (2019). Rosenberg self-esteem scale in Uruguay. <https://doi.org/10.17605/OSF.IO/H5G36>
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

- Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.
<https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P.M., MacKenzie, S.B., & Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- *Polner, B., Hupuczi, E., Kéri, S., & Kállai, J. (2021). Adaptive and maladaptive features of schizotypy clusters in a community sample. *Scientific Reports, 11*, 16653.
<https://doi.org/10.1038/s41598-021-95945-0>
- *Ponce, F.P., Torres Irribarra, D., Vergés, D., & Arias, V.B. (2021). Wording effects in assessment: Missing the trees for the forest. *Multivariate Behavioral Research*.
<https://doi.org/10.1080/00273171.2021.1925075>
- *Portes, A., & Rumbaut, R.G. (2018). Children of Immigrants Longitudinal Study (CILS), San Diego, California, Ft. Lauderdale and Miami, Florida, 1991-2006 (ICPSR 20520; Version V3) [Data set]. ICPSR.
<https://doi.org/10.3886/ICPSR20520.v3>
- Quilty, L.C., Oakman, J.M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling, 13*(1), 99-117.
https://doi.org/10.1207/s15328007sem1301_5
- Quinn, H.O. (2014). *Bifactor models, explained common variance (ECV), and the usefulness of scores from unidimensional Item Response Theory analyses* (Publication No. 1557177). [Master's Thesis, University of North Carolina at Chapel Hill]. ProQuest Dissertations and Theses Global.
- Rauch, W.A., Schweizer, K., & Moosbrugger, H. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality in LOT-R scores. *Personality and Individual Differences, 42*(8), 1597-1607.
<https://doi.org/10.1016/j.paid.2006.10.035>

- Raykov, T., & Pohl, S. (2013). Essential unidimensionality examination for multicomponent scales: An interrelationship decomposition approach. *Educational and Psychological Measurement, 73*(4), 581-600.
<https://doi.org/10.1177/0013164412470451>
- Raykov, T., Marcoulides, G.A., Menold, N., & Harrison, M. (2019). Revisiting the bifactor model: Can mixture modeling help assess its applicability? *Structural Equation Modeling, 26*(1), 110-118.
<https://doi.org/10.1080/10705511.2018.1436441>
- Rhemtulla, M., Brosseau-Liard, P.E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354–373. <https://doi.org/10.1037/a0029315>
- Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667-696.
<http://dx.doi.org/10.1080/00273171.2012.715555>
- Reise, S.P., Bonifay, W.E., & Haviland, M.G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129-140.
<https://doi.org/10.1080/00223891.2012.725437>
- *Reise, S.P., Kim, D.S., Mansolf, M., & Widaman, K.F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg self-esteem scale. *Multivariate Behavioral Research, 51*(6), 818-838.
<https://doi.org/10.1080/00273171.2016.1243461>
- Reise, S.P., Moore, T.M., & Haviland, M.G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559.
<http://dx.doi.org/10.1080/00223891.2010.496477>

- Reise, S.P., Morizot, J., & Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16, 19-31. <https://doi.org/10.1007/s11136-007-9183-7>
- Reise, S.P., Scheines, R., Widaman, K.F., & Haviland, M.G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73, 5-26. <https://doi.org/10.1177/0013164412449831>
- Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, 47, 493-504. <https://doi.org/10.1016/j.jrp.2013.04.012>
- Revelle, W., & Zinbarg, R.E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74, 145-154. <http://dx.doi.org/10.1007/s11336-008-9102-z>
- Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354-373. <https://doi.org/10.1037/a0029315>
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223-237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rodríguez, A., Reise, S.P., & Haviland, M.G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137-150. <https://doi.org/10.1037/met0000045>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- *Rufino, S. & Holden, C.J. (2020). Contingent self-esteem, borderline personality features, and mate retention behavior. retrieved from <https://osf.io/zkp5g/>
- *Salerno, L., Ingoglia, S., & Lo Coco, G. (2017). Competing factor structures of the Rosenberg Self-Esteem Scale (RSES) and its measurement invariance across clinical

- and non-clinical samples. *Personality and Individual Differences*, 113(15), 13-19.
<https://doi.org/10.1016/j.paid.2017.02.063>
- *Sallay, V., Martos, T., Földvári, M., Szabó, T., Ittész, A. (2014). A Rosenberg
 Önértékelés Skála (RSES-H): Alternatív fordítás, strukturális invariancia és validitás
 [Hungarian version of the Rosenberg Self-esteem Scale (RSES-H): An alternative
 translation, structural invariance, and validity]. *Mentálhigiéne és Pszichoszomatika*,
 15(3), 259-275. <https://doi.org/10.1556/mental.15.2014.3.7>
- Savalei, V., & Falk, C.F. (2014). Recovering substantive factor loadings in the presence
 of acquiescence bias: A comparison of three approaches. *Multivariate Behavioral
 Research*, 49(5), 407-424. <https://doi.org/10.1080/00273171.2014.931800>
- *Scheve, K., & Stasavage, D. (2021). Equal treatment and the inelasticity of tax policy
 to rising inequality. <https://doi.org/10.7910/DVN/BWCZKE>
- Schmitt, D.P., & Allik, J. (2005). Simultaneous administration of the Rosenberg self-
 esteem scale in 53 nations: Exploring the universal and culture-specific features of
 global self-esteem. *Journal of Personality and Social Psychology*, 89(4), 623-642.
<https://doi.org/10.1037/0022-3514.89.4.623>
- Schriesheim, C.A., Eisenbach, R.J., & Hill, K.D. (1991). The effect of negation and
 polar opposite item reversals on questionnaire reliability and validity: An
 experimental investigation. *51*(1), 67-78. <https://doi.org/10.1177/0013164491511005>
- Schmucker, C.M., Blümle, A., Schell, L.K., Schwarzer, G., Oeller, P., Cabrera, L., et al.
 (2017). Systematic review finds that study data not published in full text articles
 have unclear impact on meta-analyses results in medical research. *PLoS ONE*, 12(4),
 e0176210. <https://doi.org/10.1371/journal.pone.0176210>
- *Sharrat, K., Boduszek, D., Jones, A., & Gallagher, B. (2014). Construct validity,
 dimensionality and factorial invariance of the Rosenberg self-esteem scale: A
 bifactor modelling approach among children of prisoners. *Current Issues in
 Personality Psychology*, 2(4), 228-236. <https://doi.org/10.5114/cipp.2014.47447>
- *Shimizu, K. & Yoshida, K. (2008). Rosenberg 自尊感情尺度のモデル化 – wording
 と項目配置の影響の検討 – [Modeling of the Rosenberg self-esteem Scale:

- Identifying the wording and arrangement effects among items]. 関西大学 『社会学部紀要』 , 39(2), 69-97.
- *Shin, J. & Lee, T. (2017). 쌍요인 (Bifactor) 모형을 이용한 심리척도의 측정적 속성 연구방법 개관 [Bifactor modeling approach to investigate studying of psychometric properties of psychological measures]. *Korean Journal of Psychology: General*, 36(4), 477-504. <http://dx.doi.org/10.22257/kjp.2017.09.36.4.477>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107-120. <https://doi.org/10.1007/s11336-008-9101-0>
- Simms, L.J., Grös, D.F., Watson, D., & O'Hara, M.W. (2008). Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depression and Anxiety*, 25(7), E34-E46. <http://dx.doi.org/10.1002/da.20432>
- *Silverstein, M. & Bengtson, V.L. (2019). Longitudinal Study of Generations, California, 1971, 1985, 1988, 1991, 1994, 1997, 2000, 2005 (ICPSR 22100; Version V5) [Data set]. ICPSR. <https://doi.org/10.3886/ICPSR22100.v5>
- *Skylark, W., & Callan, M.J. (2021). Personal relative deprivation and reward-based eating: Two exploratory studies. <https://doi.org/10.31234/osf.io/7b59a>
- *Stafford, M.R., Cooper, M., Barkham, M., Beecham, J., Bower, P., Cromarty, K., Fugard, A.J.B., Jackson, C., Pearce, P., Ryder, R., & Street, C. (2018). Effectiveness and cost-effectiveness of humanistic counselling in schools for young people with emotional distress (ETHOS): Study protocol for a randomised controlled trial. *Trials*, 19, 175. <https://doi.org/10.1186/s13063-018-2538-2>
- Steinman, I., Sánchez, D., van Laar, S., & Braeken, J. (2021). The impact of inconsistent responders to mixed-worded scales on inferences in international large-scale assessments. *Assessment in Education*. <https://doi.org/10.1080/0969594X.2021.2005302>

- Steinmann, I., Strietholt, R., & Braeken, J. (2021). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological Methods*. <https://doi.org/10.1037/met0000392>
- Stucky, B.D., Edelen, M.O., Vaughan, C.A., Tucker, J.S., & Butler, J. (2014). The psychometric development and initial validation of the DCI-A short form for adolescent therapeutic community treatment process. *Journal of Substance Abuse Treatment*. 46(4), 516-521. <https://doi.org/10.1016/j.jsat.2013.12.005>
- Syropoulou, A., Vernadakis, N., Papastergiou, M., & Kourtessis, T. (2021). Psychometric evaluation of the Rosenberg Self-Esteem Scale in primary school students with mild intellectual disability: First evidence. *Research in Developmental Disabilities*, 114, 103964. <https://doi.org/10.1016/j.ridd.2021.103964>
- Ten Berge, J.M.F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69, 613-625. <https://doi.org/10.1007/BF02289858>
- Tomás, J.M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling*, 6(1), 84-98. <https://doi.org/10.1080/10705519909540120>
- *Tomás, J.M., Oliver, A., Hontangas, P.M., Sancho, P., & Galiana, L. (2015). Method effects and gender invariance of the Rosenberg self-esteem scale: A study on adolescents. *Acta de Investigación Psicológica*, 5(3), 2194-2203. [https://doi.org/10.1016/S2007-4719\(16\)30009-6](https://doi.org/10.1016/S2007-4719(16)30009-6)
- *Tomás, J.M., Sancho, P., Galiana, L., Oliver, A., Hontangas, P. (2015). Measuring self-esteem in Spanish adolescents: Equivalence across gender and educational levels. *Avaliação Psicológica*, 14(3), 385-393. <http://dx.doi.org/10.15689/ap.2015.1403.10>
- Tourangeau, R. (2017, May 11). The survey response process from a cognitive viewpoint [Conference presentation]. OECD Seminar on Improving the Quality of Data Collection in Large Scale. Assessments, Paris, France. <https://www.oecd.org/skills/piaac/>

- Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511819322>
- *Urbán, R., Szigeti, S., Kökönyei, G., & Demetrovics, Z. (2014). Global self-esteem and method effects: Competing factor structures, longitudinal invariance, and response styles in adolescents. *Behavior Research Methods*, 46, 488-498.
<https://doi.org/10.3758/s13428-013-0391-5>
- *US Department of Labor (2019a). National longitudinal survey of youth 1979. Retrieved from <https://www.nlsinfo.org/content/cohorts/NLSY79>
- *US Department of Labor (2019b). National longitudinal survey of youth 1979: Child and young adult. Retrieved from <https://www.nlsinfo.org/content/cohorts/nlsy79-children>
- *Vaughn, L.A. (2019). Distinguishing between need support and regulatory focus with LIWC. *Collabra: Psychology*, 5(1), 1-20. <https://doi.org/10.1525/collabra.185>
- *Ventura, J., Caycho, T., Barboza, M. (2018). Evidencias psicométricas de la escala de autoestima de Rosenberg en adolescentes limeños [Psychometric evidence of the Rosenberg's self-esteem scale with adolescents from Lima]. *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, 52(1), 44-60.
<https://doi.org/10.30849/rip/ijp.v52i1.363>
- *Wang, Y., Kong, F., Huang, L., & Liu, J. (2016). Neural correlates of biased responses: The negative method effect in the Rosenberg self-esteem scale is associated with right amygdala volume. *Journal of Personality*, 84(5), 623-632.
<https://doi.org/10.1111/jopy.12185>
- Watts, A.L., Poore, H.E., Waldman, I.D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 7(6), 1285-1303. <https://doi.org/10.1177/2167702619855035>
- *Welsh Assembly Government (2011). National Survey for Wales, 2009-2010: Pilot Study. (UK Data Service. SN: 6720) [Data set]. <http://doi.org/10.5255/UKDA-SN-6720-1>

- *Wehner, C., Maaß, U., Leckelt, M., Back, M.D., & Ziegler, M. (2020). Validation of the short dark triad in a German sample: Structure, nomological network, and an ultrashort version. *European Journal of Psychological Assessment*.
<https://doi.org/10.1027/1015-5759/a000617>
- Wood, D., Harms, P.D., Lowman, G.H., & DeSimone, J.A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science*, 8(4), 454–464.
<https://doi.org/10.1177/1948550617703168>
- *Wu, Y., Zuo, B., Wen, F., & Yan, L. (2017). Rosenberg self-esteem scale: Method effects, factorial structure and scale invariance across migrant child and urban child populations in China. *Journal of Personality Assessment*, 99(1), 83-93.
<https://doi.org/10.1080/00223891.2016.1217420>
- *Xu, M., & Leung, S.-O. (2016). Bifactor structure for the categorical Chinese Rosenberg self-esteem scale. *Spanish Journal of Psychology*, 19, e67.
<https://doi.org/10.1017/sjp.2016.66>
- *Yovel, I., Aviram, G., Kahana, N., & Katz, B.A. (2021). Testing a new indirect measure of general self-worth: The self-esteem questionnaire-based implicit association test. *British Journal of Social Psychology*.
<https://doi.org/10.1111/bjso.12472>
- *Yin, M. (2019). Parental involvement, students' self-esteem and academic achievement in immigrant families in the United States [Master's thesis, Florida State University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/parental-involvement-students-self-esteem/docview/2287470662/se-2?accountid=14512>
- *Zhang, X., & Savalei, V. (2016). Improving the factor structure of psychological scales: The expanded format as an alternative to the Likert scale format. *Educational and Psychological Measurement*, 76(3), 357-386.
<https://doi.org/10.1177/0013164415596421>

- *Zhang, X., Tse, W.W.-Y., & Savalei, V. (2019). Improved properties of the Big Five Inventory and the Rosenberg self-esteem scale in the expanded format relative to the Likert format. *Frontiers in Psychology*, 10, 1286.
<https://doi.org/10.3389/fpsyg.2019.01286>
- Zinbarg, R.E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70, 123-133.
<http://dx.doi.org/10.1007/s11336-003-0974-7>
- Zinbarg, R.E., Yovel, I., Revelle, W., & McDonald, R.P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h . *Applied Psychological Measurement*, 30(2), 121-144. <http://dx.doi.org/10.1177/0146621605278814>
- *Zitelny, H., Dror, T., Altman, S., & Bar-Anan, Y. (2021). The relation between gender identity and well-being. *Personality and Social Psychology Bulletin*.
<https://doi.org/10.1177/01461672211002362>

Tabla 1. Resumen de los registros seleccionados.

Autores	Mod	País	n	Edad	Sexo (H/M%)	B-FI	B-FID
Gana et al. (2005)	1	Francia	864	60-95a (M=72,71; DE=5,98)	386H(44,7%)/478M(55,3%)		x
DiStefano & Motl (2006) ^d	2	Estados Unidos	757	16-75a (M=22,40; DE=7,20)	H(42,0%)/M(48,0%)	x	
Martin et al (2007)	3	España	420	17-58a (M=21,29; DE=4,96)	124H(29,5%)/296M(70,5%)	x	
CentERdata (2008) ^{b,d}	4,5	Países Bajos	5598	16-96a (M=46,96; DE=16;59)	2543H(45,4%)/2942M(52,6%)	x	x
	6,7	Países Bajos	5270	16-98a (M=49,94; DE=17,43)	2362H(44,8%)/2779M(52,7%)	x	x
	8,9	Países Bajos	5119	16-93a (M=51,37; DE=17,69)	2339H(45,7%)/2705M(52,8%)	x	x
	10,11	Países Bajos	5865	16-100a (M=51,38; DE=18,29)	2721H(46,4%)/3144M(53,6%)	x	x
	12,13	Países Bajos	4971	16-102a (M=53,18; DE=18,40)	2308H(46,4%)/2663M(53,6%)	x	x
	14,15	Países Bajos	5308	16-103a (M=53,05; DE=18,64)	2449H(46,1%)/2859M(53,9%)	x	x
Shimizu & Yoshida (2008)	16	Japón	617	18-62a (M=18,55; DE=2,20)	206H(33,4%)/411M(66,6%)	x	
DiStefano & Motl (2009b) ^d	17,18	Estados Unidos	680	16-75a (M=22,40; DE=7,20)	316H(46,5%)/361M(53,1%)	x	
Barnett (2010) ^b	19	Estados Unidos	600	22-49a (M=34,60; DE=4,59)	300H(50,0%)/300M(50,0%)	x	x
	20	Estados Unidos	586	22-49a (M=34,61; DE=4,58)	292H(49,8%)/294M(50,2%)	x	x
	21	Estados Unidos	549	22.49a (M=34,72; DE=4,53)	271H(49,4%)/278M(50,6%)	x	x
National Centre for Social Research (2011) ^b	22	Escocia	4159	16-72a (M=35,02; DE=8,24)	4159M(100%)	x	x
	23	Escocia	4350	16-73a (M=35,16; DE=8,27)	4350M(100%)	x	x
	24	Escocia	7416	16-74a (M=37,13; DE=7,98)	7416M(100%)	x	x
Welsh Assembly Government (2011) ^b	25	Gales	3066	16-96a (M=56,10; DE=18,66)	1411H(46,0%)/1655M(54,0%)	x	x
Gana et al. (2013) ^c	26	Francia					x
Hyland et al. (2014)	27	Estados Unidos	5895	>18a	H(37,5%)/M(62,5%)		x
McKay et al. (2014)	28	Reino Unido	3862	12-16a	1975H(51,1%)/1887M(48,9%)		x
Sallay et al. (2014) ^d	29	Hungría	1702	>18a	674H(39,6%)/1025M(60,2%)		x
Sharratt et al. (2014)	30	Pooled	724	7-17a (M=11,27; DE=3,12)	393H(54,3%)/331M(45,7%)		x
Urbán et al. (2014)	31	Hungría	2513	(M=15,70; DE=0,55)	H(49,0%)/M(51,0%)		x
	32	Hungría	2370	(M=16,40; DE=0,68)	H(48,0%)/M(52,0%)		x
Kocalevent et al. (2015) ^a	33	Alemania	5036	14-92a (M=48,37; DE=17,97)	2334H(46,3%)/2702M(53,7%)	x	x
Tomás, Sancho et al. (2015)	34-37	España	525	(M=17,80; DE=3,69)	H(38,7%)/M(61,3%)	x	
Tomás, Oliver et al. (2015)	38,39	España	386	(M=17,80; DE=3,69)	167H(43,3%)/219M(56,7%)	x	
Demetrovics et al. (2016) ^a	40	Hungría	5005	15-23a (M=16,36; DE=0,87)	2523H(50,4%)/2482M(49,6%)	x	x
Donnellan et al. (2016) ^a	41	Estados Unidos	1127	18-52a (M=18,31; DE=1,56)	H(55,0%)/M(45,0%)	x	x
Gómez et al. (2016) ^d	42	España	506	18-72a (M=32,21; DE=10,68)	195M(38,5%)/304F(60,1%)		x
	43	Colombia	633	18-70a (M=33,93; DE=12,71)	234M(37,0%)/392F(61,9%)		x
Michaelides, Koutsogiorgi et al. (2016)	44	Grecia	205	17-25a (M=21,35; DE=1,72)	53H(25,9%)/152M(74,1%)		x
Michaelides, Zenger et al. (2016)	45	Alemania	346	38-39a	159H(45,9%)/187M(54,1%)		x
Reise et al. (2016), Kam (2021) ^{a,d}	46	Pooled	45546	9-100a (M=26,63; DE=12,39)	17264H(37,1%)/28338M(60,9%)	x	x
Wang et al. (2016)	47,48	China	272	(M=22,64; DE=1,03)	122H(44,9%)/150M(55,1%)		x
Xu & Leung (2016) ^d	49,50	China	1734	Estudiantes (5to Grado)	738H(42,6%)/992M(57,2%)	x	x
Zhang & Savalei (2016)	51	Canadá	641	(M=20,00; DE=2,75)	135H(21,0%)/506M(79,0%)	x	

Alessandri et al. (2017)	52	Agrupados	6138	(M=21,57; DE=3,80)	Agrupados				
Corker et al. (2017) ^{a,d}	53	Estados Unidos	7787	>18a (M=20,35; DE=3,41)	2096H(26,9%)/5663M(72,7%)	x		x	
Salerno et al. (2017)	54,55	Italia	855	18-64a (M=40,03; DE=12,87)	855F(100%)				x
Shin & Lee (2017)	56	Corea del Sur	223	(M=22,81; DE=5,53)	H(31,3%)/M(68,7%)				x
Wu et al. (2017) ^d	57	China	982	(M=12,93; DE=2,10)	474H(48,3%)/470M(47,9%)				x
Gnambs et al. (2018)	58	Agrupados	140671	16-68a (M=28,05; DE=12,95)	Agrupados	x			x
Goldberg (2018) ^b	59	Estados Unidos	734	18-85a (M=51,50; DE=12,68)	312H(42,5%)/422M(57,5%)	x			x
Holden et al. (2018) ^a	60	Estados Unidos	829	>18a (M=20,64; DE=4,28)	142H(17,1%)/687M(82,9%)	x			x
Masselink et al. (2018) ^a	61	Bélgica	943	(M=15,79; DE=1,31)	315H(33,4%)/628M(66,6%)	x			x
	62	Bélgica	1069	(M=16,62; DE=1,22)	373H(34,9%)/696M(65,1%)	x			x
	63	Bélgica	942	(M=17,39; DE=1,27)	323H(34,3%)/619M(65,7%)	x			x
Orsini et al. (2018) ^a	64	Chile	924	18-44a (M=22,83; DE=3,36)	341H(36,9%)/583M(63,1%)	x			x
Portes & Rumbaut (2018) ^b	65,66	Estados Unidos	5006	12-18a (M=14,23; DE=0,86)	2444H(48,8%)/2562M(51,2%)	x			x
Stafford et al (2018) ^{a,d}	67	Reino Unido	329	8-11a (M=9,27; DE=0,80)	69H(21,2%)/256M(78,8%)	x			x
	68	Reino Unido	314	13-16a (M=13,79; DE=0,79)	69H(22,0%)/241M(76,8%)	x			x
	69	Reino Unido	312	13-16a (M=13,78; DE=0,79)	65H(20,8%)/243M(77,9%)	x			x
Ventura et al. (2018) ^a	70	Perú	931	11-18a (M=15,01; DE=1,51)	480H(51,6%)/451M(48,4%)	x			x
Csajbók et al. (2019)	71	Hungría	1131	18-45a (M=23,19; DE=5,01)	H(37,3%)/M(62,7%)				x
Ortuño et al. (2019) ^a	72	Uruguay	322	18-79a (M=24,09; DE=7,70)	141H(43,8%)/181M(56,2%)	x			x
Silverstein & Bengtson (2019) ^{b,d}	73,74	Estados Unidos	1445	19-95a (M=50,16; DE=15,98)	601H(41,6%)/844M(58,4%)	x			x
	75,76	Estados Unidos	1417	14-93a (M=48,05; DE=18,24)	575H(40,6%)/842M(59,4%)	x			x
	77,78	Estados Unidos	1669	15-102a (M=48,06; DE=19,20)	671H(40,2%)/997M(57,8%)	x			x
	79,80	Estados Unidos	1674	16-96a (M=47,32; DE=19,37)	695H(41,5%)/979M(58,5%)	x			x
	81,82	Estados Unidos	1870	16-97a (M=46,23; DE=18,82)	803H(42,9%)/1067M(57,1%)	x			x
US Department of Labor (2019a) ^b	83,84	Estados Unidos	11992	15-23a (M=18,83; DE=2,30)	6021H(50,2%)/5971M(49,8%)	x			x
	85,86	Estados Unidos	10485	22-30a (M=25,80; DE=2,27)	5045H(48,8%)/5295M(51,2%)	x			x
US Department of Labor (2019b) ^b	87,88	Estados Unidos	1668	14-26a (M=17,26; DE=2,05)	859H(51,5%)/809M(48,5%)	x			x
	89,90	Estados Unidos	1308	15-30a (M=18,70; DE=4,10)	663H(50,7%)/645M(49,3%)	x			x
	91,92	Estados Unidos	5010	17-36a (M=22,74; DE=3,91)	2483H(49,6%)/2527M(50,4%)	x			x
	93,94	Estados Unidos	6293	15-38a (M=23,15; DE=4,86)	3119H(49,6%)/3174M(50,4%)	x			x
	95,96	Estados Unidos	1140	15-42a (M=26,98; DE=8,51)	579H(50,8%)/561M(49,2%)	x			x
	97,98	Estados Unidos	2032	14-34a (M=26,18; DE=6,39)	959H(47,2%)/1073M(52,8%)	x			x
Vaughn (2019) ^d	99	Estados Unidos	240	18-75a (M=35,75; DE=12,94)	97H(40,4%)/142M(59,2%)	x			x
	100	Estados Unidos	252	18-70a (M=33,05; DE=11,47)	118H(46,8%)/130M(51,6%)	x			x
Yin (2019)	101	Estados Unidos	1070	(M=17,00)	516H(48,2%)/554M(51,8%)				x
Zhang et al. (2019) ^a	102	Canadá	289	(M=20,00; DE=2,55)	H(22,0%)/M(78,0%)	x			x
Bákker et al. (2020) ^{a,d}	103	Etiopía	403	20-40a (M=23,68; DE=2,62)	154H(38,2%)/248M(61,5%)	x			x
Bueno et al. (2020)	104	Ecuador	535	>18a (M=21,80; DE=4,70)	130H(24,4%)/405M(75,6%)	x			x
Donnellan & Rakhshani (2020) ^{a,d}	105	Angloparlantes	386	18-65a	177H(45,9%)/196M(50,8%)	x			x
	106	Angloparlantes	363	18-65a	189H(52,1%)/169M(46,6%)	x			x
	107	Angloparlantes	401	18-65a	210H(52,4%)/183M(45,6%)	x			x

	108	Angloparlantes	386	18-65a	192H(49,7%)/191M(49,5%)	x	x
	109	Angloparlantes	516	18-52a	135H(26,2%)/380M(73,6%)	x	x
	110	Angloparlantes	476	18-52a	127H(26,7%)/344M(72,3%)	x	x
Gnambs & Schroeders (2020)	111	Alemania	12437	(M=14,68; DE=0,69)	H(50,0%)/M(50,0%)	x	x
Golec de Zavala et al. (2020) ^a	112	Polonia	427	18-80a (M=43,67; DE=15,30)	207H(48,5%)/220M(51,5%)	x	x
	113	Polonia	506	18-76a (M=43,15; DE=15,37)	233H(46,0%)/273M(54,0%)	x	x
	114	Polonia	1059	18-84a (M=44,26; DE=15,58)	494H(46,6%)/565M(53,4%)	x	x
Hussey & Hughes (2020) ^a	115	Pooled	7478	18-65a (M=31,77; DE=11,74)	2620H(35,0%)/4858M(65,0%)	x	x
Jabłońska & Zajdel (2020) ^a	116	Polonia	974	18-49a	974M(100,0%)	x	x
Nguyen (2020) ^{a,d}	117	Estados Unidos	1260	18-30a (M=24,41; DE=3,87)	310H(24,6%)/914M(72,5%)	x	x
Wehner et al. (2020) ^a	118	Alemania	834	16-65a (M=23,06; DE=5,88)	304H(36,5%)/530M(63,5%)	x	x
Deutsches Zentrum für Altersfragen (2021) ^b	119,120	Alemania	8196	40-93a (M=62,87; DE=11,65)	3106H(51,3%)/2949M(48,7%)	x	x
	121,122	Alemania	3972	43-95a (M=62,25; DE=10,77)	1978H(49,8%)/1994M(50,2%)	x	x
	123,124	Alemania	7998	40-95a (M=64,53; DE=11,24)	3919H(49,0%)/4079M(51,0%)	x	x
	125,126	Alemania	5593	43-97a (M=67,19; DE=10,66)	2796H(50,0%)/2797M(50,0%)	x	x
Dueber et al. (2021) ^{a,d}	127	Estados Unidos	507	>18a (M=38,40; DE=12,12)	188H(37,1%)/312M(61,5%)	x	x
	128	Estados Unidos	503	>18a (M=38,58; DE=12,31)	180H(35,8%)/320M(63,6%)	x	x
García et al. (2021) ^a	129	Rep. Dominicana	632	18-63a (M=29,60; DE=10,28)	265H(41,9%)/367M(58,1%)	x	x
Lifshin et al. (2021) ^a	130	Estados Unidos	1433	18-42a (M=18,52; DE=1,59)	490H(34,2%)/943M(65,8%)	x	x
Lo Coco et al. (2021)	131	Italia	533	18-64a (M=42,59; DE=12,78)	533M(100%)	x	x
Lodi-Smith et al. (2021) ^{b,d}	132	Estados Unidos	267	60-97a (M=71,54; DE=7,76)	115H(43,1%)/151M(56,6%)	x	x
Medeiro et al. (2021)	133	Brasil	288	18-61a (M=36,00; DE=8,67)	288M(100%)	x	
Polner et al. (2021) ^a	134	Hungría	639	18-49a (M=25,69; DE=7,89)	182H(28,5%)/457M(71,5%)	x	x
Ponce et al. (2021) ^{a,d}	135	Países Bajos	6762	15-108a (M=45,81; DE=15,96)	2885H(42,7%)/3485M(51,5%)	x	x
Rufino & Holden (2021) ^a	136	Estados Unidos	204	21-70a (M=36,93; DE=10,41)	M(50,0%)/F(50,0%)	x	x
Scheve & Stasavage (2021) ^a	137	Reino Unido	1913	19-88a (M=49,35; DE=17,16)	875H(45,7%)/1038M(54,3%)	x	x
Skylark & Callan (2021) ^{a,d}	138	Estados Unidos	845	19-75a (M=36,31; DE=11,45)	462H(54,7%)/380M(45,0%)	x	x
Yovel et al. (2021) ^{a,d}	139	Angloparlantes	241	18-69a (M=37,02; DE=13,13)	88H(36,5%)/152M(63,1%)	x	x
	140	Angloparlantes	350	18-77a (M=37,18; DE=12,87)	132H(37,7%)/215M(61,4%)	x	x
Zitelny et al. (2021) ^a	141	Estados Unidos	1127	15-81a (M=37,12; DE=14,25)	1127M(100%)	x	x
	142	Estados Unidos	1115	16-79a (M=33,57; DE=14,47)	1115M(100%)	x	x
	143	Estados Unidos	1124	17-81a (M=39,58; DE=14,24)	1124H(100%)	x	x
	144	Estados Unidos	1140	17-80a (M=31,79; DE=13,77)	1140M(100%)	x	x

Nota. B-FI=Modelo bifactor con un factor específico de ítems inversos; B-FID=Modelo bifactor con factores específicos de ítems directos e inversos.

^a A partir de la disponibilidad de los datos, los modelos fueron nuevamente estimados para corregir parámetros adicionales originalmente especificados (e.g., unicidades correlacionadas; Ponce et al., 2021; Reise et al., 2016), corregir errores (e.g., datos no recodificados; Ventura et al., 2018) o estimar los modelos bifactor.

^b Archivos de datos empleados para estimar los modelos confirmatorios bifactor.

^c Los autores reportaron cargas factoriales promedio.

^d No todos los participantes reportaron su sexo o reportaron una identidad sexual no binaria.

Figura 1. Modelos bifactor examinados.

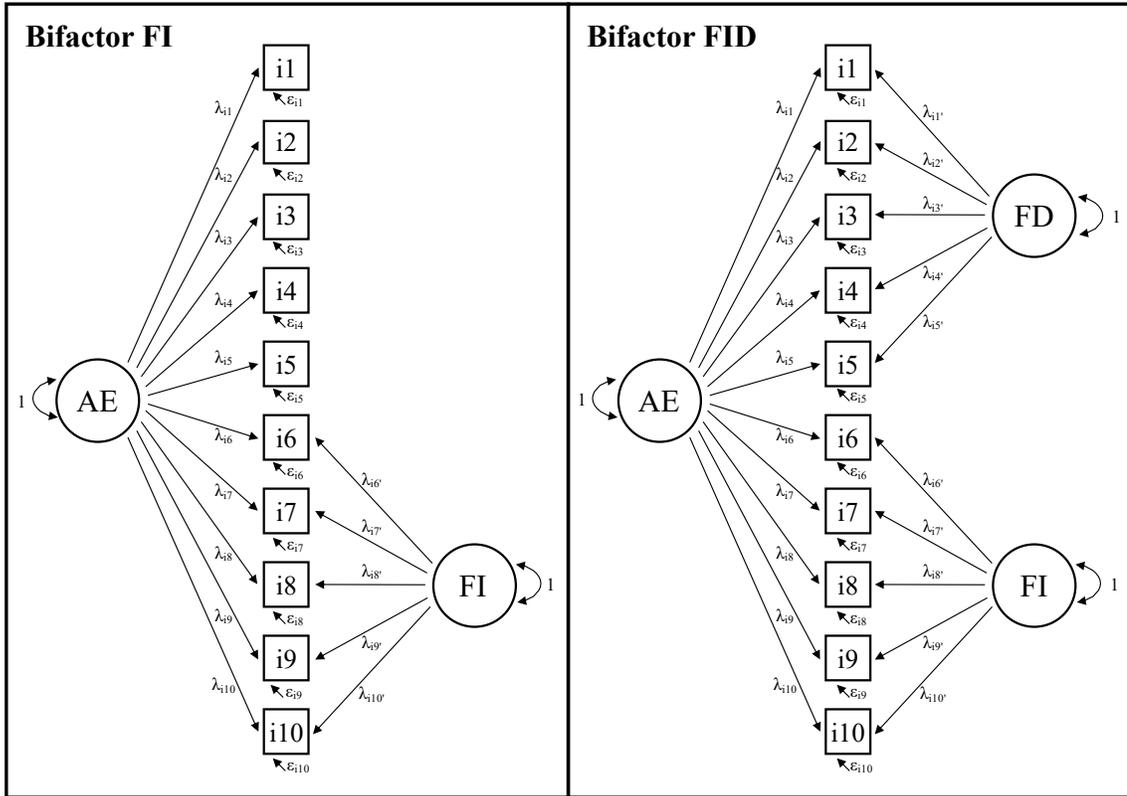


Figura 2. Diagrama de flujo del proceso de revisión, identificación y selección de los estudios.

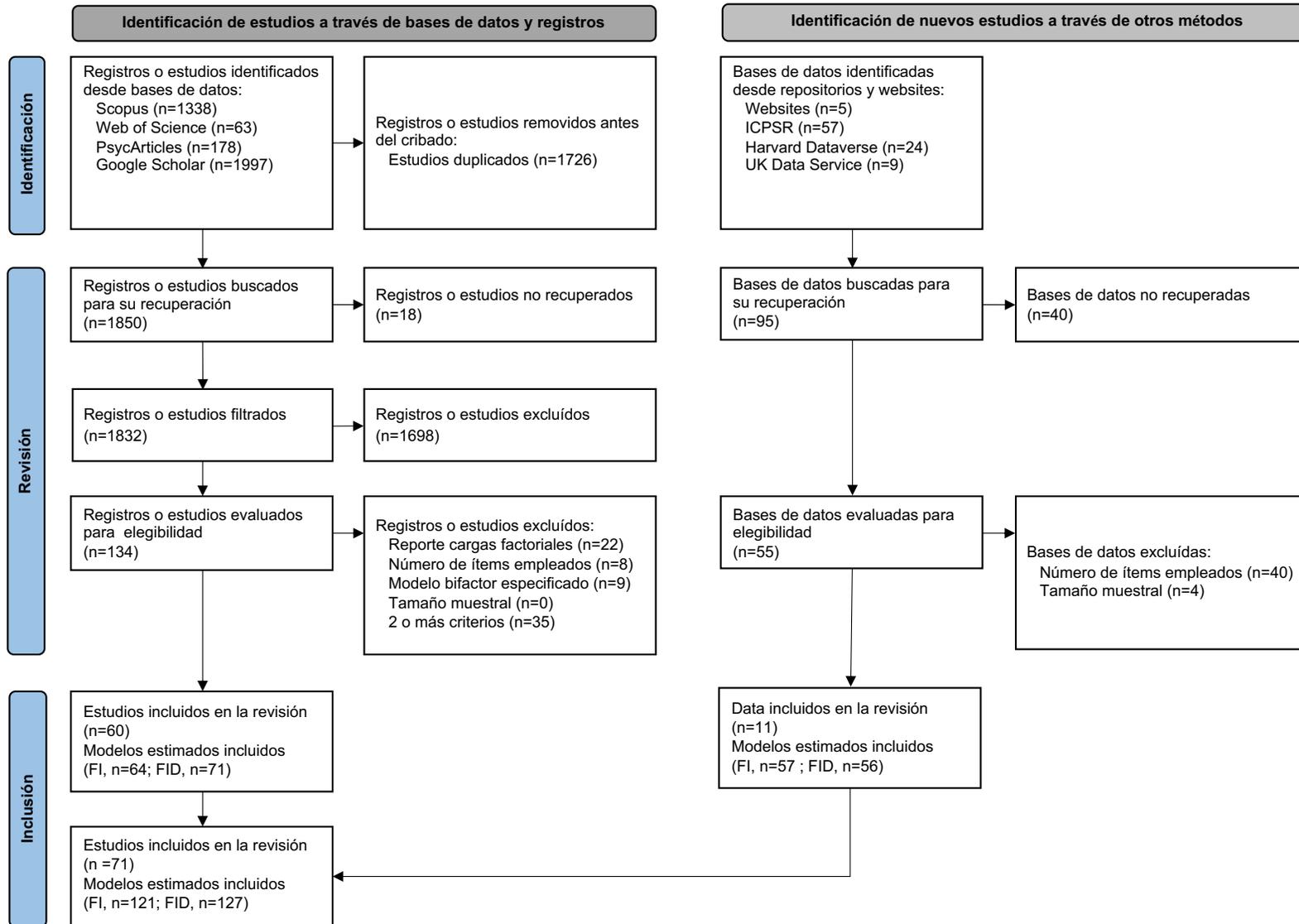


Figura 3. Resumen evaluación del riesgo de sesgo.

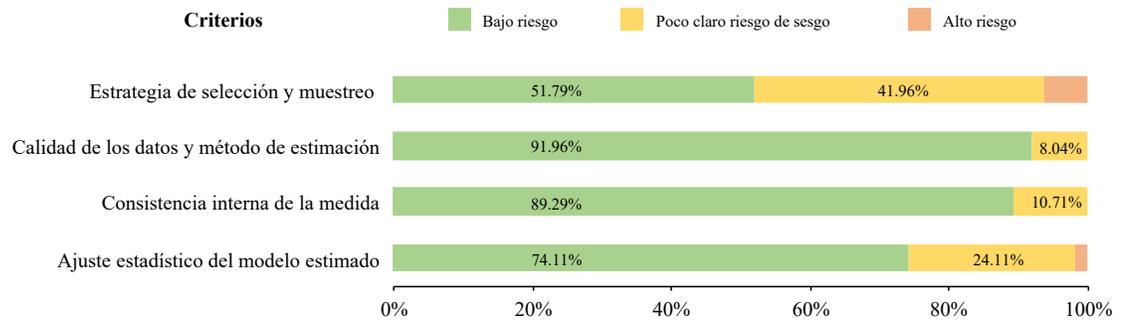


Figura 4. Coeficientes omega (ω) y omega jerárquico (ω_h) de los factores general de autoestima (círculos grises) y específico de redacción asociado a ítems inversos (círculos blancos), estimados del modelo Bifactor-FI.

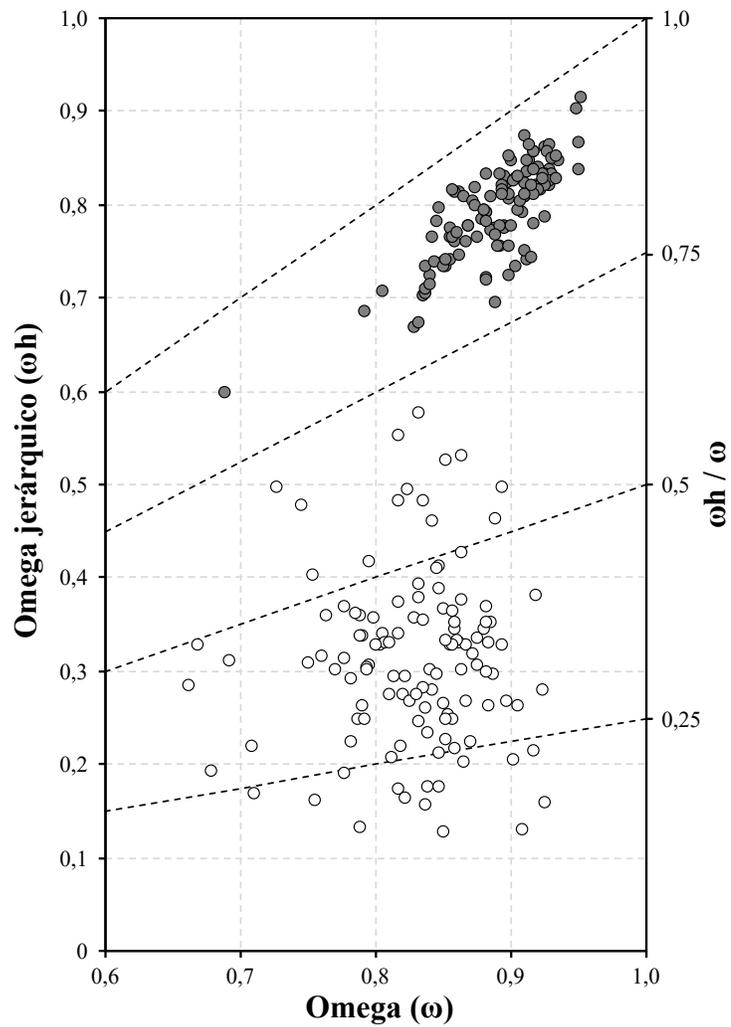


Figura 5. Coeficientes omega (ω) y omega jerárquico (ω_h) de los factores general de autoestima (círculos grises) y específicos de redacción asociado a ítems directos (círculos negros) e inversos (círculos blancos), estimados del modelo Bifactor-FID.

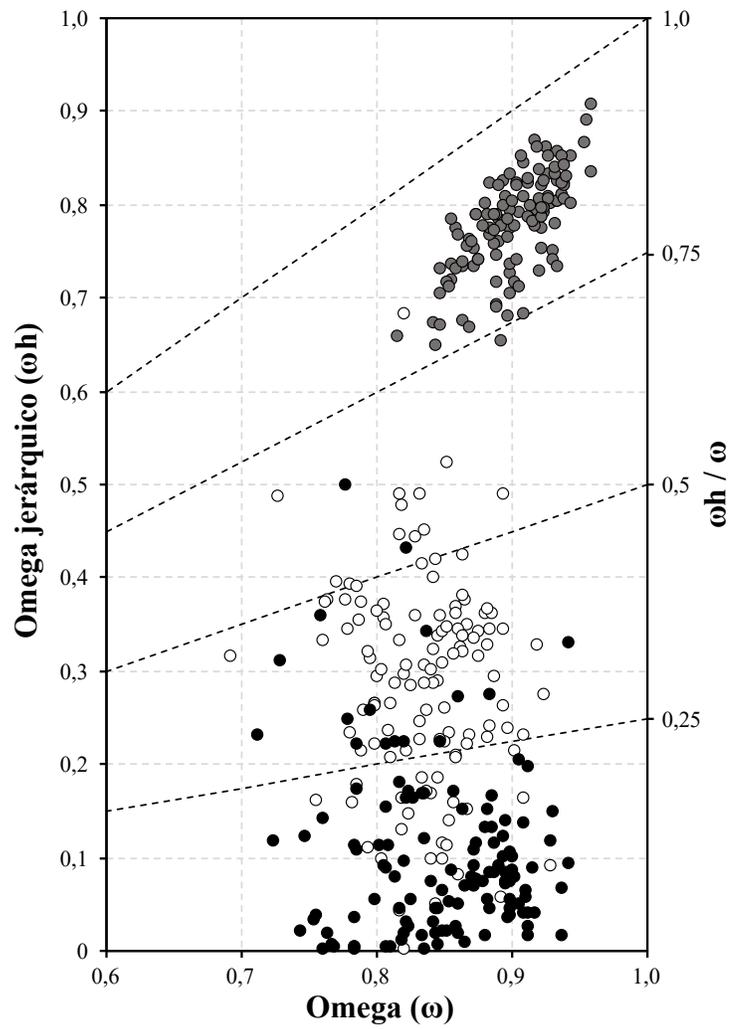


Figura 6. Valores del índice H por factor y modelo bifactor.

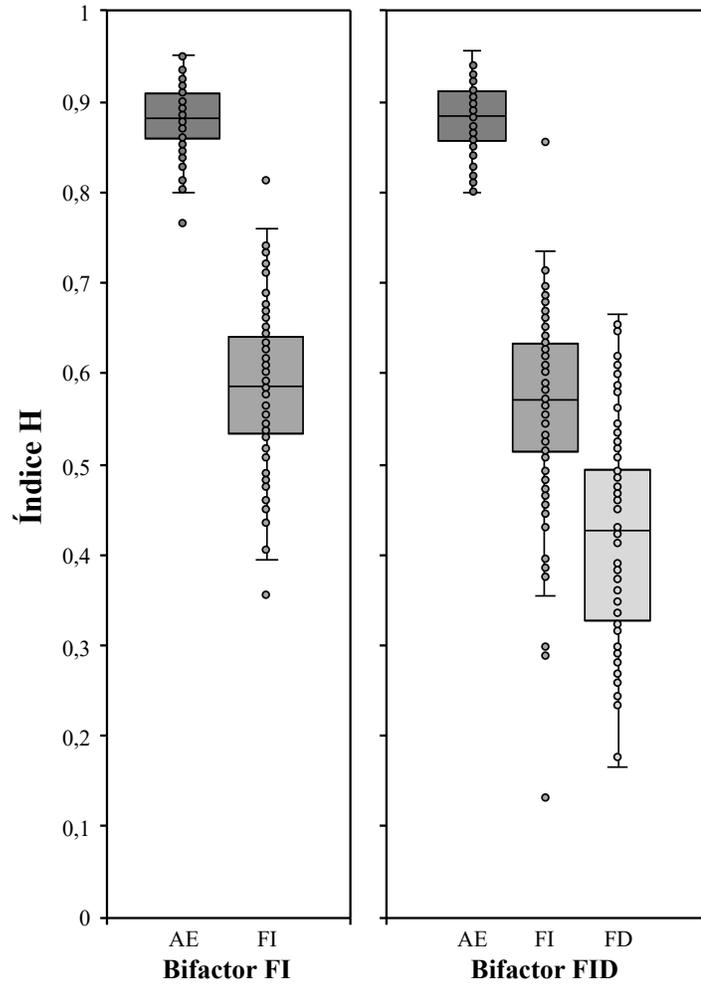


Figura 7. Valores de índice de determinación del factor (DF) por factor y modelo bifactor.

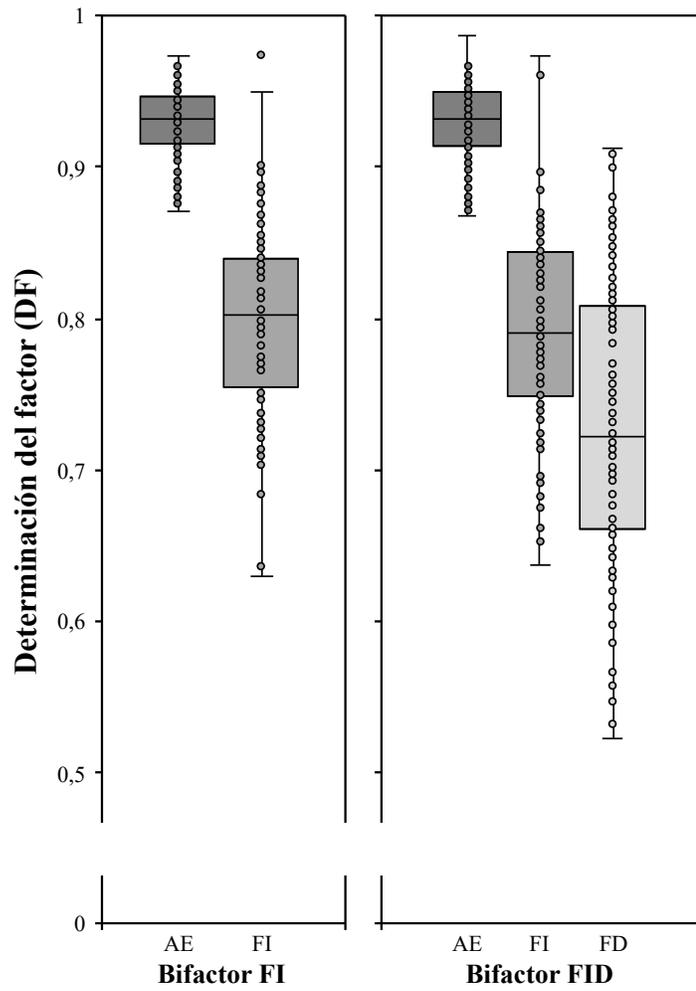


Figura 8. Panel izquierdo. Valores de varianza común explicada (ECV) por factor y modelo bifactor. Panel derecho. Relación entre los valores de varianza común explicada y sesgo de parámetro relativo promedio (ARPB) para los modelos Bifactor-FI (B-FI, círculos negros) y Bifactor-FID (B-FID, círculos blancos).

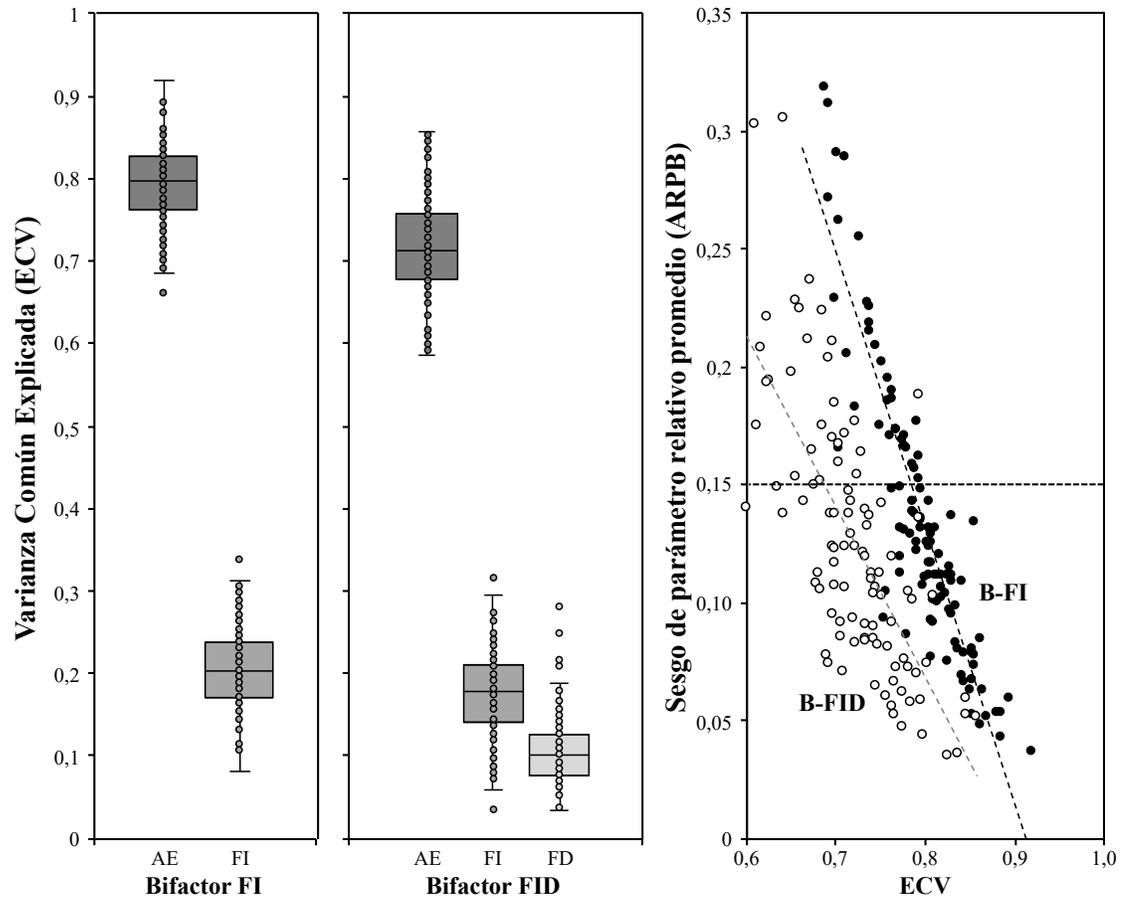
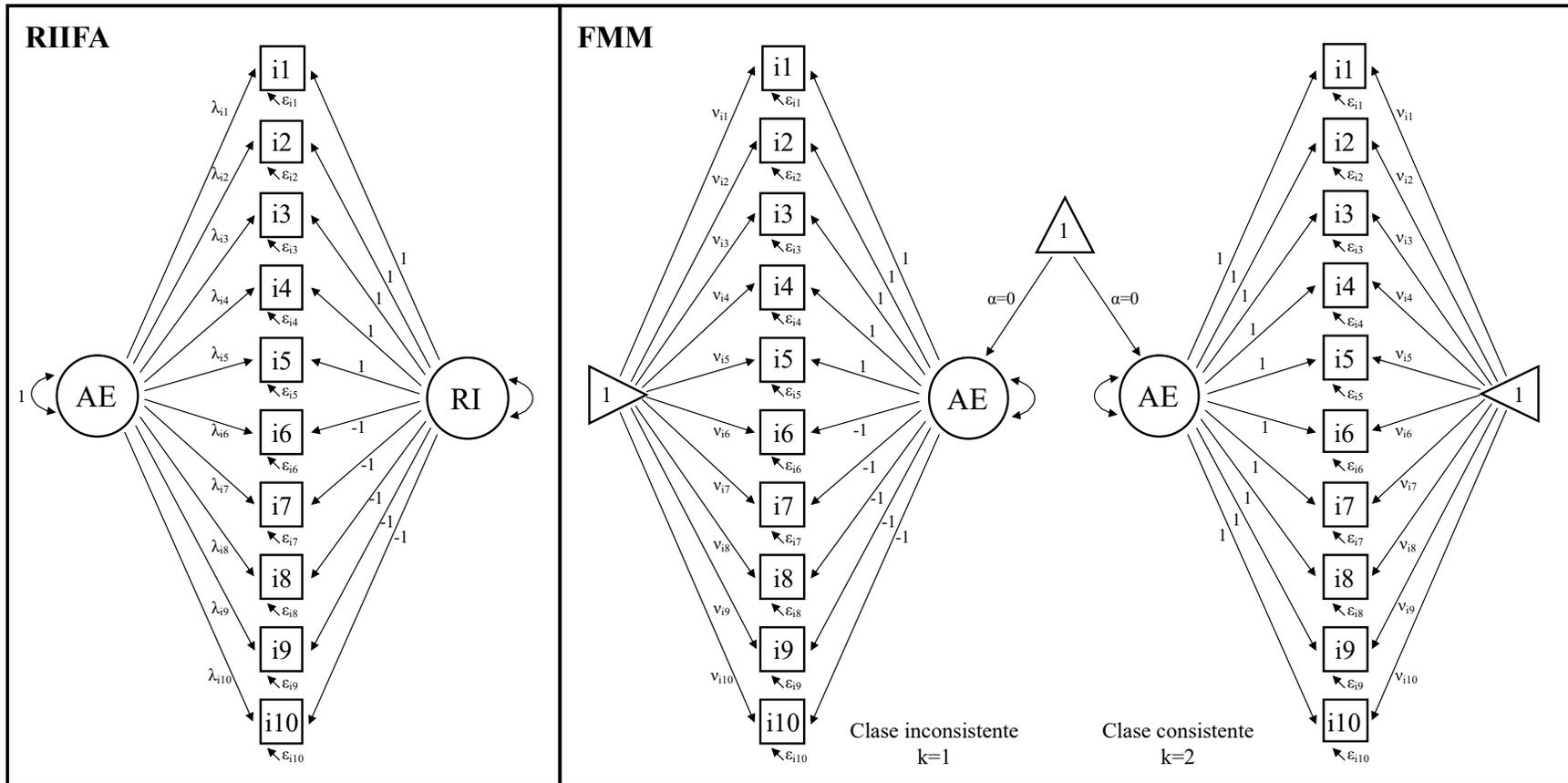


Figura 9. Modelos RIIFA y FMM asumiendo ítems inversos recodificados.



Material Suplementario

Sobre el uso y abuso del factor de redacción en la RSES: Una revisión sistemática y estimación de índices derivados.

Estimación de modelos bifactor

A partir de las bases de datos identificadas y los estudios que compartieron sus archivos de datos (ver Tabla 1), examinamos la dimensionalidad de la RSES, utilizando el modelo de un factor y los modelos Bifactor-FI y Bifactor-FID, tradicionalmente empleados para examinar los efectos de redacción (Alessandri et al., 2015; DiStefano & Motl, 2006; Gnambis et al., 2018; Marsh et al., 2010; Reise et al., 2016). Todos los modelos fueron estimados utilizando Robust Maximum Likelihood (MLR) como método de estimación y Full-Information Maximum Likelihood (FIML) como método para abordar la pérdida de datos.

La bondad de ajuste de los modelos estimados fue evaluada en todos los casos utilizando los índices CFI, TLI y RMSEA. Para los índices CFI y TLI, valores estimados mayores a 0,90 y 0,95 indican niveles de ajuste aceptables y buenos, respectivamente (Hu & Bentler, 1999; Marsh et al., 2004). Para los valores de RMSEA, valores iguales o inferiores a 0,05 y 0,08 fueron considerados como buenos y aceptables, respectivamente (Browne & Cudeck, 1992; Hu & Bentler, 1999). Todos los análisis fueron llevados a cabo utilizando Mplus v.7.4 (Muthén & Muthén, 2014).

Protocolo PRISMA.

Aquí, anexamos las listas de verificación de los criterios del protocolo PRISMA incorporados en el presente estudio (Tablas Suplementarias 1 y 2), el desglose por estudio de la evaluación del riesgo de sesgo de acuerdo a los criterios especificados en el manuscrito y la lista completa de los estudios y las magnitudes de los indicadores derivadas calculados a partir de las cargas factoriales estandarizadas (Tabla Suplementaria 3).

Tabla Suplementaria S1. Lista de verificación PRISMA.

Sección	Tema	Ítem	Descripción	Localiz.
Título	Título	1	Identificar la publicación/reporte como una revisión sistemática.	p.4
Resumen	Resumen	2	Vea la lista de verificación para resúmenes estructurados de la declaración PRISMA 2020 (ver, Tabla Suplementaria 2)	Tabla Supl.2
Introducción	Justificación del estudio	3	Describa la justificación de la revisión en el contexto del conocimiento existente.	p.6,18
	Objetivos	4	Proporcione una declaración explícita de los objetivos o las preguntas que aborda la revisión.	p.18
Método	Criterios de elegibilidad	5	Especifique los criterios de inclusión y exclusión de la revisión y cómo se agruparon los estudios para la síntesis.	p.20
	Fuentes de información	6	Especifique todas las bases de datos, registros, sitios web, organizaciones, listas de referencias y otros recursos de búsqueda o consulta para identificar los estudios. Especifique la fecha en la que cada recurso se buscó o consultó por última vez.	p.19-20
	Estrategia de búsqueda	7	Presente las estrategias de búsqueda completas de todas las bases de datos, registros y sitios web, incluyendo cualquier filtro y los límites utilizados.	p.19-20
	Proceso de selección de los estudios	8	Especifique los métodos utilizados para decidir si un estudio cumple con los criterios de inclusión de la revisión, incluyendo cuántos autores de la revisión revisaron cada registro y cada publicación recuperada, si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	p.20-21
	Proceso de extracción de los datos	9	Indique los métodos utilizados para extraer los datos de los registros o publicaciones, incluyendo cuántos revisores recopilaron datos de cada publicación, si trabajaron de manera independiente, los procesos para obtener o confirmar los datos por parte de los investigadores del estudio y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	p.20-21
	Lista de datos	10a	Enumere y defina todos los resultados para los que se buscaron los datos. Especifique si se buscaron todos los resultados compatibles con cada dominio del resultado (<i>e.g.</i> , para todas las escalas de medida, puntos temporales, análisis) y, de no ser así, los métodos utilizados para decidir los resultados que se debían recoger.	p.21
		10b	Enumere y defina todas las demás variables para las que se buscaron datos (<i>e.g.</i> , características de los participantes y de la intervención, fuentes de financiación). Describa todos los supuestos formulados sobre cualquier información ausente (<i>missing</i>) o incierta.	p.21
	Evaluación del riesgo de sesgo de los estudios	11	Especifique los métodos utilizados para evaluar el riesgo de sesgo de los estudios incluidos, incluyendo detalles de las herramientas utilizadas, cuántos autores de la revisión evaluaron cada estudio y si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	p.21
	Medidas del efecto	12	Especifique, para cada resultado, las medidas del efecto (<i>e.g.</i> , razón de riesgo, diferencia de medias) empleadas en la síntesis o presentación de los resultados.	p.21-23
	Métodos de síntesis	13a	Describa el proceso utilizado para decidir qué estudios eran elegibles para cada síntesis.	p.21-23
		13b	Describa los métodos requeridos para preparar los datos para su presentación o síntesis, tales como el manejo de los datos perdidos en los estadísticos de resumen o las conversiones de datos.	p.21-23

Sección	Tema	Ítem	Descripción	Localiz.
Método		13c	Describa los métodos utilizados para resumir o presentar visualmente los resultados de los estudios individuales y su síntesis.	p.21-23
	Evaluación del sesgo en la publicación	14	Describa los métodos utilizados para evaluar el riesgo de sesgo debido a resultados faltantes en una síntesis (derivados de los sesgos en las publicaciones).	N/A
Resultados	Evaluación de certeza	15	Describa los métodos utilizados para evaluar la certeza en el cuerpo de la evidencia para cada resultado.	N/A
	Selección de los estudios	16a	Describa los resultados de los procesos de búsqueda y selección, desde el número de registros identificados en la búsqueda hasta el número de estudios incluidos en la revisión, idealmente utilizando un diagrama de flujo.	Figura 2
		16b	Cite los estudios que aparentemente cumplían con los criterios de inclusión, pero que fueron excluidos, y explique por qué fueron excluidos.	p.20, 32
	Características de los estudios	17	Cite cada estudio incluido y presente sus características.	Tabla 1
	Riesgo de sesgo de los estudios individuales	18	Presente las evaluaciones del riesgo de sesgo para cada uno de los estudios incluidos.	p.23
	Resultados de los estudios individuales	19	Presente, para todos los resultados y para cada estudio: a) los estadísticos de resumen para cada grupo (si procede) y b) la estimación del efecto y su precisión (<i>e.g.</i> , intervalo de confianza), idealmente utilizando tablas estructuradas o gráficos.	Tabla Supl.3
	Resultados de la síntesis	20a	Para cada síntesis, resuma brevemente las características y el riesgo de sesgo entre los estudios contribuyentes.	Figura 3, Fig. Supl. 1
		20b	Presente los resultados de todas las síntesis estadísticas realizadas. Si se ha realizado un meta-análisis, presente para cada uno de ellos el estimador de resumen y su precisión (<i>e.g.</i> , intervalo de confianza) y las medidas de heterogeneidad estadística. Si se comparan grupos, describa la dirección del efecto.	p.24-26 Figura 4-8
		20c	Presente los resultados de todas las investigaciones sobre las posibles causas de heterogeneidad entre los resultados de los estudios.	N/A
		20d	Presente los resultados de todos los análisis de sensibilidad realizados para evaluar la robustez de los resultados sintetizados.	N/A
	Sesgos en la publicación	21	Presente las evaluaciones del riesgo de sesgo debido a resultados faltantes (derivados de los sesgos en las publicaciones) para cada síntesis evaluada.	N/A
	Certeza de la evidencia	22	Presente las evaluaciones de la certeza (o confianza) en el cuerpo de la evidencia para cada resultado evaluado.	N/A
	Discusión	Discusión	23a	Proporcione una interpretación general de los resultados en el contexto de otras evidencias.
23b			Discutir las limitaciones de la evidencia incluida en la revisión.	p.31-32
23c			Discutir las limitaciones de los procesos de revisión utilizados.	p.31-32
23d			Discutir las implicaciones de los resultados para la práctica, las políticas y las futuras investigaciones.	p.32-33
Información adicional	Registro y protocolo	24	Proporcione la información del registro de la revisión, incluyendo el nombre y el número de registro, o declare que la revisión no ha sido registrada.	p.19
	Financiamiento	25	Describa las fuentes de apoyo financiero o no financiero para la revisión y el papel de los financiadores o patrocinadores en la revisión.	p.33
	Conflictos de interés	26	Declare los conflictos de intereses de los autores de la revisión.	p.33

Sección	Tema	Ítem	Descripción	Localiz.
	Disponibilidad de datos, códigos y otros materiales	27	Especifique qué elementos de los que se indican a continuación están disponibles al público y dónde se pueden encontrar: Plantillas de formularios de extracción de datos, datos extraídos de los estudios incluidos, datos utilizados para todos los análisis, código de análisis, cualquier otro material utilizado en la revisión.	p.33

Tabla Suplementaria S2. Lista de verificación PRISMA 2021 para resúmenes estructurados.

Sección	Tema	Ítem	Descripción	Localiz.
Título	Título	1	Identificar la publicación/reporte como una revisión sistemática.	p.4
Antecedentes	Objetivos	2	Proporcione una declaración explícita de los objetivos o las preguntas que aborda la revisión.	p.5,lin.9-12
Método	Criterios de elegibilidad	3	Especifique los criterios de inclusión y exclusión de la revisión.	p.5, lin.12-13
	Fuentes de información	4	Especifique las fuentes de información (<i>e.g.</i> , bases de datos, registros) utilizadas para identificar los estudios y la fecha de la última búsqueda en cada una de estas fuentes.	p.5, lin.12-15
Resultados	Riesgo de sesgo de los estudios individuales	5	Especifique los métodos utilizados para evaluar el riesgo de sesgo de los estudios individuales incluidos.	
	Síntesis de los resultados	6	Especifique los métodos utilizados para presentar y sintetizar los resultados.	
	Estudios incluidos	7	Proporcione el número total de estudios incluidos y de participantes y resuma las características relevantes de los estudios.	p.5, lín.12
	Síntesis de los resultados	8	Presente los resultados principales e indique, preferiblemente, el número de estudios incluidos y los participantes en cada uno de ellos. Si se ha realizado un meta-análisis, indique el estimador de resumen y el intervalo de confianza. Si se comparan grupos, describa la dirección del efecto (<i>e.g.</i> , qué grupo se ha visto favorecido).	p.5, lín.12-16
Discusión	Limitaciones de la evidencia	9	Proporcione un breve resumen de las limitaciones de la evidencia incluida en la revisión (<i>e.g.</i> , riesgo de sesgo, inconsistencia –heterogeneidad– e imprecisión).	
Otros	Interpretación	10	Proporcione una interpretación general de los resultados y sus implicaciones importantes.	p.2, lín.17-19
	Financiación	11	Especifique la fuente principal de financiación de la revisión.	
	Registro	12	Proporcione el nombre y el número de registro.	p.2, lín.19

Figura Suplementaria S1. Representación tabular del riesgo de sesgo en estudios individuales.

Autores	Mod	Crit1	Crit2	Crit3	Crit4	Autores	Mod	Crit1	Crit2	Crit3	Crit4
Gana et al. (2005)	1	?	+	+	+	Silverstein & Bengtson (2019) ^e	73,74	+	+	+	+
DiStefano & Motl (2006)	2	?	?	?	+		75,76	+	+	+	+
Martin et al (2007)	3	?	?	+	?		77,78	+	+	+	+
CentERdata (2008) ^e	4,5	+	+	+	+		79,80	+	+	+	+
	6,7	+	+	+	+		81,82	+	+	+	+
	8,9	+	+	+	+	US Department of Labor (2019a) ^e	83,84	+	+	+	+
	10,11	+	+	+	+		85,86	+	+	+	+
	12,13	+	+	+	+	US Department of Labor (2019b) ^e	87,88	+	+	+	+
	14,15	+	+	+	+		89,90	+	+	+	+
Shimizu & Yoshida (2008)	16	?	?	?	?		91,92	+	+	+	+
DiStefano & Motl (2009b)	17,18	?	?	?	+		93,94	+	+	+	+
Barnett (2010) ^e	19	+	+	+	+		95,96	+	+	+	+
	20	+	+	+	+		97,98	+	+	+	+
	21	+	+	+	+	Vaughn (2019)	99	?	+	+	+
National Centre for Social Research (2011) ^e	22	+	+	+	+		100	?	+	+	+
	23	+	+	+	+	Yin (2019)	101	?	+	+	?
	24	+	+	+	+	Zhang et al. (2019) ^a	102	?	+	+	?
Welsh Assembly Government (2011) ^e	25	+	+	+	+	Bækker et al. (2020) ^a	103	?	+	?	?
Gana et al. (2013) ^b	26	+	?	+	+	Bueno et al. (2020)	104	?	+	+	+
Hyland et al. (2014)	27	+	+	?	+	Donnellan & Rakhshani (2020) ^a	105	-	+	+	?
McKay et al. (2014)	28	+	+	+	+		106	-	+	+	?
Sallay et al. (2014)	29	+	?	+	+		107	-	+	+	?
Sharratt et al. (2014)	30	+	?	+	+		108	-	+	+	?
Urbán et al. (2014)	31	?	+	+	+		109	?	+	+	?
	32	?	+	+	+		110	?	+	+	?
Kocalevent et al. (2015) ^a	33	+	+	+	+	Gnams & Schroeders (2020)	111	+	+	+	+
Tomás, Sancho et al. (2015)	34-37	?	+	?	+	Golec de Zavala et al. (2020) ^a	112	+	+	+	+
Tomás, Oliver et al. (2015)	38,39	?	+	?	+		113	+	+	+	+
Demetrovics et al. (2016) ^a	40	+	+	+	+		114	+	+	+	+
Donnellan et al. (2016) ^a	41	?	+	+	?	Hussey & Hughes (2020) ^a	115	+	+	+	+
Gómez et al. (2016)	42	?	+	?	?	Jabłońska & Zajdel (2020) ^a	116	?	+	+	?
	43	?	+	+	?	Nguyen (2020) ^a	117	?	+	+	+
Michaelides, Koutsogiorgi et al. (2016)	44	-	+	?	+	Wehner et al. (2020) ^a	118	?	+	+	+
Michaelides, Zenger et al. (2016)	45	+	+	+	+	Deutsches Zentrum für Altersfragen (2021) ^e	119,120	+	+	+	+
Reise et al. (2016), Kam (2021) ^a	46	+	+	+	+		121,122	+	+	+	+
Wang et al. (2016)	47,48	?	+	+	+		123,124	+	+	+	+
Xu & Leung (2016)	49,50	?	?	+	+		125,126	+	+	+	+
Zhang & Savalei (2016)	51	?	+	+	?	Dueber et al. (2021) ^a	127	+	+	+	+
Alessandri et al. (2017)	52	?	+	+	?		128	+	+	+	+
Corker et al. (2017) ^a	53	?	+	+	?	García et al. (2021) ^a	129	?	+	+	?
Salerno et al. (2017)	54,55	+	+	+	+	Lifshin et al. (2021) ^a	130	?	+	+	?
Shin & Lee (2017)	56	-	+	+	+	Lo Coco et al. (2021)	131	?	+	+	?
Wu et al. (2017)	57	?	+	?	+	Lodi-Smith et al. (2021) ^e	132	?	+	?	?
Gnams et al. (2018)	58	+	+	+	+	Medeiro et al. (2021)	133	?	?	+	-
Goldberg (2018) ^e	59	+	+	?	?	Polner et al. (2021) ^a	134	?	+	+	+
Holden et al. (2018) ^a	60	?	+	+	?	Ponce et al. (2021) ^a	135	-	+	+	?
Masselink et al. (2018) ^a	61	+	+	+	+	Rufino & Holden (2021) ^a	136	-	+	+	-
	62	+	+	+	+	Scheve & Stasavage (2021) ^a	137	+	+	+	+
	63	+	+	+	+	Skylark & Callan (2021) ^a	138	+	+	+	+
Orsini et al. (2018) ^a	64	?	+	?	?	Yovel et al. (2021) ^a	139	?	+	+	+
Portes & Rumbaut (2018) ^e	65,66	+	+	+	+		140	?	+	+	+
Stafford et al (2018) ^a	67	+	+	+	+	Zitely et al. (2021) ^a	141	?	+	+	+
	68	+	+	+	?		142	?	+	+	+
	69	+	+	+	?		143	?	+	+	+
Ventura et al. (2018) ^a	70	?	+	+	?		144	?	+	+	+
Csajbók et al. (2019)	71	?	+	+	+						
Ortuño et al. (2019) ^a	72	?	+	?	?						

Tabla Suplementaria S3. Resumen indicadores derivados de los modelos bifactor recolectados.

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB		
Gana et al. (2005)		1	B-FID	AE	0.649	0.930	0.752	0.808	0.903	0.930	0.556			
				FI	0.199	0.865	0.376	0.435	0.643	0.845				
				FID	0.152	0.911	0.197	0.217	0.593	0.824				
DiStefano & Motl (2006)		2	B-FI	AE	0.769	0.894	0.777	0.868	0.872	0.924	0.778	0.173		
				FI	0.231	0.832	0.379	0.456	0.607	0.790				
				AE	0.998	0.879	0.878	0.999	0.915	0.956			0.778	
Martín-Albo et al. (2007)		3	B-FI	FI	0.002	0.868	0.002	0.003	0.008	0.097	0.778			
				AE	0.762	0.909	0.791	0.870	0.895	0.938			0.778	0.170
				FID	0.238	0.864	0.376	0.436	0.666	0.851				
CentERdata (2008)	2009, Hombres	4	B-FID	AE	0.674	0.922	0.776	0.842	0.905	0.948	0.556	0.164		
				FI	0.215	0.863	0.383	0.443	0.668	0.858				
				FID	0.111	0.884	0.085	0.096	0.478	0.800				
				AE	0.800	0.919	0.821	0.894	0.912	0.949			0.778	0.111
				FI	0.200	0.862	0.330	0.383	0.623	0.836				
				FID	0.110	0.901	0.101	0.113	0.510	0.839				
	2009, Mujeres	5	B-FI	AE	0.710	0.929	0.801	0.862	0.917	0.951	0.556	0.106		
				FI	0.180	0.862	0.325	0.377	0.618	0.835				
				FID	0.110	0.901	0.101	0.113	0.510	0.839				
				AE	0.805	0.913	0.821	0.900	0.899	0.939			0.778	0.132
				FI	0.195	0.864	0.302	0.350	0.604	0.826				
				FID	0.121	0.895	0.074	0.082	0.531	0.870				
2011, Hombres	6	B-FI	AE	0.700	0.927	0.808	0.871	0.909	0.953	0.556	0.137			
			FI	0.179	0.863	0.321	0.372	0.617	0.845					
			FID	0.121	0.895	0.074	0.082	0.531	0.870					
			AE	0.796	0.922	0.821	0.890	0.913	0.948			0.778	0.131	
			FI	0.204	0.876	0.336	0.384	0.627	0.839					
			FID	0.101	0.901	0.081	0.090	0.480	0.830					
2011, Mujeres	7	B-FI	AE	0.712	0.933	0.805	0.863	0.920	0.955	0.556	0.123			
			FI	0.187	0.876	0.342	0.391	0.631	0.846					
			FID	0.101	0.901	0.081	0.090	0.480	0.830					
			AE	0.788	0.910	0.808	0.888	0.897	0.937			0.778	0.157	
			FI	0.212	0.867	0.329	0.379	0.624	0.833					
			FID	0.104	0.881	0.057	0.065	0.476	0.842					
2013, Hombres	8	B-FI	AE	0.697	0.923	0.798	0.864	0.904	0.950	0.556	0.169			
			FI	0.198	0.866	0.350	0.405	0.639	0.850					
			FID	0.104	0.881	0.057	0.065	0.476	0.842					
			AE	0.819	0.928	0.839	0.904	0.923	0.955			0.778	0.101	
			FI	0.181	0.876	0.307	0.350	0.598	0.832					
			FID	0.172	0.875	0.317	0.363	0.607	0.840					
2013, Mujeres	9	B-FID	AE	0.743	0.936	0.824	0.880	0.928	0.959	0.556	0.103			
			FI	0.172	0.875	0.317	0.363	0.607	0.840					
			FID	0.084	0.910	0.065	0.072	0.432	0.817					

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB		
CentERdata (2008)	2017, Hombres	10	B-FI	AE	0.789	0.910	0.808	0.888	0.902	0.942	0.778	0.137		
				FI	0.211	0.860	0.333	0.387	0.625	0.834				
			B-FID	AE	0.695	0.923	0.791	0.857	0.910	0.951			0.556	0.137
				FI	0.195	0.860	0.346	0.402	0.634	0.845				
	2017, Mujeres	11	B-FI	AE	0.810	0.926	0.832	0.899	0.924	0.956	0.778	0.101		
				FI	0.190	0.873	0.319	0.366	0.614	0.841				
			B-FID	AE	0.740	0.936	0.826	0.882	0.929	0.965	0.556	0.112		
				FI	0.181	0.872	0.335	0.384	0.626	0.854				
	2019, Hombres	12	B-FI	AE	0.787	0.922	0.816	0.885	0.910	0.946	0.778	0.158		
				FI	0.213	0.880	0.346	0.393	0.646	0.858				
			B-FID	AE	0.705	0.933	0.805	0.863	0.917	0.956	0.556	0.167		
				FI	0.201	0.880	0.362	0.412	0.658	0.871				
	2019, Mujeres	13	B-FI	AE	0.803	0.930	0.832	0.895	0.921	0.953	0.778	0.125		
				FI	0.197	0.883	0.330	0.374	0.635	0.857				
			B-FID	AE	0.737	0.938	0.824	0.878	0.927	0.963	0.556	0.132		
				FI	0.189	0.883	0.345	0.391	0.646	0.870				
	2021, Hombres	14	B-FI	AE	0.792	0.928	0.822	0.886	0.915	0.949	0.778	0.152		
				FI	0.208	0.886	0.352	0.397	0.637	0.848				
			B-FID	AE	0.723	0.937	0.812	0.867	0.922	0.959	0.556	0.154		
				FI	0.197	0.885	0.363	0.410	0.646	0.861				
2021, Mujeres	15	B-FI	AE	0.790	0.929	0.825	0.888	0.925	0.957	0.778	0.125			
			FI	0.210	0.882	0.352	0.399	0.659	0.872					
		B-FID	AE	0.732	0.938	0.821	0.875	0.930	0.967	0.556	0.139			
			FI	0.202	0.882	0.368	0.417	0.670	0.885					
Shimizu & Yoshida (2008)	16	B-FI	AE	0.797	0.855	0.766	0.895	0.857	0.917	0.778	0.107			
			FI	0.203	0.751	0.309	0.412	0.562	0.789					
DiStefano & Motl (2009)	Hombres	17	B-FI	AE	0.763	0.896	0.774	0.864	0.875	0.927	0.778			
	Mujeres	18	B-FI	AE	0.761	0.900	0.779	0.865	0.876	0.926	0.778			
				FI	0.239	0.847	0.388	0.459	0.633	0.809				

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB	
Barnett (2010)	1989, Ola 1	19	B-FI	AE	0.852	0.847	0.798	0.942	0.853	0.917	0.778	0.052	
				FI	0.148	0.756	0.162	0.215	0.435	0.686			
		1990, Ola 2	20	B-FID	AE	0.745	0.859	0.774	0.901	0.857	0.918	0.556	0.064
					FI	0.131	0.756	0.162	0.215	0.432	0.684		
	1991, Ola 3		21	B-FI	FID	0.124	0.784	0.114	0.146	0.428	0.687		
					AE	0.818	0.856	0.776	0.906	0.862	0.919	0.778	0.106
		1999, Ola 1	22	B-FID	FI	0.182	0.787	0.250	0.318	0.491	0.721		
					AE	0.682	0.872	0.733	0.841	0.862	0.913	0.556	0.105
	2000, Ola 2		23	B-FI	FI	0.138	0.788	0.214	0.272	0.456	0.696		
					FID	0.180	0.808	0.223	0.276	0.566	0.763		
		2001, Ola 2	24	B-FI	AE	0.853	0.874	0.820	0.938	0.870	0.923	0.778	0.067
					FI	0.147	0.818	0.174	0.213	0.459	0.709		
2001, Ola 2	25		B-FID	AE	0.733	0.887	0.790	0.890	0.875	0.926	0.556	0.084	
				FI	0.124	0.819	0.166	0.202	0.445	0.698			
	National Center for Social Research (2005)	22	B-FI	FID	0.143	0.807	0.156	0.193	0.494	0.738			
				AE	0.701	0.882	0.723	0.821	0.850	0.909	0.778	0.291	
2000, Ola 2		23	B-FID	FI	0.299	0.841	0.462	0.549	0.668	0.820			
				AE	0.622	0.889	0.693	0.780	0.845	0.888	0.556	0.193	
	2001, Ola 2	24	B-FI	FI	0.242	0.842	0.402	0.477	0.625	0.778			
				FID	0.135	0.816	0.182	0.223	0.479	0.686			
Welsh Assembly Government (2011)		25	B-FI	AE	0.736	0.892	0.755	0.847	0.864	0.917	0.778	0.226	
				FI	0.264	0.847	0.414	0.489	0.647	0.814			
	Gana et al. (2013)	26	B-FID	AE	0.656	0.899	0.728	0.810	0.865	0.905	0.556	0.153	
				FI	0.213	0.847	0.359	0.424	0.604	0.777			
2001, Ola 2		24	B-FI	FID	0.131	0.835	0.170	0.203	0.482	0.705			
				AE	0.739	0.891	0.757	0.849	0.862	0.916	0.778	0.225	
	2001, Ola 2	25	B-FID	FI	0.261	0.845	0.410	0.486	0.640	0.807			
				AE	0.665	0.898	0.736	0.819	0.860	0.899	0.556	0.143	
2001, Ola 2		26	B-FI	FI	0.201	0.845	0.338	0.400	0.583	0.756			
				FID	0.134	0.834	0.170	0.204	0.489	0.709			
	2001, Ola 2	27	B-FI	AE	0.774	0.889	0.776	0.873	0.867	0.920	0.778	0.169	
				FI	0.226	0.829	0.357	0.431	0.591	0.782			
2001, Ola 2		28	B-FID	AE	0.711	0.898	0.775	0.863	0.870	0.926	0.556	0.171	
				FI	0.208	0.829	0.361	0.435	0.591	0.786			
	2001, Ola 2	29	B-FI	FID	0.081	0.842	0.032	0.038	0.364	0.751			
				AE	0.609	0.864	0.677	0.783	0.835	0.893	0.556		
2001, Ola 2		30	B-FID	FI	0.239	0.770	0.395	0.514	0.629	0.806			
				FID	0.152	0.823	0.172	0.209	0.492	0.714			

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB
Hyland et al. (2014)		27	B-FID	AE	0.761	0.873	0.789	0.904	0.852	0.917	0.556	
				FI	0.193	0.799	0.265	0.332	0.560	0.774		
				FID	0.046	0.783	0.006	0.008	0.181	0.531		
McKay et al. (2014)		28	B-FID	AE	0.761	0.873	0.789	0.904	0.852	0.917	0.556	
				FI	0.193	0.799	0.265	0.332	0.560	0.774		
				FID	0.046	0.783	0.006	0.008	0.181	0.531		
Sallay et al. (2014)		29	B-FID	AE	0.585	0.897	0.682	0.760	0.866	0.909	0.556	
				FI	0.250	0.843	0.420	0.498	0.686	0.844		
				FID	0.165	0.846	0.225	0.266	0.590	0.773		
Sharrat et al. (2014)		30	B-FID	AE	0.611	0.869	0.670	0.770	0.831	0.886	0.556	
				FI	0.107	0.840	0.099	0.118	0.432	0.718		
				FID	0.282	0.776	0.501	0.645	0.655	0.806		
Urban et al. (2014)	Ola 1	31	B-FID	AE	0.661	0.889	0.718	0.807	0.875	0.930	0.556	
				FI	0.274	0.819	0.478	0.584	0.669	0.836		
				FID	0.065	0.843	0.046	0.055	0.271	0.609		
	Ola 2	32	B-FID	AE	0.618	0.899	0.705	0.784	0.882	0.922	0.556	
				FI	0.256	0.829	0.445	0.537	0.695	0.860		
				FID	0.126	0.863	0.152	0.176	0.500	0.746		
Kocalevent et al. (2015)		33	B-FI	AE	0.727	0.898	0.755	0.841	0.862	0.914	0.778	0.255
				FI	0.273	0.864	0.427	0.494	0.661	0.818		
				FID	0.120	0.827	0.165	0.199	0.452	0.669		
			B-FID	AE	0.675	0.903	0.742	0.821	0.856	0.887	0.556	0.150
				FI	0.205	0.864	0.338	0.391	0.586	0.742		
				FID	0.120	0.827	0.165	0.199	0.452	0.669		
Tomas, Sancho et al. (2015)	Muestra 1, Hombres	34	B-FI	AE	0.811	0.842	0.766	0.910	0.851	0.919	0.778	
				FI	0.189	0.661	0.285	0.431	0.512	0.736		
	Muestra 1, Mujeres	35	B-FI	AE	0.812	0.862	0.815	0.945	0.884	0.938	0.778	
				FI	0.188	0.678	0.193	0.285	0.721	0.975		
	Muestra 2, Hombres	36	B-FI	AE	0.704	0.829	0.670	0.808	0.817	0.898	0.778	
				FI	0.296	0.745	0.479	0.643	0.619	0.795		
	Muestra 2, Mujeres	37	B-FI	AE	0.806	0.858	0.815	0.950	0.880	0.937	0.778	
				FI	0.194	0.710	0.170	0.240	0.747	0.978		
Tomas, Oliver et al. (2015)	Hombres	38	B-FI	AE	0.718	0.806	0.709	0.880	0.804	0.891	0.778	
				FI	0.282	0.668	0.328	0.491	0.814	0.949		
	Mujeres	39	B-FI	AE	0.832	0.866	0.809	0.935	0.885	0.938	0.778	
				FI	0.168	0.709	0.219	0.309	0.610	0.850		

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB
Demetrovics et al. (2016)		40	B-FI	AE	0.704	0.882	0.719	0.815	0.871	0.927	0.778	0.262
				FI	0.296	0.823	0.495	0.601	0.672	0.833		
			B-FID	AE	0.625	0.888	0.690	0.777	0.876	0.916	0.556	0.194
				FI	0.125	0.840	0.168	0.200	0.458	0.682		
				FD	0.250	0.821	0.432	0.526	0.647	0.819		
Donnellan et al. (2016)		41	B-FI	AE	0.764	0.905	0.794	0.877	0.890	0.936	0.778	0.148
				FI	0.236	0.850	0.368	0.433	0.674	0.852		
			B-FID	AE	0.698	0.915	0.787	0.860	0.897	0.940	0.556	0.116
				FI	0.203	0.849	0.343	0.405	0.652	0.845		
				FD	0.098	0.873	0.077	0.088	0.427	0.766		
Gómez-Lugo et al. (2016)	Colombia	42	B-FID	AE	0.854	0.883	0.823	0.932	0.908	0.936	0.556	
				FI	0.075	0.819	0.131	0.160	0.288	0.675		
				FD	0.071	0.785	0.109	0.139	0.268	0.549		
	España	43	B-FID	AE	0.759	0.895	0.808	0.902	0.880	0.921	0.556	
	FI			0.098	0.861	0.081	0.094	0.466	0.899			
	FD			0.143	0.785	0.222	0.282	0.507	0.758			
Michaelides, Koutsogiorgi et al. (2016)		44	B-FID	AE	0.693	0.918	0.795	0.866	0.919	0.956	0.556	
				FI	0.173	0.835	0.308	0.368	0.608	0.844		
				FD	0.134	0.894	0.102	0.114	0.626	0.871		
Michaelides, Zenger et al. (2016)		45	B-FID	AE	0.693	0.889	0.747	0.840	0.872	0.912	0.556	
				FI	0.143	0.831	0.228	0.274	0.473	0.690		
				FD	0.165	0.814	0.226	0.278	0.581	0.797		
Reise et al. (2016) Kam (2020)		46	B-FI	AE	0.855	0.926	0.863	0.932	0.917	0.951	0.778	0.134
				FI	0.145	0.870	0.224	0.258	0.540	0.806		
			B-FID	AE	0.794	0.933	0.857	0.918	0.920	0.957	0.556	0.136
				FI	0.138	0.869	0.232	0.267	0.546	0.816		
				FD	0.068	0.898	0.038	0.042	0.328	0.732		
Wang et al. (2016)	Estudio 1	47	B-FID	AE	0.721	0.921	0.805	0.874	0.915	0.964	0.556	
				FI	0.221	0.859	0.371	0.432	0.714	0.899		
				FD	0.058	0.880	0.018	0.020	0.281	0.756		
	Estudio 2	48	B-FID	AE	0.799	0.943	0.853	0.904	0.940	0.958	0.556	
	FI			0.157	0.893	0.263	0.295	0.619	0.865			
	FD			0.044	0.910	0.058	0.064	0.246	0.556			

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB	
Xu & Leung (2016)	Estimador=ML	49	B-FI	AE	0.733	0.850	0.740	0.870	0.820	0.891	0.778		
				FI	0.267	0.790	0.337	0.427	0.645	0.815			
			B-FID	AE	0.696	0.891	0.760	0.853	0.853	0.898			0.556
				FI	0.227	0.843	0.324	0.384	0.663	0.832			
				FD	0.076	0.805	0.091	0.113	0.319	0.597			
				AE	0.743	0.885	0.773	0.873	0.859	0.915			
	B-FID	FI	0.257	0.836	0.355	0.425	0.689	0.855					
		AE	0.677	0.856	0.720	0.841	0.814	0.871	0.556				
	B-FID	FI	0.229	0.801	0.295	0.368	0.612	0.785					
		FD	AE	0.094	0.747	0.123	0.165	0.325	0.565				
	Zhang & Savalei (2016)		51	B-FI	AE	0.816	0.935	0.848	0.907	0.939	0.967	0.778	0.120
		FI			0.184	0.855	0.328	0.384	0.652	0.869			
Alessandri et al. (2017)	52	B-FID	AE	0.707	0.872	0.753	0.864	0.858	0.926	0.556			
			FI	0.223	0.763	0.376	0.493	0.610	0.802				
			FD	0.070	0.833	0.017	0.020	0.262	0.633				
Corker et al. (2017)	53	B-FI	AE	0.692	0.912	0.740	0.812	0.891	0.938	0.778	0.271		
			FI	0.308	0.863	0.532	0.616	0.743	0.877				
		B-FID	AE	0.635	0.919	0.730	0.794	0.877	0.908	0.556		0.148	
			FI	0.236	0.863	0.426	0.494	0.679	0.829				
Salerno et al. (2017)	Estimador=MLR	54	B-FID	AE	0.672	0.886	0.758	0.855	0.871	0.941	0.556		
				FI	0.214	0.780	0.394	0.504	0.631	0.833			
				FD	0.114	0.859	0.028	0.032	0.412	0.784			
Estimador=RWLS	55	B-FID	AE	0.672	0.917	0.777	0.848	0.903	0.957	0.556			
			FI	0.214	0.833	0.415	0.498	0.671	0.874				
			FD	0.114	0.899	0.046	0.051	0.465	0.836				
Shin & Lee (2017)	56	B-FID	AE	0.793	0.894	0.826	0.924	0.905	0.950	0.556	0.188		
			FI	0.168	0.821	0.214	0.261	0.642	0.961				
			FD	0.038	0.820	0.019	0.023	0.178	0.522				
Wu et al. (2017)	57	B-FID	AE	0.592	0.892	0.655	0.735	0.875	0.936	0.556			
			FI	0.317	0.820	0.685	0.835	0.855	0.974				
			FD	0.091	0.883	0.045	0.051	0.460	0.900				
Gnambs et al. (2018)	58	B-FI	AE	0.806	0.882	0.791	0.897	0.865	0.919	0.778	0.125		
			FI	0.194	0.813	0.295	0.363	0.539	0.751				
		B-FID	AE	0.749	0.889	0.784	0.882	0.869	0.920	0.556		0.112	
			FI	0.176	0.813	0.287	0.353	0.531	0.748				
			FD	0.074	0.826	0.057	0.069	0.301	0.621				

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB
Goldberg (2018)		59	B-FI	AE	0.836	0.897	0.824	0.918	0.902	0.943	0.778	0.080
				FI	0.164	0.833	0.246	0.295	0.522	0.766		
			B-FID	AE	0.706	0.914	0.799	0.875	0.908	0.947	0.556	0.085
				FI	0.140	0.832	0.248	0.298	0.519	0.768		
Holden et al. (2018)		60	B-FI	AE	0.806	0.910	0.824	0.905	0.908	0.947	0.778	0.092
				FI	0.194	0.840	0.301	0.358	0.618	0.846		
			B-FID	AE	0.719	0.921	0.807	0.876	0.916	0.953	0.556	0.093
				FI	0.176	0.840	0.303	0.360	0.615	0.848		
Masselink et al. (2018)	Ola 1	61	B-FI	AE	0.804	0.898	0.807	0.898	0.888	0.933	0.778	0.116
				FI	0.196	0.845	0.298	0.352	0.577	0.783		
		B-FID	AE	0.745	0.905	0.793	0.876	0.890	0.932	0.556	0.105	
			FI	0.182	0.846	0.291	0.344	0.575	0.781			
Ola 2	62	B-FI	AE	0.796	0.907	0.805	0.888	0.894	0.936	0.778	0.135	
			FI	0.204	0.857	0.328	0.383	0.592	0.795			
		B-FID	AE	0.734	0.913	0.788	0.863	0.896	0.934	0.556	0.119	
			FI	0.189	0.857	0.319	0.372	0.588	0.791			
Ola 3	63	B-FI	AE	0.812	0.899	0.813	0.904	0.887	0.933	0.778	0.111	
			FI	0.188	0.842	0.281	0.334	0.569	0.784			
		B-FID	AE	0.700	0.916	0.782	0.854	0.892	0.929	0.556	0.107	
			FI	0.166	0.842	0.287	0.341	0.573	0.783			
Orsini et al. (2018)		64	B-FI	AE	0.745	0.834	0.702	0.841	0.831	0.895	0.778	0.209
				FI	0.255	0.798	0.357	0.447	0.565	0.752		
			B-FID	AE	0.683	0.841	0.675	0.803	0.826	0.877	0.556	0.151
				FI	0.175	0.798	0.263	0.330	0.473	0.661		
Portes & Rumbaut (2018)	Hombres	65	B-FI	AE	0.713	0.837	0.706	0.843	0.813	0.893	0.778	0.205
				FI	0.287	0.754	0.404	0.536	0.642	0.806		
			B-FID	AE	0.612	0.847	0.673	0.794	0.800	0.868	0.556	0.175
				FI	0.171	0.782	0.160	0.204	0.561	0.788		
				FD	0.217	0.759	0.359	0.474	0.541	0.721		

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB	
Portes & Rumbaut (2018)	Mujeres	66	B-FI	AE	0.722	0.840	0.724	0.861	0.815	0.892	0.778	0.183	
				FI	0.278	0.764	0.359	0.470	0.647	0.818			
			B-FID	AE	0.650	0.852	0.719	0.843	0.824	0.906	0.556	0.197	
				FI	0.254	0.762	0.375	0.493	0.651	0.825			
Stafford et al. (2018)	Ola 1	67	B-FI	AE	0.825	0.871	0.805	0.924	0.864	0.922	0.778	0.075	
				FI	0.175	0.782	0.226	0.289	0.540	0.773			
			B-FID	AE	0.759	0.880	0.802	0.911	0.869	0.930	0.556	0.080	
				FI	0.162	0.780	0.234	0.299	0.535	0.772			
	Ola 2	68	B-FI	AE	0.852	0.895	0.831	0.929	0.885	0.932	0.778	0.078	
				FI	0.148	0.818	0.220	0.269	0.491	0.736			
	Ola 3	69	B-FI	AE	0.852	0.913	0.848	0.928	0.904	0.943	0.778	0.080	
				FI	0.148	0.852	0.227	0.267	0.510	0.755			
	Ventura et al. (2018)		70	B-FI	AE	0.698	0.833	0.674	0.809	0.829	0.904	0.778	0.229
					FI	0.302	0.726	0.497	0.684	0.643	0.819		
				B-FID	AE	0.616	0.843	0.649	0.771	0.829	0.898	0.556	0.208
					FI	0.270	0.726	0.487	0.670	0.635	0.812		
Csajbók et al. (2019)		71	B-FID	AE	0.747	0.934	0.826	0.884	0.915	0.941	0.556		
				FI	0.145	0.882	0.229	0.260	0.558	0.808			
				FD	0.108	0.894	0.141	0.157	0.473	0.741			
Ortuño & Cabana (2019)		72	B-FI	AE	0.773	0.791	0.685	0.866	0.840	0.910	0.778	0.132	
				FI	0.227	0.692	0.312	0.451	0.482	0.709			
			B-FID	AE	0.641	0.815	0.661	0.811	0.845	0.914	0.556	0.137	
				FI	0.192	0.691	0.317	0.459	0.487	0.713			
Silverstein & Bengtson (2019)	1988, Hombres	73	B-FI	AE	0.813	0.880	0.788	0.895	0.884	0.933	0.778	0.100	
				FI	0.187	0.795	0.308	0.387	0.511	0.732			
			B-FID	AE	0.681	0.897	0.765	0.853	0.891	0.939	0.556	0.112	
				FI	0.162	0.795	0.313	0.394	0.516	0.739			
	1988, Mujeres	74	B-FI	AE	0.835	0.884	0.808	0.914	0.882	0.933	0.778	0.082	
				FI	0.165	0.790	0.264	0.334	0.492	0.722			
			B-FID	AE	0.722	0.898	0.794	0.884	0.888	0.937	0.556	0.083	
				FI	0.142	0.790	0.259	0.328	0.488	0.722			
			FD	0.136	0.871	0.092	0.106	0.540	0.835				

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB		
Silverstein & Bengtson (2019)	1991, Hombres	75	B-FI	AE	0.845	0.846	0.783	0.926	0.846	0.909	0.778	0.078		
				FI	0.155	0.777	0.192	0.247	0.438	0.684				
			B-FID	AE	0.765	0.855	0.785	0.918	0.840	0.900	0.556	0.052		
				FI	0.128	0.794	0.111	0.140	0.475	0.773				
	1991, Mujeres	76	B-FI	AE	0.885	0.857	0.815	0.952	0.857	0.917	0.778	0.052		
				FI	0.115	0.788	0.132	0.168	0.357	0.636				
			B-FID	No-Converge										
				AE	0.807	0.869	0.777	0.894	0.865	0.917	0.778	0.128		
	1994, Hombres	77	B-FI	FI	0.193	0.820	0.276	0.337	0.509	0.723				
				AE	0.716	0.883	0.769	0.871	0.870	0.928	0.556	0.147		
			B-FID	FI	0.183	0.820	0.298	0.363	0.531	0.747				
				FD	0.101	0.799	0.055	0.069	0.375	0.697				
	1994, Mujeres	78	B-FI	AE	0.813	0.880	0.794	0.902	0.874	0.924	0.778	0.111		
				FI	0.187	0.826	0.269	0.326	0.533	0.755				
			B-FID	AE	0.739	0.893	0.799	0.895	0.875	0.934	0.556	0.137		
				FI	0.176	0.826	0.285	0.345	0.546	0.770				
	1997, Hombres	79	B-FI	AE	0.786	0.859	0.761	0.885	0.841	0.904	0.778	0.143		
				FI	0.214	0.794	0.305	0.384	0.534	0.733				
			B-FID	AE	0.730	0.869	0.764	0.880	0.844	0.915	0.556	0.163		
				FI	0.204	0.794	0.320	0.404	0.545	0.747				
	1997, Mujeres	80	B-FI	AE	0.804	0.879	0.784	0.892	0.869	0.921	0.778	0.123		
				FI	0.196	0.823	0.295	0.359	0.529	0.737				
			B-FID	AE	0.751	0.888	0.789	0.888	0.872	0.930	0.556	0.141		
				FI	0.187	0.822	0.307	0.373	0.539	0.751				
2000, Hombres	81	B-FI	AE	0.795	0.882	0.791	0.897	0.873	0.922	0.778	0.135			
			FI	0.205	0.835	0.282	0.338	0.585	0.809					
		B-FID	AE	0.717	0.892	0.777	0.871	0.876	0.925	0.556	0.129			
			FI	0.189	0.834	0.287	0.343	0.587	0.812					
2000, Mujeres	82	B-FI	AE	0.769	0.875	0.765	0.874	0.858	0.914	0.778	0.172			
			FI	0.231	0.817	0.340	0.417	0.593	0.791					
		B-FID	AE	0.721	0.884	0.776	0.878	0.856	0.919	0.556	0.176			
			FI	0.211	0.817	0.333	0.408	0.590	0.794					
FD	AE	0.068	0.806	0.004	0.005	0.274	0.652							
	FI	0.068	0.806	0.004	0.005	0.274	0.652							

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB
U.S. Department of Labor (2019a)	1980, Hombres	83	B-FI	AE	0.771	0.856	0.765	0.893	0.841	0.910	0.778	0.112
				FI	0.229	0.770	0.301	0.391	0.602	0.789		
			B-FID	AE	0.689	0.867	0.757	0.874	0.833	0.894	0.556	0.077
				FI	0.172	0.786	0.180	0.229	0.568	0.782		
	1980, Mujeres	84	B-FI	AE	0.772	0.860	0.769	0.895	0.838	0.907	0.778	0.119
				FI	0.228	0.782	0.294	0.375	0.602	0.791		
			B-FID	AE	0.696	0.870	0.761	0.875	0.835	0.900	0.556	0.095
				FI	0.137	0.803	0.100	0.125	0.469	0.735		
	1987, Hombres	85	B-FI	AE	0.772	0.891	0.778	0.873	0.879	0.930	0.778	0.148
				FI	0.228	0.817	0.374	0.457	0.614	0.802		
			B-FID	AE	0.703	0.901	0.779	0.864	0.870	0.923	0.556	0.159
				FI	0.088	0.843	0.051	0.060	0.386	0.724		
	1987, Mujeres	86	B-FI	AE	0.796	0.882	0.782	0.887	0.870	0.926	0.778	0.147
				FI	0.204	0.788	0.338	0.428	0.555	0.755		
			B-FID	AE	0.715	0.897	0.784	0.874	0.872	0.932	0.556	0.137
				FI	0.200	0.804	0.356	0.443	0.594	0.791		
U.S. Department of Labor (2019b)	1996, Hombres	87	B-FI	AE	0.776	0.869	0.777	0.894	0.850	0.913	0.778	0.130
				FI	0.224	0.793	0.303	0.381	0.620	0.810		
			B-FID	AE	0.708	0.879	0.777	0.884	0.843	0.898	0.556	0.071
				FI	0.172	0.798	0.222	0.278	0.580	0.795		
	1996, Mujeres	88	B-FI	AE	0.810	0.874	0.800	0.916	0.859	0.919	0.778	0.091
				FI	0.190	0.793	0.249	0.314	0.565	0.786		
			B-FID	AE	0.741	0.882	0.789	0.894	0.857	0.919	0.556	0.109
				FI	0.101	0.816	0.045	0.055	0.375	0.717		
	2000, Hombres	89	B-FI	AE	0.841	0.892	0.833	0.934	0.882	0.932	0.778	0.068
				FI	0.159	0.811	0.209	0.257	0.521	0.770		
			B-FID	AE	0.790	0.899	0.834	0.928	0.883	0.936	0.556	0.069
				FI	0.147	0.810	0.208	0.257	0.517	0.771		
				FD	0.063	0.844	0.019	0.023	0.265	0.635		

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB
U.S. Department of Labor (2019b)	2000, Mujeres	90	B-FI	AE	0.865	0.900	0.848	0.943	0.889	0.935	0.778	0.062
				FI	0.135	0.838	0.177	0.211	0.477	0.746		
			B-FID	AE	0.795	0.908	0.845	0.931	0.889	0.936	0.556	0.058
				FI	0.122	0.838	0.171	0.204	0.470	0.748		
	2004, Hombres	91	B-FI	AE	0.807	0.880	0.795	0.904	0.863	0.919	0.778	0.116
				FI	0.193	0.810	0.276	0.341	0.550	0.767		
			B-FID	AE	0.751	0.886	0.791	0.892	0.864	0.919	0.556	0.103
				FI	0.175	0.810	0.267	0.330	0.539	0.761		
	2004, Mujeres	92	B-FI	AE	0.829	0.894	0.822	0.920	0.877	0.926	0.778	0.096
				FI	0.171	0.839	0.235	0.280	0.535	0.768		
			B-FID	AE	0.762	0.903	0.823	0.912	0.879	0.935	0.556	0.119
				FI	0.167	0.837	0.259	0.310	0.551	0.783		
	2008, Hombres	93	B-FI	AE	0.819	0.894	0.815	0.911	0.878	0.927	0.778	0.111
				FI	0.181	0.837	0.261	0.312	0.552	0.775		
			B-FID	AE	0.763	0.903	0.820	0.908	0.879	0.922	0.556	0.056
				FI	0.110	0.848	0.116	0.137	0.456	0.750		
	2008, Mujeres	94	B-FI	AE	0.869	0.899	0.854	0.949	0.889	0.936	0.778	0.051
				FI	0.131	0.837	0.157	0.188	0.467	0.738		
			B-FID	AE	0.802	0.907	0.852	0.938	0.891	0.942	0.556	0.073
				FI	0.130	0.833	0.186	0.223	0.483	0.750		
	2012, Hombres	95	B-FI	AE	0.805	0.894	0.812	0.908	0.876	0.927	0.778	0.111
				FI	0.195	0.830	0.275	0.332	0.586	0.803		
			B-FID	AE	0.747	0.900	0.803	0.892	0.872	0.915	0.556	0.081
				FI	0.119	0.848	0.100	0.118	0.471	0.777		
2012, Mujeres	96	B-FI	AE	0.884	0.910	0.874	0.960	0.903	0.944	0.778	0.043	
			FI	0.116	0.850	0.129	0.151	0.475	0.818			
		B-FID	AE	0.826	0.917	0.870	0.949	0.903	0.943	0.556	0.034	
			FI	0.104	0.853	0.113	0.132	0.467	0.830			
				FD	0.071	0.853	0.053	0.062	0.324	0.658		

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB	
U.S. Department of Labor (2019b)	2016, Hombres	97	B-FI	AE	0.825	0.904	0.828	0.915	0.886	0.930	0.778	0.111	
				FI	0.175	0.854	0.254	0.297	0.562	0.790			
			B-FID	AE	0.762	0.912	0.823	0.902	0.888	0.932	0.556	0.091	
					FI	0.150	0.853	0.233	0.274	0.532	0.774		
					FD	0.088	0.849	0.065	0.077	0.373	0.695		
		2016, Mujeres	98	B-FI	AE	0.855	0.917	0.858	0.936	0.903	0.942	0.778	0.077
				FI	0.145	0.866	0.204	0.235	0.523	0.782			
			B-FID	AE	0.797	0.925	0.862	0.931	0.904	0.940	0.556	0.043	
				FI	0.109	0.866	0.152	0.175	0.460	0.748			
				FD	0.094	0.871	0.070	0.081	0.437	0.793			
Vaughn (2019)	Estudio 1	99	B-FI	AE	0.919	0.951	0.917	0.963	0.950	0.970	0.778	0.037	
				FI	0.081	0.909	0.130	0.143	0.405	0.779			
			B-FID	AE	0.857	0.958	0.909	0.949	0.957	0.987	0.556	0.051	
					FI	0.091	0.909	0.163	0.180	0.449	0.825		
					FD	0.052	0.936	0.016	0.017	0.300	0.881		
		Estudio 2	100	B-FI	AE	0.829	0.920	0.840	0.913	0.905	0.940	0.778	0.136
				FI	0.171	0.883	0.263	0.297	0.562	0.807			
			B-FID	AE	0.775	0.927	0.832	0.898	0.912	0.939	0.556	0.061	
				FI	0.060	0.891	0.059	0.067	0.300	0.652			
				FD	0.165	0.860	0.273	0.317	0.588	0.792			
Yin (2019)		101	B-FID	AE	0.675	0.922	0.789	0.855	0.886	0.901	0.556		
				FI	0.138	0.857	0.159	0.186	0.556	0.825			
				FD	0.187	0.884	0.275	0.311	0.664	0.866			
Zhang et al. (2019)		102	B-FI	AE	0.835	0.912	0.835	0.915	0.901	0.941	0.778	0.098	
				FI	0.165	0.850	0.266	0.313	0.539	0.776			
			B-FID	AE	0.744	0.922	0.822	0.891	0.903	0.939	0.556	0.085	
				FI	0.132	0.850	0.225	0.264	0.508	0.763			
				FD	0.125	0.887	0.116	0.131	0.534	0.865			
Bakker et al. (2020)		103	B-FI	AE	0.807	0.689	0.598	0.868	0.767	0.871	0.778	0.076	
				FI	0.193	0.421	0.315	0.748	0.395	0.629			
Bueno et al. (2020)		104	B-FI	AE	0.816	0.925	0.822	0.889	0.950	0.973	0.778		
				FI	0.184	0.795	0.417	0.525	0.629	0.902			
			B-FID	AE	0.679	0.930	0.742	0.798	0.909	0.926	0.556		
				FI	0.110	0.808	0.237	0.293	0.477	0.799			
				FD	0.211	0.941	0.332	0.353	0.662	0.816			

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB	
Donnellan & Rakhshani (2020)	Estudio 2, Likert-4	105	B-FI	AE	0.710	0.910	0.751	0.825	0.909	0.947	0.778	0.289	
				FI	0.290	0.888	0.463	0.521	0.712	0.880			
	Estudio 2, Likert-5	106	B-FID	B-FI	AE	0.793	0.921	0.816	0.886	0.912	0.945	0.778	0.161
					FI	0.207	0.894	0.329	0.368	0.623	0.835		
					AE	0.686	0.938	0.806	0.860	0.919	0.962	0.556	0.175
					FI	0.188	0.894	0.346	0.387	0.637	0.858		
					FD	0.126	0.899	0.076	0.085	0.599	0.908		
	Estudio 2, Likert-6	107	B-FI	B-FID	AE	0.811	0.924	0.833	0.902	0.922	0.954	0.778	0.131
					FI	0.189	0.887	0.298	0.336	0.608	0.849		
					AE	0.723	0.937	0.823	0.878	0.926	0.958	0.556	0.123
					FI	0.166	0.887	0.294	0.332	0.601	0.852		
	Estudio 2, Likert-7	108	B-FI	B-FID	AE	0.693	0.916	0.744	0.812	0.924	0.957	0.778	0.311
					FI	0.307	0.894	0.498	0.557	0.741	0.897		
					AE	0.611	0.934	0.735	0.787	0.929	0.964	0.556	0.303
					FI	0.264	0.894	0.491	0.549	0.735	0.899		
					FD	0.126	0.898	0.107	0.119	0.569	0.848		
	Estudio 2, Likert-5	109	B-FI	B-FID	AE	0.784	0.923	0.828	0.896	0.915	0.953	0.778	0.128
					FI	0.216	0.859	0.346	0.402	0.694	0.879		
					AE	0.734	0.932	0.834	0.895	0.922	0.963	0.556	0.083
					FI	0.182	0.849	0.311	0.366	0.657	0.873		
Estudio 2, Likert-11	110	B-FI	B-FID	AE	0.780	0.934	0.828	0.887	0.936	0.965	0.778	0.086	
				FI	0.220	0.881	0.369	0.419	0.715	0.905			
				AE	0.692	0.944	0.802	0.850	0.944	0.966	0.556	0.074	
				FI	0.184	0.881	0.329	0.374	0.686	0.897			
				FD	0.124	0.930	0.150	0.161	0.562	0.881			
Gnambs & Schroeders (2020)	111	B-FI	B-FID	AE	0.778	0.867	0.760	0.877	0.850	0.909	0.778	0.165	
				FI	0.222	0.803	0.329	0.409	0.555	0.755			
				AE	0.698	0.875	0.742	0.848	0.850	0.902	0.556	0.123	
				FI	0.187	0.803	0.302	0.376	0.529	0.732			
				FD	0.115	0.801	0.114	0.143	0.422	0.679			

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB
Golec de Zavala et al. (2020)	112	B-FI	AE	0.755	0.916	0.781	0.852	0.925	0.960	0.778	0.093	
			FI	0.245	0.816	0.482	0.591	0.677	0.856			
		B-FID	AE	0.677	0.922	0.753	0.817	0.920	0.949	0.556	0.108	
			FI	0.215	0.817	0.447	0.547	0.652	0.838			
			FD	0.108	0.929	0.118	0.127	0.454	0.764			
			AE	0.757	0.925	0.789	0.853	0.930	0.962	0.778	0.104	
	113	B-FI	FI	0.243	0.835	0.484	0.579	0.680	0.859			
			B-FID	AE	0.707	0.931	0.781	0.838	0.941	0.971	0.556	0.091
				FI	0.216	0.835	0.451	0.541	0.661	0.856		
		114	B-FI	FD	0.078	0.936	0.068	0.072	0.375	0.819		
				AE	0.703	0.903	0.735	0.814	0.902	0.947	0.778	0.165
				FI	0.297	0.817	0.554	0.678	0.734	0.888		
Hussey & Hughes (2020)	115	B-FID	AE	0.600	0.909	0.683	0.751	0.883	0.910	0.556	0.140	
			FI	0.250	0.817	0.490	0.600	0.691	0.853			
			FD	0.150	0.904	0.206	0.228	0.546	0.733			
		B-FI	AE	0.844	0.913	0.849	0.930	0.908	0.947	0.778	0.066	
			FI	0.156	0.858	0.218	0.254	0.547	0.791			
			B-FID	AE	0.784	0.919	0.838	0.911	0.910	0.947	0.556	0.057
	FI	0.142		0.858	0.210	0.245	0.539	0.787				
	FD	0.074		0.865	0.070	0.081	0.330	0.658				
	Jabłońska & Zajdel (2020)	116	B-FI	AE	0.662	0.889	0.697	0.784	0.873	0.930	0.778	0.388
				FI	0.338	0.833	0.578	0.694	0.761	0.884		
				B-FID	AE	0.622	0.902	0.717	0.795	0.863	0.915	0.556
		FI	0.269		0.831	0.490	0.589	0.713	0.857			
FD		0.109	0.870		0.081	0.093	0.457	0.755				
Nguyen (2020)		117	B-FI	AE	0.686	0.899	0.724	0.806	0.878	0.930	0.778	0.318
	FI			0.314	0.852	0.526	0.617	0.717	0.862			
	B-FID			AE	0.641	0.905	0.713	0.788	0.893	0.942	0.556	0.305
		FI	0.296	0.852	0.524	0.616	0.716	0.867				
		FD	0.063	0.861	0.052	0.060	0.286	0.643				
	Wehner et al. (2020)	118	B-FI	AE	0.856	0.928	0.865	0.932	0.928	0.956	0.778	0.073
FI				0.144	0.902	0.206	0.228	0.545	0.815			
B-FID				AE	0.782	0.937	0.851	0.909	0.931	0.960	0.556	0.072
		FI	0.136	0.902	0.214	0.238	0.551	0.820				
		FD	0.082	0.878	0.075	0.086	0.392	0.718				

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB	
Deutsches Zentrum für Altersfragen (2021)	2008, Hombres	119	B-FI	AE	0.777	0.837	0.733	0.876	0.807	0.883	0.778	0.166	
				FI	0.223	0.760	0.316	0.416	0.507	0.703			
			B-FID	AE	0.699	0.848	0.732	0.864	0.811	0.894	0.556	0.184	
					FI	0.209	0.760	0.333	0.438	0.519	0.717		
					FD	0.092	0.764	0.020	0.026	0.305	0.641		
	2008, Mujeres	120	B-FI	AE	0.776	0.843	0.738	0.876	0.814	0.885	0.778	0.170	
				FI	0.224	0.778	0.314	0.404	0.520	0.714			
			B-FID	AE	0.691	0.855	0.738	0.863	0.819	0.904	0.556	0.203	
					FI	0.217	0.778	0.345	0.444	0.547	0.744		
					FD	0.092	0.766	0.007	0.009	0.316	0.687		
	2011, Hombres	121	B-FI	AE	0.737	0.837	0.711	0.849	0.799	0.876	0.778	0.218	
				FI	0.263	0.777	0.369	0.475	0.559	0.735			
			B-FID	AE	0.660	0.848	0.706	0.832	0.800	0.881	0.556	0.224	
					FI	0.241	0.777	0.377	0.486	0.563	0.740		
					FD	0.100	0.755	0.040	0.053	0.348	0.662		
	2011, Mujeres	122	B-FI	AE	0.739	0.841	0.715	0.851	0.804	0.878	0.778	0.214	
				FI	0.261	0.789	0.361	0.457	0.561	0.735			
			B-FID	AE	0.655	0.853	0.712	0.835	0.806	0.885	0.556	0.227	
					FI	0.240	0.789	0.375	0.475	0.571	0.747		
					FD	0.104	0.753	0.034	0.045	0.378	0.700		
	2014, Hombres	123	B-FI	AE	0.753	0.851	0.734	0.862	0.814	0.881	0.778	0.201	
				FI	0.247	0.806	0.341	0.423	0.559	0.734			
			B-FID	AE	0.672	0.863	0.733	0.850	0.817	0.898	0.556	0.236	
					FI	0.239	0.805	0.373	0.463	0.584	0.762		
				FD	0.089	0.761	0.003	0.004	0.335	0.703			
2014, Mujeres	124	B-FI	AE	0.759	0.850	0.735	0.864	0.815	0.881	0.778	0.195		
			FI	0.241	0.807	0.332	0.411	0.548	0.724				
		B-FID	AE	0.697	0.859	0.731	0.851	0.818	0.890	0.556	0.210		
				FI	0.233	0.807	0.350	0.433	0.563	0.741			
				FD	0.069	0.743	0.022	0.030	0.260	0.586			
2017, Hombres	125	B-FI	AE	0.763	0.856	0.742	0.867	0.820	0.885	0.778	0.189		
			FI	0.237	0.810	0.331	0.408	0.548	0.726				
		B-FID	AE	0.785	0.860	0.768	0.894	0.850	0.917	0.556	0.101		
				FI	0.035	0.820	0.004	0.005	0.133	0.461			
				FD	0.180	0.729	0.312	0.428	0.479	0.697			
2017, Mujeres	126	B-FI	AE	0.764	0.853	0.741	0.869	0.820	0.887	0.778	0.186		
			FI	0.236	0.801	0.328	0.410	0.547	0.731				
		B-FID	AE	0.685	0.864	0.738	0.855	0.825	0.905	0.556	0.224		
				FI	0.233	0.801	0.365	0.456	0.578	0.764			
				FD	0.082	0.768	0.005	0.006	0.304	0.686			

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB	
Dueber et al. (2021)	Grupo 2	127	B-FI	AE	0.842	0.950	0.868	0.914	0.940	0.962	0.778	0.108	
				FI	0.158	0.923	0.280	0.303	0.579	0.835			
		Grupo 3	128	B-FID	AE	0.809	0.954	0.868	0.910	0.940	0.966	0.556	0.102
					FI	0.150	0.923	0.275	0.298	0.574	0.839		
	128		B-FI	AE	0.862	0.926	0.857	0.925	0.929	0.948	0.778	0.084	
				FI	0.138	0.917	0.215	0.235	0.495	0.755			
	García-Batista et al. (2021)	129	B-FID	AE	0.847	0.928	0.852	0.919	0.927	0.942	0.556	0.052	
				FI	0.080	0.853	0.140	0.164	0.354	0.636			
		129	B-FI	AE	0.073	0.893	0.123	0.138	0.321	0.558			
				FI	0.759	0.861	0.746	0.866	0.843	0.907	0.778	0.185	
	Lifshin et al. (2021)	130	B-FID	AE	0.241	0.786	0.362	0.461	0.583	0.775			
				FI	0.669	0.875	0.741	0.847	0.854	0.929	0.556	0.211	
130		B-FI	AE	0.228	0.785	0.392	0.499	0.608	0.802				
			FI	0.103	0.823	0.026	0.031	0.392	0.785				
Lo Coco et al. (2021)	131	B-FID	AE	0.849	0.882	0.833	0.944	0.872	0.926	0.778	0.062		
			FI	0.151	0.822	0.164	0.199	0.492	0.755				
	131	B-FI	AE	0.775	0.891	0.821	0.921	0.874	0.926	0.556	0.047		
			FI	0.130	0.823	0.147	0.178	0.473	0.745				
Aging and Autism Study (2021)	132	B-FID	AE	0.095	0.806	0.089	0.111	0.377	0.659				
			FI	0.678	0.886	0.772	0.871	0.861	0.932	0.556			
	132	B-FI	AE	0.207	0.786	0.356	0.453	0.644	0.851				
			FI	0.115	0.848	0.022	0.026	0.432	0.792				
Medeiro et al. (2021)	133	B-FID	AE	0.863	0.918	0.858	0.936	0.931	0.962	0.778	0.048		
			FI	0.137	0.847	0.212	0.251	0.519	0.793				
	133	B-FI	AE	0.756	0.931	0.842	0.904	0.937	0.968	0.556	0.060		
			FI	0.126	0.846	0.226	0.267	0.524	0.798				
Polner et al. (2021)	134	B-FID	AE	0.118	0.915	0.089	0.097	0.525	0.884				
			FI	0.749	0.889	0.768	0.863	0.863	0.915	0.778	0.175		
	134	B-FI	AE	0.251	0.857	0.364	0.424	0.652	0.823				
			FI	0.785	0.917	0.812	0.886	0.906	0.945	0.778	0.138		
Polner et al. (2021)	134	B-FID	AE	0.215	0.858	0.354	0.412	0.640	0.849				
			FI	0.717	0.927	0.808	0.871	0.913	0.956	0.556	0.142		
	134	B-FI	AE	0.200	0.858	0.363	0.423	0.646	0.860				
			FI	0.082	0.899	0.038	0.042	0.423	0.866				

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ω_h	ω/ω_h	H	DF	PUC	ARPB
Ponce et al. (2021)		135	B-FI	AE	0.790	0.911	0.811	0.891	0.902	0.943	0.778	0.121
				FI	0.210	0.852	0.334	0.392	0.632	0.839		
			B-FID	AE	0.700	0.923	0.796	0.862	0.911	0.953	0.556	0.122
				FI	0.194	0.851	0.347	0.408	0.639	0.849		
Rufino & Holden (2021)		136	B-FI	AE	0.790	0.950	0.837	0.881	0.938	0.964	0.778	0.176
				FI	0.210	0.918	0.381	0.415	0.676	0.884		
			B-FID	AE	0.732	0.959	0.835	0.870	0.945	0.972	0.556	0.121
				FI	0.167	0.918	0.328	0.358	0.627	0.872		
Scheve & Stasavage (2021)		137	B-FI	AE	0.805	0.916	0.821	0.896	0.903	0.940	0.778	0.142
				FI	0.195	0.883	0.301	0.341	0.600	0.815		
			B-FID	AE	0.733	0.923	0.797	0.864	0.901	0.928	0.556	0.090
				FI	0.151	0.884	0.241	0.273	0.544	0.772		
Skylark & Callan (2021)		138	B-FI	AE	0.894	0.949	0.904	0.952	0.943	0.964	0.778	0.059
				FI	0.106	0.925	0.160	0.173	0.485	0.817		
			B-FID	AE	0.836	0.955	0.892	0.934	0.946	0.963	0.556	0.035
				FI	0.072	0.929	0.093	0.100	0.395	0.785		
Yovel et al. (2021)	Estudio 1	139	B-FI	AE	0.828	0.930	0.849	0.913	0.926	0.954	0.778	0.114
				FI	0.172	0.906	0.262	0.290	0.597	0.854		
			B-FID	AE	0.742	0.940	0.831	0.884	0.925	0.948	0.556	0.090
				FI	0.143	0.908	0.232	0.256	0.572	0.837		
	Estudio 2	140	B-FI	AE	0.830	0.933	0.853	0.915	0.922	0.953	0.778	0.108
				FI	0.170	0.898	0.269	0.300	0.598	0.845		
			B-FID	AE	0.778	0.939	0.842	0.897	0.924	0.949	0.556	0.076
				FI	0.145	0.897	0.240	0.267	0.565	0.825		
			FD	0.077	0.891	0.091	0.102	0.380	0.687			

Autores	Subgrupo	Mod	B-Mod	Factor	ECV	ω	ωh	$\omega/\omega h$	H	DF	PUC	ARPB
Zitelny et al. (2021)	Estudios 1-2	141	B-FI	AE	0.824	0.917	0.838	0.913	0.902	0.941	0.778	0.103
				FI	0.176	0.868	0.269	0.310	0.583	0.808		
			B-FID	AE	0.767	0.924	0.826	0.895	0.903	0.935	0.556	0.066
				FI	0.141	0.868	0.222	0.255	0.534	0.777		
	Estudio 6	142	B-FI	AE	0.880	0.914	0.864	0.946	0.909	0.947	0.778	0.053
				FI	0.120	0.847	0.177	0.208	0.451	0.738		
			B-FID	AE	0.847	0.918	0.862	0.940	0.911	0.951	0.556	0.059
				FI	0.119	0.846	0.186	0.220	0.460	0.746		
	Estudios 7-9	143	B-FI	AE	0.830	0.903	0.825	0.914	0.891	0.933	0.778	0.111
				FI	0.170	0.857	0.250	0.292	0.537	0.773		
			B-FID	AE	0.768	0.909	0.810	0.891	0.891	0.926	0.556	0.072
				FI	0.137	0.858	0.209	0.243	0.493	0.740		
	Estudios 11-13	144	B-FI	AE	0.830	0.906	0.831	0.918	0.895	0.937	0.778	0.095
				FI	0.170	0.851	0.249	0.292	0.550	0.791		
			B-FID	AE	0.781	0.912	0.827	0.907	0.898	0.943	0.556	0.104
FI				0.166	0.851	0.262	0.308	0.560	0.801			
			FD	0.054	0.852	0.023	0.027	0.244	0.620			

Referencias

- Alessandri, G., Vecchione, M., Eisenberg, N., & Laguna, M. (2015). On the factor structure of the Rosenberg (1965) General Self-Esteem Scale. *Psychological Assessment, 27*(2), 621–536. <https://doi.org/10.1037/pas0000073>
- Browne, M.W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258. <https://doi.org/10.1177/0049124192021002005>
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*(3), 440–464. https://doi.org/10.1207/s15328007sem1303_6
- Gnambs, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg self-esteem scale: A cross-cultural meta-analysis. *Zeitschrift für Psychologie, 226*(1), 14–29. <https://doi.org/10.1027/2151-2604/a000317>
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Marsh, H.W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Marsh, H.W., Scalas, L.F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22*(2), 366–381. <https://doi.org/10.1037/a0019225>
- Muthén, L.K., & Muthén, B.O. (1998-2014). *Mplus user's guide*. (7th ed.). Muthén & Muthén.
- Reise, S.P., Kim, D.S., Mansolf, M., & Widaman, K.F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of Iteratively Reweighted Least Squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Research, 51*(6), 818–838. <https://doi.org/10.1080/00273171.2016.1243461>

3.2. MANUSCRITO 2:

Efectos de redacción en evaluación: Cuando el bosque no deja ver los árboles

Traducción del manuscrito publicado en Multivariate Behavioral Research

<https://doi.org/10.1080/00273171.2021.1925075>

Fernando P. Ponce¹, David Torres Iribarra¹, Alvaro Vergés¹, & Víctor B. Arias²

¹ Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile.

² Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Facultad de Psicología, Universidad de Salamanca

Resumen

Este artículo examina el efecto de redacción cuando se incluyen ítems redactados positiva y negativamente en evaluación psicológica. Los efectos de redacción han sido analizados en la literatura utilizando enfoques estadísticos basados en el supuesto de homogeneidad poblacional (*i.e.*, CFA, SEM), adoptando comúnmente modelos bifactor para separar la varianza del rasgo de la varianza del efecto de la redacción. Este artículo presenta un enfoque alternativo, al modelar, explícitamente, la heterogeneidad de la población a través de un modelo de perfiles latentes, basado en la idea de que un subconjunto de individuos son los que exhiben efectos de redacción. Este tipo de modelo de mezclas permite simultáneamente clasificar a los participantes, caracterizar sustancialmente las diferencias en sus perfiles de respuesta e informar los resultados de los encuestados de manera comparable. En tres estudios, en los que utilizamos los datos de la escala de autoestima de Rosenberg del panel LISS (N=6762), identificamos un subgrupo de participantes que responden de manera diferente según la redacción de los ítems y examinamos el impacto de sus respuestas, en términos de ajuste global e individual, en la estimación del modelo de medición para la RSES, bajo modelamiento de un-factor y bifactor.

Los resultados de estos análisis apoyan la interpretación de los efectos de redacción en términos de un patrón diferencial de respuesta a los elementos redactados positiva y negativamente, introduciendo una valiosa herramienta para examinar la interpretación artefactual o sustantiva de tales efectos de redacción.

Palabras clave: Efectos de redacción, análisis de perfiles latentes, escala de autoestima de Rosenberg, análisis factorial confirmatorio, mínimos cuadrados iterativamente ponderados.

El efecto de redacción es uno de los sesgos de respuesta más comunes que se observan en las escalas compuestas por ítems positiva y negativa redactados. La investigación basada en la teoría del estilo de respuesta (Jackson & Messick, 1958), ha investigado los efectos de la redacción de una variedad de constructos psicológicos, utilizando enfoques estadísticos que asumen la homogeneidad de la población. Sin embargo, reciente evidencia (Kam & Fan, 2018; Reise et al., 2016) ha cuestionado esta suposición y su corpus de evidencia. El presente estudio tiene como objetivo modelar directamente la heterogeneidad de la población mediante análisis de perfiles latentes en orden de identificar y caracterizar el subgrupo específico de participantes quienes exhiben efectos de redacción. Para ilustrar este enfoque, examinamos la heterogeneidad en la estimación de la dimensionalidad de la escala de autoestima de Rosenberg (RSES), una de las escalas más utilizadas en los estudios de los efectos de redacción. Este enfoque permitirá mejorar los procedimientos empíricos para examinar la naturaleza de este sesgo de respuesta en la investigación psicológica.

Efectos de redacción

Muchos de los constructos investigados en psicología se representan como dimensiones bipolares (*e.g.*, extroversión-introversión). En la práctica, esta bipolaridad se traslada a los test a través de ítems redactados positiva y negativamente en escalas balanceadas. Esta estrategia se basa en el supuesto de que los ítems positivos y negativos representan muestras psicométricamente equivalentes del dominio del constructo (Quilty et al., 2006). Además, los ítems redactados negativamente pueden reducir los efectos de patrones de respuestas sesgados (*e.g.*, aquiescencia o sesgo de acuerdo; Cronbach, 1950), dado que los ítems negativos actúan como "obstáculos" cognitivos que requieren que las personas se comprometan en un procesamiento cognitivo controlado (Podsakoff et al., 2003).

Sin embargo, diferentes líneas de investigación han demostrado que los participantes brindan respuestas menos consistentes a ítems redactados negativamente (Holden & Fekken, 1990; Schriesheim et al., 1991) y tienden a responder a ítems redactados de manera positiva y negativa diferencialmente (Kamoen et al., 2011, 2013;

Quilty et al., 2006). Esta diferencia en las respuestas introduciría un sesgo sistemático ligado a los procedimientos de medición (*i.e.*, efectos de redacción) en el análisis, produciendo alteraciones sustanciales en la calidad de los datos que pueden derivar en: (a) relaciones espurias entre variables, que de otro modo no estarían correlacionadas (Huang et al., 2015); (b) una inflación/deflación de la consistencia interna (Wood et al., 2017); (c) un desajuste en modelos unidimensionales y la aparición de factores de método en escalas balanceadas (Greenberger et al., 2003; Kam & Meyer, 2015); y (d) un sesgo sistemático que socava la efectividad de las manipulaciones experimentales (Maniaci & Rogge, 2014), por mencionar algunas consecuencias. En este sentido, la interpretación y manejo de tales elementos sistemáticos implican desafíos tanto conceptuales como analíticos para la calidad de la evaluación con medidas de auto-reporte.

Efectos de redacción como un estilo de respuesta

Una de las cuestiones más relevantes sobre los efectos de redacción es determinar si su naturaleza es puramente artificial (*i.e.*, sesgo de medición atribuible al instrumento) o sustantiva (*i.e.*, un rasgo que subyace a las respuestas a los ítems positivos y/o negativos, pero que no es especificado en el modelo teórico). Marsh (1996) consideró los efectos de redacción como “efectos de método sustancialmente irrelevantes” (p. 815), que consisten en un sesgo sistemático que contaminaría la medición, requiriendo así especificaciones estadísticas para controlar y minimizar sus efectos. Es decir, la varianza sistemática capturada por ítems positivos y negativos representa varianza “efímera” (Marsh et al., 2010). Por otro lado, Tomás y Oliver (1999) sugirieron que los efectos de redacción representan una característica estable como un estilo de respuesta en sí mismo de relevancia sustantiva (*i.e.*, un rasgo individual), más que un mero artefacto metodológico. Según este punto de vista, un estilo de respuesta asume que es una "consistencia conductual que opera a través de medidas de varios rasgos de contenido conceptualmente distintos", generalmente, reflejada en situaciones de evaluación por consistencias en respuesta a las características de los ítems, que pueden ser "identificadas por la existencia de una variable latente" (Bentler et al., 1971; p.188; véase también Jackson & Messick, 1958). En consecuencia, Tomás y Oliver (1999) sugirieron que un estilo de respuesta

sustantivo asociado con efectos de redacción implica: (a) una distinción entre rasgos y efectos de redacción (*i.e.*, una variable latente del factor de redacción), (b) estabilidad temporal en los factores de redacción a partir de medidas obtenidas en diferentes olas de medición, (c) evidencia de un factor de redacción más prominente para los individuos que son más altos en rasgos de personalidad específicos, y (d) la presencia de factores de redacción en diferentes rasgos, con asociaciones significativas entre ellos.

Basados en esta conceptualización, Christine DiStefano (DiStefano & Motl, 2006; 2009a,b; Horan et al., 2003) ha utilizado el marco analítico de rasgos correlacionados/métodos correlacionados (CTCM), en la RSES para separar y estimar componentes sustantivos y de redacción del instrumento (Marsh, 1989; Marsh & Grayson, 1995). El CTCM permite el estudio de los efectos de la redacción como una entidad distinta que puede tener relaciones empíricas sustantivas con otras variables (Horan et al., 2003). Usando este marco analítico, DiStefano y sus colegas (DiStefano & Motl, 2006; 2009a,b; Horan et al., 2003; Motl & DiStefano, 2002) reportaron cuatro fuentes de evidencia que apoyan la idea de los efectos de redacción como un estilo de respuesta, en línea con las implicaciones descritas por Tomás y Oliver (1999).

La primera fuente de evidencia corresponde a la mejora en el ajuste del modelo, cuando se compara un modelo de varianza de un solo rasgo (*e.g.*, un factor; modelo izquierdo en la Figura 1a) y un modelo que separa la varianza del rasgo y de los efectos de redacción (*e.g.*, bifactor; modelo derecho en la Figura 1a). La segunda fuente de evidencia involucra la estabilidad temporal de los efectos de la redacción entre los ítems redactados positiva y/o negativamente ($r_{wpt1, wpt2}$ y/o $r_{wnt1, wnt2}$ en la Figura 1b), después de controlar la varianza única de los mismos ítems (Marsh & Grayson, 1994), en orden de examinar la estacionalidad (*i.e.*, la misma estructura a través de distintas olas) y la estabilidad (*i.e.*, el mismo orden relativo de individuos a través de las olas) de los factores de redacción. La tercera fuente de evidencia tiene que ver con la asociación del factor de redacción con variables latentes que representan diferencias individuales en variables sustantivas ($r_{Y, W_{PX}}$ y/o $r_{Y, W_{NX}}$ en la Figura 1c), tales como habilidades cognitivas o rasgos de personalidad. La última fuente de evidencia consiste en la asociación entre los factores

de redacción extraídos de diferentes constructos ($r_{wnx, wny}$ y/o $r_{wpx, wpy}$ en la Figura 1d), más allá de la asociación entre sus rasgos sustantivos ($r_{X,Y}$ en la Figura 1d).

[Insertar Figura 1 aproximadamente aquí]

El número de hallazgos que apoyan a cada una de estas fuentes de evidencia es variable. La primera fuente de evidencia, en relación con la separación de rasgos y efectos de redacción, ha sido reportada para medidas de autoestima (Alessandri et al., 2015; DiStefano & Motl, 2006; DiStefano & Motl, 2009a,b; Gnambs et al., 2018; Horan et al., 2003), ansiedad físico social (Motl & Conroy, 2000; Motl et al., 2000), auto-motivación (Motl et al., 2003), miedo al fracaso (Conroy, 2001), y auto-evaluaciones centrales (Arias & Arias, 2017; Gu et al., 2015), por mencionar algunos atributos. Del mismo modo, diversos estudios han reportado evidencia del factor de redacción y su relación con criterios externos, específicamente sobre diferencias en la tendencia a prestar atención a pensamientos y sentimientos internos y externos (DiStefano & Motl, 2006; 2009b; Michaelides et al., 2016), tendencias motivacionales y afectivas (Alessandri et al., 2015; Arias & Arias, 2017; DiStefano & Motl, 2009b; Lindwall et al., 2012; Quilty et al., 2006; Tomás et al., 2013), rasgos de personalidad (Arias et al., 2018; Donnellan et al., 2016; Michaelides et al., 2016) y habilidades verbales (Dunbar et al., 2000; Gnambs & Schroeders, 2020; Marsh, 1996; Weems et al., 2006).

En contraste, para los efectos de estabilidad temporal, solo unos pocos conjuntos de estudios han reportado apoyo empírico para la estacionalidad y estabilidad (Motl & DiStefano, 2002; Gana et al., 2013; Marsh et al., 2010). Finalmente, con respecto a la relación entre los factores de redacción extraídos de diferentes rasgos, solo un estudio reportó evidencia que apoya esta predicción (DiStefano & Motl, 2006).

Estilo de respuesta y heterogeneidad

La evidencia empírica descrita anteriormente, sobre los efectos de redacción como un potencial estilo de respuesta, se ha obtenido utilizando el modelo bifactor para separar la varianza sustantiva de la que se atribuye a la redacción, estimándose un marco estadístico (*i.e.*, análisis factorial confirmatorio y modelado de ecuaciones estructurales), que asume homogeneidad de la población (todos los participantes pertenecen a una

población homogénea). Este enfoque analítico implica explicar los efectos de redacción a partir de la estimación de un conjunto común de parámetros para toda la población. Sin embargo, ambos supuestos, la representación bifactor y la homogeneidad poblacional, son controversiales (Kam & Fan, 2018; Reise et al., 2016).

Reise et al. (2016), utilizando mínimos cuadrados iterativamente ponderados (IRLS; ver Yuan & Bentler, 2000), informaron que en circunstancias en que el modelo bifactor muestra un mejor ajuste que una solución de un factor, este modelo aumenta la predictibilidad de los patrones de respuesta del individuo sólo en el 3% de la muestra en comparación con la predictibilidad observada con el modelo de un factor. Además, pese al ajuste global mejorado, el 11% de los casos no se modelaba adecuadamente incluso con el enfoque bifactor. Estos hallazgos apuntan a la posibilidad de que la selección de modelos puede ser dirigido por una pequeña minoría de casos que tienen patrones de respuesta anormales, llevando a Reise et al. a concluir que el mejor ajuste del enfoque bifactor se debe a la "capacidad del modelo para acomodar mejor patrones de respuesta inverosímiles y, posiblemente, inválidos y no necesariamente porque explica mejor los efectos de la dirección de la redacción" (Reise et al., 2016, p.818). Esta capacidad podría deberse a la alta propensión del modelo bifactor a ajustar a cualquier dato posible, dada la forma funcional compleja derivada de la forma en que sus parámetros y variables aleatorias se combinan y especifican en las ecuaciones del modelo (Bonifay & Cai, 2017).

Además, más allá de la cuestión de si el modelo bifactor es la mejor opción para dar cuenta de los efectos de redacción, un grupo específico de participantes puede exhibir dificultades para lidiar con ítems redactados negativamente. De ser así, toda la evidencia reportada basada en el supuesto de homogeneidad de la población es "potencialmente engañosa, si no fundamentalmente defectuosa" (Kam & Fan, 2018; p.3). En vista de lo anterior, la inspección de la naturaleza de los efectos de la redacción requiere enfoques de modelado que asuman heterogeneidad poblacional para identificar a aquellos encuestados que exhiben respuestas diferenciales de acuerdo a la dirección de los ítems y examinar las implicancias descritas por Tomás y Oliver (1999). De hecho, existe cierta evidencia de que las diferencias individuales pueden influir en la relación observada entre ítems

redactados de manera similar. Se ha encontrado que las respuestas inconsistentes se relacionan con constructos cognitivos y de personalidad (ver la tercera fuente de evidencia descrita anteriormente). Por lo tanto, la relación entre ítems redactados de manera similar y la dimensionalidad observada de un constructo dado puede diferir entre los individuos de una población.

Basados en esta posibilidad, utilizando modelamiento de mixtura de factores (FMM; ver Lubke & Muthén, 2005), Kam y Fan (2020) y Kam (2020) identificaron heterogeneidad en las respuestas a ítems positivos y negativos de satisfacción laboral, optimismo y autoestima. Específicamente, encontraron dos clases con un grupo de participantes que mostraron respuestas inconsistentes a preguntas redactadas negativamente, operacionalizado como la diferencia observada entre los valores de respuesta máximo y mínimo dentro del subconjunto de ítems negativos. Finalmente, después de excluir del análisis a este grupo específico de encuestados, se encontró una solución esencialmente unidimensional para estos rasgos. Sin embargo, cabe mencionar que, esta evidencia que explica el impacto del subgrupo identificado también ha sido reportada en estudios que, utilizando las mismas escalas, no excluyeron ninguna clase de participantes (Alessandri et al., 2011; 2015, por mencionar algunos).

Dado lo anterior, una forma más apropiada de describir el impacto de este grupo eliminado habría sido examinar la mejora en la estructura esencialmente unidimensional, mediante la estimación de indicadores derivados del modelo bifactor, como la variación en: (a) la proporción de varianza capturada por el factor de rasgo del modelo bifactor (calculado a partir de la varianza común explicada [ECV]; Brunner et al., 2012), (b) la fiabilidad compuesta multidimensional ponderada por unidad para el factor general (calculada a partir del coeficiente omega; McDonald, 1999) y, (c) la proporción de la varianza en la puntuación bruta compuesta ponderada por unidad atribuible al factor general, o la proporción de varianza confiable en las puntuaciones compuestas debido al factor general (ambos calculados a partir del coeficiente omega jerárquico; Zinbarg et al., 2005), por mencionar algunas posibilidades.

Estos estudios identifican una proporción de participantes cuyos patrones de respuesta desafían los modelos definidos teóricamente para escalas con ítems redactados de manera positiva y negativa, lo que genera preguntas sobre el supuesto de homogeneidad de la población en las respuestas a las medidas de auto-reporte. No obstante, cabe señalar que Reise et al. (2016) y Kam y Fan (2018; Kam, 2020) han desarrollado procedimientos para caracterizar la heterogeneidad de la población que se basan en la especificación de modelos factoriales y el uso de un proceso de varios pasos para identificar subconjuntos de casos que no se ajustan a la estructura factorial especificada. Reise et al. (IRLS; 2016) utilizó un procedimiento de dos pasos, basado en indicadores de métodos robustos, para describir la idoneidad de un patrón de respuesta dado un modelo confirmatorio (*i.e.*, un factor, dos factores y bifactor); Kam y Fan (FFM; 2018; Kam, 2020) utilizaron un procedimiento de comparaciones secuenciales entre modelos mixtos para identificar subgrupos latentes derivados de un modelo específico de dos factores correlacionados. Bajo ambos enfoques, IRLS y FMM, una vez que se detectan los casos que contribuyen a la heterogeneidad de la población, no existe una forma clara de caracterizarlos de manera sustantiva o reportar los resultados para este subconjunto de la población de una manera comparable con los resultados de los encuestados que están mejor alineados con un modelo homogéneo⁶.

Heterogeneidad latente

Como señalan Kam y Fan (2018), la heterogeneidad de la población puede estar presente pero ser indetectable en el contexto de los estudios de efectos de redacción porque la información sobre las subpoblaciones es inobservable o latente. En otras palabras, no es posible identificar directamente qué participantes, en una población, son responsables de producir patrones que son inconsistentes con un modelo teórico

⁶ El enfoque IRLS utilizado por Reise et al. (2016) descansa en la estimación de un modelo específico (e.g., bifactor) para posteriormente identificar en una segunda etapa los casos que desajustan, pero una vez los casos identificados, no ofrece una alternativa obvia inmediata para reportar los resultados de este subgrupo de la población. El enfoque FMM adoptado por Kam y Fan (2020) genera estimaciones para cada clase, pero estas no son comparables dada la ausencia de parámetros comunes que permitan equiparar las escalas en las que se expresan los resultados de las diferentes clases.

esperado (*e.g.*, un factor), y emplean modelos de variables latentes para estimar patrones de respuesta erráticos a un modelo específico.

Para abordar este problema, es necesario utilizar una técnica estadística capaz de identificar distintos subgrupos latentes en los datos de una muestra, basados en patrones de respuesta (Marsh et al., 2009). El análisis de perfiles latentes (LPA) es una técnica sencilla de modelado de variables latentes que asume que los encuestados provienen de grupos no observados en la población, con una proporción diferente cada uno (Lazarsfeld & Henry, 1968; Marsh et al., 2009). Cada grupo de encuestados, se caracteriza por valores similares en un conjunto de indicadores observados (Pastor et al., 2007), modelando explícitamente la heterogeneidad poblacional. En el contexto de los efectos de redacción, LPA podría ofrecer la posibilidad de identificar perfiles de personas cuyas respuestas difieren en un conjunto específico de ítems (*e.g.*, ítems redactados positiva versus negativamente; Kamoen et al., 2011; Kamoen et al., 2013), sin asumir ningún modelo factorial específico. En este sentido, LPA es una forma más apropiada de modelar variables latentes asociadas con tendencias de respuesta (como lo sugieren Bentler et al., 1971; Mislevy & Verhelst, 1990), ya que ayuda a diferenciar grupos de individuos con un patrón de respuesta sensible a los efectos de la redacción, lidiando explícitamente con la heterogeneidad de una población y produciendo una taxonomía interpretable sustantivamente de los encuestados.

Hasta donde sabemos, Bandalos et al. (citado en Reise et al., 2016), en un artículo no publicado, describieron una solución de tres clases latentes en dos muestras de datos de la RSES, con un subgrupo de participantes (17% y 12% en cada muestra, respectivamente) que respondieron de manera diferencial a los ítems redactados negativamente. Este estudio, no publicado, ilustra este enfoque estadístico para identificar a los participantes sesgados.

En este artículo, aplicamos el LPA clásico para identificar un grupo específico de individuos con patrones de respuestas cualitativamente distintos a los ítems positivos y negativos y para examinar más a fondo el impacto de este subgrupo en la estimación de

la dimensionalidad de la escala de autoestima de Rosenberg, una de las escalas más utilizadas para estudiar los efectos de la redacción.

El presente estudio

Este artículo presenta tres estudios que abordan cómo modelar la heterogeneidad de la población puede ayudar a comprender los efectos de redacción. Nuestros objetivos son (a) identificar y caracterizar un subgrupo de participantes con un patrón de respuesta asimétrico, utilizando el marco de variables latentes categóricas, (b) examinar el impacto de este subgrupo, en su conjunto, sobre la unidimensionalidad de la RSES, y (c) cuantificar el efecto que tiene cada participante en el ajuste (o la falta de éste) del modelo unidimensional de la RSES.

En el Estudio 1, identificamos, utilizando LPA, un subgrupo de participantes que exhiben un patrón de respuesta asimétrico sesgado en la redacción. Hipotetizamos que un factor de redacción negativo y/o positivo en una solución bifactor refleja la presencia de un grupo específico de participantes que exhiben un patrón diferencial de respuesta entre ítems positivos y negativos (un subgrupo sesgado). Siguiendo a Reise et al. (2016), esperábamos que entre el 3% y el 14% de los participantes proporcionaran respuestas que obliguen a especificar una solución bifactor como modelo de medición general.

En el Estudio 2, estimamos modelos de medición para la RSES luego de controlar la presencia de este subgrupo identificado. En particular, estimamos un modelo de un factor y uno bifactor, utilizando la muestra completa (MC) y comparamos esas soluciones con modelos estimados con una muestra en la que se eliminaron los participantes que exhibían un patrón diferencial de respuesta (*i.e.*, muestra completa menos participantes identificados con efecto de método; MC-M). Dado que estos individuos identificados revelan un patrón sesgado de respuestas, con una naturaleza difícil de modelar (como afirman Reise et al., 2016), planteamos la hipótesis de que el mejor ajuste que presentaría el modelo bifactor en la condición MC, se reduciría considerablemente o incluso desaparecería en la condición MC-M. Este resultado revelaría un refinamiento en la estructura unidimensional, consistente con la evidencia parcial reportada por Kam y Fan (2020).

Finalmente, en el Estudio 3, caracterizamos los patrones de respuesta de los individuos contenidos en los perfiles identificados, en términos de adecuación a los modelos de un factor y bifactor (descrito como "modelabilidad" en Reise et al., 2016), así como su contribución individual a la prueba de razón de verosimilitud global entre ambos modelos (Reise & Widaman, 1999). En relación con la modelabilidad, a partir de los hallazgos hipotetizados en el Estudio 2, esperamos que los patrones de respuesta contenidos en los perfiles que no corresponden al efecto de redacción pudieran ser modelados por el modelo de un factor, similar a los hallazgos reportados por Reise et al. (2016). Por el contrario, esperamos que los patrones de respuesta de los participantes del grupo de método fueran modelados en su mayoría por el modelo bifactor o sean lo suficientemente inconsistentes como para considerar que no pueden modelarse adecuadamente. Finalmente, con respecto a la contribución a la prueba de razón de verosimilitud general entre modelos, hipotetizamos que los participantes del grupo de método contribuyen, principalmente, a una mayor diferencia a favor del modelo bifactor. Por el contrario, los valores de diferencia de los otros perfiles no descienden mucho por debajo de cero, con una pequeña contribución a la prueba de diferencia general de chi-cuadrado.

Método

Datos y Participantes

Para estos estudios utilizamos los datos panel LISS de CentERdata, un panel representativo para estudios longitudinales de Internet en ciencias sociales, administrado por el CentERdata en Tilburg, Países Bajos. El panel LISS consta de 5000 hogares (N=6808 participantes), seleccionados aleatoriamente de los registros municipales de los Países Bajos. Para la recolección de datos, se invita a los miembros del hogar a completar cuestionarios en línea todos los meses, asignándoles acceso a Internet y/o a una computadora personal cuando sea necesario (Scherpenzeel, 2011).

Considerando el tamaño muestral, para facilitar los análisis descartamos 46 individuos (0,68%) con datos de respuesta perdidos, lo que dejó la muestra utilizada para

los análisis en 6762 participantes (51,54% mujeres; 5,27% no informó su sexo), cuyas edades fluctuaron entre 15 y 108 años (Me=45,81; DE=15,96; 5,79% no informo su edad).

Medidas

Escala de autoestima de Rosenberg (RSES). La RSES es una escala balanceada de 10 ítems que miden autoestima global (Rosenberg, 1965). Los ítems fueron evaluados en una escala tipo-Likert de 7 puntos (*i.e.*, 1 “totalmente en desacuerdo” a 7 “totalmente de acuerdo”). Los ítems redactados negativamente se invirtieron de modo que las puntuaciones más altas en la RSES corresponden a una mayor autoestima. Una examinación preliminar utilizando diferentes estimadores e inspeccionando sus respectivos índices de modificación revelaron que, al igual que Reise et al. (2016), un par de pares de ítems (*i.e.*, i1 e i2; i6 e i7) exhibieron residuos correlacionados⁷. Por lo tanto, en nuestro análisis posterior, incorporamos estas especificaciones.

Estudio 1: Identificación de respuestas asimétricas

Análisis de los datos

Para evaluar la consistencia de la estimación de perfiles (Collins & Lanza, 2010), dividimos la muestra total en dos muestras aleatorias (n=3381 cada una). El análisis de perfiles latentes se realizó con ambas muestras utilizando Mplus v.7.4 (Muthén & Muthén, 1998-2014).

Se utilizaron consideraciones teóricas y estadísticas (*i.e.*, índices de bondad de ajuste) para identificar el número de perfiles que se ajustan adecuadamente a los datos observados (Muthén & Muthén, 2000). Los modelos de perfiles latentes se compararon utilizando el criterio de información de Akaike (AIC; Akaike, 1974), el criterio de información bayesiano (BIC; Schwarz, 1978), el BIC ajustado por tamaño de muestra (adjBIC; Nylund et al., 2007), y la prueba de razón de verosimilitud Lo-Mendell-Rubin (LMR-LRT; Lo et al., 2001). Al comparar los modelos, los valores más pequeños en los valores AIC, BIC y adjBIC indican un mejor ajuste del modelo, y un valor LMR-LRT significativo indica una mejora significativa del ajuste del modelo con k respecto a uno

⁷ Los ítems 1 y 2 son los primeros ítems a los que se respondió, ambos están redactados positivamente, e incluyen la frase común “Siento que...”. Los ítems 6 y 7 están redactados positivamente y ambos refieren a una evaluación positiva general de uno mismo.

con $k-1$ perfiles. Adicionalmente, se consideró un criterio de parsimonia en la solución de perfiles en base a una prevalencia mínima del 5% (Osborne & Sibley, 2017).

Luego de definir la solución óptima, examinamos la precisión de clasificación para cada solución en cada muestra utilizando tanto los estadísticos de entropía como las probabilidades de clasificación de perfil latente promedio (ALPCP). Los valores de entropía varían de 0 a 1, y los valores más altos indican una mayor grado de separación de entre las clases (Celeux & Soromenho, 1996; Lubke & Muthén, 2007). Los valores de ALPCP varían de 0 a 1, y los valores más altos indican niveles más bajos de clasificaciones erróneas (Vermunt & Magidson, 2002; Neely-Barnes, 2010).

Resultados

Análisis descriptivos

La Tabla 1 presenta los descriptivos para las dos muestras utilizadas en este estudio. Como puede verse, no se observan diferencias estadísticamente significativas entre las muestras (las pruebas t de muestras independientes variaron entre $t [6760]=-1,008; p=0,314$ a $t [6760]=1,170; p=0,242$).

[Insertar Tabla 1 aproximadamente aquí]

Perfiles latentes de autoestima y prevalencia.

Los índices de ajuste para los modelos de dos a seis perfiles se presentan en la Tabla 2. Aunque el AIC, BIC y adjBIC no respaldaron un modelo específico, los resultados con el test de LMR-LRT sugirieron la solución de cinco perfiles en ambas muestras. Sin embargo, el quinto perfil incluyó menos del 5% de la muestra. De este modo, en base al estadístico de LMR-LRT y la parsimonia del número de grupos, seleccionamos la solución de cuatro perfiles como modelo final.

La Figura 2 describe las medias no estandarizadas de los ítems directos e inversos estimadas en cada muestra. Como se puede observar, ambas muestras presentan patrones similares, donde tres de las cuatro clases presentan un patrón ordenado que se pueden interpretar claramente como una clase de alta autoestima (AE_{ALTA} : 52,14% en la muestra 1 y 52,23% en la muestra 2; *vid.*, Tabla 3), una clase de autoestima moderada o media (AE_{MED} : 28,99% en la muestra 1 y 29,45% en la muestra 2) y una clase de baja autoestima

(AE_{BAJA} : 9,70% en la muestra 2 y 10,15% en la muestra 1). La cuarta y última clase, sin embargo, presenta un patrón de puntuación asimétrico que se acerca a la clase de alta autoestima para los ítems redactados positivamente y cercano a la clase de baja autoestima para los ítems redactados negativamente ($AE_{MÉTODO}$). La prevalencia de este perfil fue 8,58% en la muestra 2 y 8,73% en la muestra 1 (*vid.*, Tabla 3). Vale la pena señalar que el uso de LPA permite dar cuenta de la heterogeneidad de la población al tiempo que proporciona clases distintas y sustancialmente interpretables que podrían usarse para informar los resultados de todos los encuestados bajo este marco.

[Insertar Figura 2 aproximadamente aquí]

El patrón de respuesta observado en $AE_{MÉTODO}$ fue claramente diferente del patrón visto en los otros perfiles, los cuales, más allá de algunas diferencias entre ítems, muestran monotonicidad de clase, lo que los convierte en clases latentes bien ordenadas (Croon, 1990; 1991; 2002) y consistentes con la noción de una estructura unidimensional de autoestima. Además, el patrón asimétrico de respuestas observado en el perfil $AE_{MÉTODO}$ sugiere que (a) los participantes pueden responder diferencialmente a los ítems redactados negativamente, expresando una opinión menos favorable en comparación con los ítems positivos; y (b) tal como se propuso inicialmente, este patrón asimétrico, en un grupo específico de ítems, viola el comportamiento esperado de las clases consistente con una estructura unidimensional de la autoestima general, induciendo potencialmente la presencia de variables latentes adicionales en un modelo factorial para dar cuenta del patrón de respuesta diferencial en este subgrupo.

Estudio 2: El efecto de las respuestas sesgadas en los modelos de medición

La identificación de un subgrupo que muestra un patrón de respuesta asimétrica podría estar asociado con los hallazgos de dimensionalidad reportados con la RSES, que indican consistentemente un mejor ajuste de una estructura multidimensional (*i.e.*, dos factores, bifactor), en comparación con un rasgo de autoestima general.

Para explorar esta posibilidad, examinamos el impacto del patrón de respuesta identificado como un perfil $AE_{MÉTODO}$ en el ajuste del modelo estimado en una solución de un factor y una solución bifactor.

Análisis de datos

Para examinar el efecto de las respuestas asimétricas identificadas en el Estudio 1 en la dimensionalidad de la RSES, comparamos los resultados de ajuste de diferentes estructuras dimensionales mientras (a) incluimos los casos asociados con la clase que patrones de respuesta asimétricos, (b) removiendo esos casos, y (c) removiendo una muestra aleatoria de casos. Para lograr esto, utilizamos un modelo de un factor y dos enfoques bifactor alternativos a menudo considerados en la literatura: (a) un modelo bifactor con un factor general de autoestima y un factor específico de redacción negativa (Bifactor-FI; DiStefano & Motl, 2006; Marsh et al., 2010), y (b) un bifactor con un factor general de autoestima y dos factores específicos de efectos de redacción negativos y positivos (Bifactor-FID; Alessandri et al., 2015; Gnambs et al., 2018; Reise et al., 2016). Todos los modelos se estimaron utilizando Maximum Likelihood Robust (MLR) como método de estimación. La bondad de ajuste se evaluó en todos los casos utilizando el índice de ajuste comparativo (CFI), el índice de Tucker-Lewis (TLI), la raíz del error cuadrático medio de aproximación (RMSEA) y la raíz cuadrática media residual estandarizada (SRMR). Para los índices CFI y TLI, los valores estimados por encima de 0,90 y 0,95 indican grados de ajuste aceptable y bueno, respectivamente (Hu & Bentler, 1999; Marsh et al., 2004). Para el SRMR y RMSEA, los valores iguales o inferiores a 0,05 y 0,08 se consideran buenos y aceptables, respectivamente (Browne & Cudeck, 1992; Hu & Bentler, 1999). Todos los análisis se realizaron utilizando Mplus v.7.4 (Muthén & Muthén, 1998-2014).

Las estimaciones se realizaron tanto en la muestra total como en ambas submuestras en tres condiciones: (a) Muestras completa (condición MC; n=3381 en las muestras 1 y 2, y n=6762 en la muestra total), (b) Muestra completa menos el subgrupo identificado con una respuesta asimétrica a los ítems positivos y negativos (AEMÉTODO, condición MC-M; n=3086 en la muestra 1, n=3091 en la muestra 2 y n=6177 en la muestra total), y (c) Muestra completa menos un número seleccionado al azar de participantes para igualar el tamaño de la muestra con la condición MC-M (condición MC-R; n=3086 en la muestra 1, n=3091 en la muestra 2 y n=6177 en la muestra total).

Resultados

Los índices de ajuste de los modelos por muestra y condición se presentan en la Tabla 4. Aquí, describimos los resultados para la muestra completa, pero notamos que éstos fueron consistentes en ambas muestras. Como se ha reportado previamente en la literatura, el modelo de un factor presentó un ajuste pobre ($\chi^2=2442,182$; RMSEA=0,102 [IC 90%=0,099, 0,106], CFI=0,870; TLI=0,828) mientras que el modelo Bifactor-FI ($\chi^2=778,602$; RMSEA=0,063 [90% CI=0,059, 0,067]; CFI=0,959; TLI=0,935) y el modelo Bifactor-FID ($\chi^2=284,372$; RMSEA=0,041 [90% CI= 0,037, 0,045]; CFI= 0,986; TLI= 0,972) presentaron un ajuste sustancialmente mejor. Se encontraron resultados similares en la condición MC-R (ver Tabla 4).

Por otro lado, el ajuste del modelo de un factor mejoró sustancialmente mediante la exclusión de los participantes del perfil AEMÉTODO (condición MC-M), alcanzando niveles aceptables de ajuste ($\chi^2=1292,359$; RMSEA=0,077 [90% CI=0,074, 0,081], CFI=0,927; TLI=0,903). El Bifactor-FI todavía muestra un mejor ajuste ($\chi^2=610,087$; RMSEA=0,058 [90% CI=0,054, 0,062], CFI=0,966; TLI=0,946) que el modelo de un factor, mientras que la estimación del modelo Bifactor-FID, después de este arreglo de muestreo, no convergió. Estos efectos se encontraron consistentemente en cada muestra generada al azar⁸.

[Inserte la Tabla 4 aproximadamente aquí]

A partir de las cargas factoriales estandarizadas (ver Material Suplementario, Tablas S1-S3), calculamos índices derivados del modelo bifactor, como la varianza común explicada (ECV; Brunner et al., 2012), el coeficiente omega (ω ; McDonald, 1999) y el coeficiente omega jerárquico (ω_h ; Zinbarg et al., 2005), y comparamos estos índices a

⁸ Para explorar los factores asociados con la falta de convergencia, llevamos a cabo análisis adicionales que examinaron los efectos del número de iteraciones, el criterio de convergencia, y la proporción del grupo de método reducido en el set de datos analizados. En general, encontramos que la falta de convergencia siguió siendo un problema hasta que el 60 % de la muestra del método se retuvo como parte de la muestra completa o hasta que el criterio de convergencia se elevó a 0,005. Hemos incluido los resultados de convergencia para estas condiciones en el Material Suplementario (Tablas S4 y S5). Esta falta de convergencia vendría dada por la incapacidad del algoritmo de estimación para obtener un conjunto de parámetros que permita reproducir, a través de un modelo ortogonal de tres factores (Modelo Bifactor-FID), un conjunto de datos predominantemente unidimensional (Tablas S6 y S7). Esto también se conoce como sobrefactorización (Barendse et al. 2015; Garrido et al. 2016).

través de condiciones bajo el modelo Bifactor-FI (que fue el único modelo bifactor que convergió). Para las condiciones MC y MC-R, la varianza común explicada del factor general fue .80 ($ECV_{MC}=.801$; $ECV_{MC-R}=.802$), lo que indica que el 80% de la varianza común se debe al factor general, mientras que en la condición MC-M este valor aumentó a 89% ($ECV_{MC-M}=.887$). Esta mejora en la estructura unidimensional también se obtuvo en términos de la confiabilidad compuesta multidimensional ponderada por unidad para el factor general ($\omega_{MC}=.907$; $\omega_{MC-R}=.908$; $\omega_{MC-M}=.913$), la proporción de la varianza en puntuación bruta compuesta ponderada por unidad atribuible al factor general ($\omega h_{MC}=\omega h_{MC-R}=.817$; $\omega h_{MC-M}=.869$), y la proporción de varianza confiable en las puntuaciones compuestas debido al factor general ($\omega h_{MC} / \omega_{MC}=\omega h_{MC-R} / \omega_{MC-R}=.900$; $\omega h_{MC-M} / \omega_{MC-M}=.952$).

Tomados en conjunto, estos resultados sugieren que los participantes con patrones de respuesta asimétricos a ítems positivos y negativos, previamente identificados en el Estudio 1, contribuyen al ajuste deficiente obtenido en el modelo de un factor, reduciendo la proporción de varianza capturada y la varianza confiable atribuible al factor de rasgo general, y a la convergencia del modelo Bifactor-FID. Estos resultados son consistentes con los reportados previamente por Reise et al. (2016), quienes identificaron una proporción de participantes cuyos patrones de respuesta serían "inmodelables" con una estructura factorial deficiente.

Estos hallazgos apuntan a un efecto general del perfil $AE_{MÉTODO}$ en el ajuste o desajuste del modelo unidimensional de la RSES y la ventaja relativa, o incluso la viabilidad, de utilizar una interpretación específica del modelo bifactor. El tercer y último estudio intenta cuantificar la contribución de casos específicos en las diferentes clases a estos resultados.

Estudio 3: Patrones de respuesta y ajuste individual

El Estudio 2 permitió evaluar el impacto de remover el subgrupo de individuos con respuestas asimétricas sobre el ajuste para un modelo de un factor. Este resultado refleja el efecto colectivo de este perfil.

Para cuantificar el impacto individual sobre el ajuste o desajuste de los modelos de un factor y Bifactor-FID, el Estudio 3 buscó identificar la proporción de participantes del perfil $AE_{MÉTODO}$ cuyos patrones de respuesta contribuyen a reproducir la matriz de varianza-covarianza implícita en cada modelo (ajuste individual), así como la proporción de participantes cuyos patrones de respuesta no están modelados adecuadamente por el modelo.

Además, el Estudio 3 tiene como objetivo evaluar la contribución individual de los participantes de $AE_{MÉTODO}$ a la diferencia en el ajuste a favor del modelo Bifactor-FID en comparación con el modelo de un factor.

Análisis de datos

Primero, para estimar el ajuste de un patrón de respuesta individual a un modelo específico, empleamos IRLS (Reise et al., 2016; Yuan & Bentler, 2000) utilizando la muestra completa ($n=6762$), bajo las especificaciones del modelo de un factor y Bifactor-FID mencionadas anteriormente. A partir de este análisis, obtuvimos la “distancia de inmodelabilidad” (d_r) para describir, en términos del tamaño de la expresión residual del modelo, qué tan predecible es un patrón de respuesta del individuo dado un modelo específico.

Basados en esta conceptualización, definimos un patrón de respuesta como menos predecible si el valor de d_r fue estadísticamente significativo (*i.e.*, tiene un valor- p , de una cola, menor a 0,025; Reise et al., 2016). Dada esta definición, la forma de juzgar el valor de modelos cada vez más complejos es calcular el porcentaje creciente de individuos modelados adecuadamente a medida que se aplican modelos más complejos.

En segundo lugar, para identificar la contribución individual a la prueba de razón de verosimilitud general de las diferencias entre modelos, comenzamos calculando un valor de log-likelihood para cada individuo bajo los modelos de un factor y Bifactor-FID (estimados previamente en el Estudio 2). La suma de estos valores da como resultado el valor de log-likelihood del modelo (-99963,313 en el modelo de un factor y -98308,200 en el modelo Bifactor-FID). Luego, la diferencia en estos valores de log-likelihood se multiplica por -2, lo que arroja la contribución de cada individuo a la prueba de razón de

verosimilitud global de las diferencias entre modelos (*i.e.*, IND_{CHI} ; Reise et al., 2016; Reise & Widaman, 1999). Valores positivos de IND_{CHI} reflejan un caso que contribuye a que el modelo Bifactor-FID se considere superior (un mejor ajuste) en relación con el modelo de un factor. Los valores de IND_{CHI} cercanos a cero indican un caso cuya contribución a un modelo específico es relativamente pequeña. Valores negativos indican que el patrón de respuesta es más probable bajo un modelo de un factor que el modelo Bifactor-FID (al reducir la diferencia de chi-cuadrado).

En ambos análisis, comparamos la proporción de desajustes de modelabilidad y la distribución de valores de IND_{CHI} entre los perfiles con patrones de respuestas simétricas (AE_{ALTA} , AE_{MED} , y AE_{BAJA}) y asimétricas ($AE_{MÉTOD}$).

Resultados

El panel superior de la Figura 3 muestra la distancia de inmodelabilidad con los modelos de un factor y Bifactor-FID por subconjuntos de perfiles con patrones de respuesta ordenados (AE_{ALTA} , AE_{MED} y AE_{BAJA}) y desordenados ($AE_{MÉTOD}$), y el valor crítico de chi-cuadrado (a nivel de $\alpha=0,025$) con 9 y 7 grados de libertad; el porcentaje de participantes con residuales significativos corresponde a la proporción de individuos que se consideran no modelados adecuadamente bajo un modelo específico.

[Inserte la Figura 3 aproximadamente aquí]

En el subconjunto de perfiles ordenados, el 37,2% de los participantes tenían residuos significativos y, por lo tanto, se consideraría modelado de manera inapropiada según el modelo de un factor, y un 5,3% adicional se modeló adecuadamente según el modelo Bifactor-FID. Por lo tanto, concluimos que el 31,9% de la muestra no se modela adecuadamente ni siquiera con el modelo más flexible. Por otro lado, solo el 13% de los participantes del perfil $AE_{MÉTOD}$ tienen residuos bajos en el modelo de un factor, y este valor sólo aumenta al 40% bajo el modelo Bifactor-FID. Más importante, estos resultados revelan que la mayoría (60%) de los participantes en el perfil $AE_{MÉTOD}$ no están bien modelados incluso por el modelo Bifactor-FID.

Con respecto a la contribución individual a la diferencia de chi-cuadrado entre los modelos de un factor y bifactor, la distribución de frecuencia de los valores de IND_{CHI} por

subconjuntos de perfiles con patrones de respuesta ordenadas (AE_{ALTA} , AE_{MED} , y AE_{BAJA}) y desordenadas ($AE_{MÉTODO}$) se muestran en el panel inferior de la Figura 3. Para el grupo de perfiles ordenados, el diagrama superior muestra una distribución simétrica de valores de IND_{CHI} (Asimetría=-0.188), y cuyos valores se aproximan a cero (Media=0,303; DE=1,690). Por lo tanto, es posible asumir una baja contribución de estos perfiles al mejor ajuste del modelo bifactor. Por el contrario, el diagrama inferior muestra una distribución sesgada a la derecha de los valores de IND_{CHI} para los participantes del perfil $AE_{MÉTODO}$ (Asimetría=1,411) con una mayor proporción de valores positivos (Media=2,456; DE=4,402), lo que indica que los participantes del perfil $AE_{MÉTODO}$ contribuyen significativamente a la mejora del ajuste del modelo bifactor (al aumentar la diferencia de chi-cuadrado).

Discusión

La presente serie de estudios ilustró un enfoque valioso para abordar los efectos de redacción basado en el supuesto de que estos efectos son producto de la heterogeneidad de la población. En el Estudio 1, mostramos que LPA podría revelar un subgrupo de participantes que exhiben un patrón de respuestas consistente con la noción de un efecto de redacción. Este hallazgo indica que los efectos de redacción pueden producir un patrón específico de respuestas que configura un grupo identificable. Esto implica que el LPA puede convertirse en una poderosa herramienta para detectar fácilmente (y eventualmente excluir o asignar un tratamiento analítico diferente) a los participantes que responden de manera inconsistente a ítems redactados de manera positiva y negativa.

Esta implicancia se investigó en el Estudio 2, en el que la exclusión de los participantes que exhibieron un efecto de redacción llevó a un cambio en el modelo de medición que mejor se ajusta a los datos. Es decir, el modelo que proporciona un mejor ajuste en la muestra completa (*i.e.*, el modelo bifactor) parece constituir, en gran medida, una acomodación a la anomalía generada en los datos por este grupo particular de participantes, pero no necesariamente refleja la verdadera naturaleza del constructo subyacente.

Más evidencia para esta conclusión fue aportada en el Estudio 3, en el que se demostró que los participantes que exhiben un efecto de redacción contribuyen principalmente al mejor ajuste general del modelo Bifactor-FID, aunque para la mayoría de los individuos en este grupo el patrón de respuesta inconsistente no puede ser modelados por ninguno de los modelos sugeridos en la literatura. En otras palabras, aunque un análisis tradicional que asuma homogeneidad poblacional podría indicar que el efecto de redacción está siendo adecuadamente capturado a través del modelo bifactor, nuestros hallazgos sugieren que la preferencia por el modelo bifactor puede conceptualizarse como el resultado del sobreajuste estadístico. Además, al ignorar el hecho de que el efecto de redacción emerge en un grupo particular de individuos, nos estamos quedando con un conjunto sustancial de participantes que no fueron modelados correctamente.

Sin embargo, esta investigación tiene potenciales limitaciones. Los efectos reportados aquí han sido identificados utilizando datos de panel representativos recolectados por Internet (panel LISS, Universidad de Tilburg). A pesar de su composición representativa y su estrategia de recopilación de datos (cada vez más común en la investigación psicológica), los resultados sobre la heterogeneidad en los efectos de redacción que se reportan aquí son incipientes y su generalización a otras poblaciones necesita ser examinada a fondo. Más allá de estas limitaciones, el conjunto actual de estudios tiene implicancias importantes para la investigación futura sobre los efectos de redacción que se analizan en la siguiente sección.

Modelos de heterogeneidad latente y naturaleza de los efectos de redacción

Con respecto a la naturaleza de los efectos de redacción, la identificación de un grupo específico de individuos quienes expresan un patrón de respuesta consistente con la noción de efectos de redacción propuesta previamente (Kamoen, Holleman, Mak, et al., 2011; Kamoen, Holleman, Nouwen, et al., 2011; Kamoen et al., 2013; Quilty et al., 2006), y cuyo impacto sobre el impacto en el desajuste global e individual es consistente con los hallazgos reportados en investigaciones previas, ofrece la oportunidad de examinar los efectos de redacción desde un enfoque estadístico alternativo a aquellos previamente

empleadas (que usan el modelo bifactor para separar la varianza de rasgo y de método). En los siguientes párrafos, ampliaremos esta descripción.

De manera más general, este estudio ilustra las ventajas de adoptar modelamiento de heterogeneidad latente para examinar los efectos de redacción y efectos asociados. La relevancia de este tipo de métodos se puede ver en al menos tres aspectos: (a) su considerable repertorio de extensiones que podrían aplicarse tanto al estudio de los efectos del estilos de respuesta como a la evaluación de la naturaliza sustantiva o efímera de los efectos de redacción, (b) una integración más estrecha entre los modelos de rasgos latentes cuantitativos y categóricos para dar cuenta de este tipo de efectos, y (c) su utilidad para generar resultados interpretables para generar resultados interpretables en contextos prácticos y, al mismo tiempo, dar cuenta de la heterogeneidad de la población.

Con respecto al primero de estos tres aspectos, este tipo de métodos incluye un conjunto amplio de extensiones estadísticas (Lanza et al., 2003; Porcu & Giambona, 2017) que pueden usarse para examinar la estabilidad del efecto de redacción en el tiempo (*i.e.*, utilizando análisis de transiciones latentes para estimar las probabilidades de transición entre los subgrupos $A_{EMÉTODO}$ identificados); la asociación del efecto de redacción con rasgos individuales (*i.e.*, utilizando regresión logística multinomial para estimar la probabilidad de pertenecer al subgrupo $A_{EMÉTODO}$ dado rasgos específicos como covariables); y la consistencia en la identificación de este subgrupo utilizando distintos atributos evaluados con escalas que contienen ítems redactados positiva y negativamente (*i.e.*, estimar las probabilidades de pertenecer al subgrupo $X_{MÉTODO}$ del atributo X dado que el individuo pertenece al subgrupo $Y_{MÉTODO}$ del atributo Y).

Si tal sesgo de respuesta es el resultado de un proceso de respuesta “similar a un rasgo”, se esperaría que exhibiera altas probabilidades de transición entre estados de subgrupos sesgados (al menos comparables con las transiciones entre estados que describen el rasgo específico de interés). Además, esperaríamos encontrar evidencia de detección consistente de participantes sesgados en escalas balanceadas y, más críticamente, una asociación consistente del efecto de redacción, utilizando diferentes escalas balanceadas.

El segundo aspecto donde podemos ver las ventajas de adoptar modelos de LCA y LPA es su potencial integración con modelos factoriales tradicionales. Los marcos de modelado global como FMM (Lubke & Muthén, 2005) o GLLAMM (Rabe-Hesketh et al., 2004; Skrondal & Rabe-Hesketh, 2004) permiten en principio combinar en un modelo de mixtura diferentes estructuras latentes dentro de diferentes clases. Esta flexibilidad podría usarse para acomodar modelos similares al modelo híbrido de Yamamoto, que se basa en una mezcla de un modelo de variable latente cuantitativa para dar cuenta del rasgo relevante que se está modelando, y una clase latente para capturar (en la aplicación original) la conducta de adivinación (Yamamoto, 1989; Yamamoto & Everson, 1995). Este tipo de estructura podría aplicarse a esta situación, donde basados en los resultados de este estudio y aquellos reportados en las Tablas Suplementarias S6 y S7, sería razonable esperar que los encuestados que fueron clasificados como parte de los grupos de autoestima baja, media y alta se asignarían en dicho modelo a una clase con alta probabilidad, mientras que una clase tridimensional describiría mejor a los encuestados asociados al subgrupo de método. Los modelos que aprovechan esta flexibilidad pueden representar más fielmente, en términos estadísticos, las distinciones conceptuales referidas a los efectos de sesgo de respuesta, y el uso de un modelo LPA, en este estudio, ilustra las ventajas potenciales de considerar este tipo de estructuras en esta área.

Finalmente, el tercer aspecto donde el uso de este tipo de modelos de mixturas ofrece ventajas está fuera del contexto de investigación sobre estilos de respuesta y sesgos, ya que puede proporcionar un marco relativamente simple para generar resultados interpretables en el ámbito clínico y, al mismo tiempo, dar cuenta del tema de la heterogeneidad poblacional (y más ampliamente con los problemas asociados a una simple interpretación unidimensional de la RSES).

Como lo señalaron Reise et al. (2016; p.836), enfoques alternativos como IRLS “son complementarios o compiten con el modelamiento de mixturas latentes, según el uso particular y la perspectiva teórica de cada uno sobre la naturaleza de los rasgos psicológicos”. Este artículo ilustra un área clara en la que el LPA puede servir como un prometedor complemento a los enfoques tradicionales, al mismo tiempo que muestra

algunas de las áreas en las que presenta ventajas sobre los métodos alternativos, particularmente en la presentación de informes de resultados en el contexto clínico.

Esperamos que estas líneas de investigación complementen aún más el correcto manejo estadístico de los efectos de redacción, de modo que la identificación y exploración de participantes con patrones de respuesta aberrantes o inconsistentes se convierta en una práctica estándar para el análisis estadístico en el futuro cercano.

Referencias

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. <http://dx.doi.org/10.1109/TAC.1974.1100705>
- Alessandri, G., Vecchione, M., Eisenberg, N., & Laguna, M. (2015). On the factor structure of the Rosenberg (1965) General Self-Esteem Scale. *Psychological Assessment*, 27(2), 621-536. <https://doi.org/10.1037/pas0000073>
- Alessandri, G., Vecchione, M., Tisak, J., & Barbaranelli, C. (2011). Investigating the nature of method factors through multiple informants: Evidence for a specific factor? *Multivariate Behavioral Research*, 46(4), 625-642. <https://doi.org/10.1080/00273171.2011.589272>
- Arias, V.B., & Arias, B. (2017). The negative wording factor of Core Self-Evaluations Scale (CSES): Methodological artifact, or substantive specific variance? *Personality and Individual Differences*, 109, 28-34. <https://doi.org/10.1016/j.paid.2016.12.038>
- Arias, V.B., Jenaro, C., & Ponce, F.P. (2018). Testing the generality of the general factor of personality: An exploratory bifactor approach. *Personality and Individual Differences*, 129, 17-23. <https://doi.org/10.1016/j.paid.2018.02.042>
- Bentler, P.M., Jackson, D.N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 76(3), 186-204. <https://doi.org/10.1037/h0031474>
- Bonifay, W., & Cai, L. (2017). On the complexity of Item Response Theory models. *Multivariate Behavioral Research*, 52(4), 465-484. <https://doi.org/10.1080/00273171.2017.1309262>
- Browne, M.W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258. <http://dx.doi.org/10.1177/0049124192021002005>
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80(4), 796-846. <https://doi.org/10.1111/j.1467-6494.2011.00749.x>

- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195-212.
<http://doi.org/10.1007/BF01246098>
- Clark, S.L. (2010). Mixture-modeling with behavioral data (Doctoral dissertation). University of California, Los Angeles, Los Angeles, CA.
- Collins, L.M., & Lanza, S.T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. New York: Wiley.
- Conroy, D.E. (2001). Progress in the development of a multidimensional measure of fear of failure: The Performance Failure Appraisal Inventory (PFAI). *Anxiety, Stress, and Coping*, 14, 431-452. <https://doi.org/10.1080/10615800108248365>
- Cronbach, L.J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3-31.
<https://doi.org/10.1177/001316445001000101>
- Croon, M. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology*, 43, 171-192.
<https://doi.org/10.1111/j.2044-8317.1990.tb00934.x>
- Croon, M. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of Mathematical and Statistical Psychology*, 44, 315-331. <https://doi.org/10.1111/j.2044-8317.1991.tb00964.x>
- Croon, M. (2002). Ordering the classes. In J.A. Hagenaars & A.L. McCutcheon (Eds.), *Applied latent class analysis*. Cambridge, UK: Cambridge University Press.
- DiStefano, C., & Motl, R.W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13(3), 440-464. https://doi.org/10.1207/s15328007sem1303_6
- DiStefano, C., & Motl, R.W. (2009a). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling*, 16, 134-146. <https://doi.org/10.1080/10705510802565403>

- DiStefano, C., & Motl, R.W. (2009b). Personality correlates of method effects due to negatively worded items on the Rosenberg self-esteem scale. *Personality and Individual Differences, 46*(3), 309-313. <https://doi.org/10.1016/j.paid.2008.10.020>
- Donnellan, M.B., Ackerman, R.A., & Brecheen, C. (2016). Extending structural analyses of the Rosenberg self-esteem scale to consider criterion-related validity: Can composite self-esteem scores be good enough? *Journal of Personality Assessment, 98*(2), 169-177. <http://dx.doi.org/10.1080/00223891.2015.1058268>
- Dunbar, M., Ford, G., Hunt, K., & Der, G. (2000). Question wording effects in the assessment of global self-esteem. *European Journal of Psychological Assessment, 16*(1), 13-19. <https://doi.org/10.1027/1015-5759.16.1.13>
- Gana, K., Saada, Y., Bailly, N., Joulain, M., Hervé, C., & Alaphilippe, D. (2013). Longitudinal factorial invariance of the Rosenberg Self-Esteem Scale: Determining the nature of method effects due to item wording. *Journal Research in Personality, 47*(4), 406-416. <https://doi.org/10.1016/j.jrp.2013.03.011>
- Gnamb, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg self-esteem scale: A cross-cultural meta-analysis. *Zeitschrift für Psychologie, 226*(1), 14-29. <https://doi.org/10.1027/2151-2604/a000317>
- Gnamb, T., Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment, 1*-15. <https://doi.org/10.1177/1073191117746503>
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S.P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem scale: Do they matter? *Personality and Individual Differences, 35*(6), 1241-1254. [https://doi.org/10.1016/S0191-8869\(02\)00331-8](https://doi.org/10.1016/S0191-8869(02)00331-8)
- Gu, H., Wen, Z., & Fan, X. (2015). The impact of wording effect on reliability and validity of the Core Self-Evaluation Scale (CSES): A bi-factor perspective. *Personality and Individual Differences, 83*, 142-147. <https://doi.org/10.1016/j.paid.2015.04.006>

- Holden, R.R., & Fekken, G.C. (1990). Structured psychopathological test item characteristics and validity. *Psychological Assessment*, 2(1), 35-40.
<http://dx.doi.org/10.1037/1040-3590.2.1.35>
- Horan, P.M., DiStefano, C., & Motl, R.W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling*, 10(3), 435-455. https://doi.org/10.1207/S15328007SEM1003_6
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. <http://dx.doi.org/10.1080/10705519909540118>
- Huang, J.L., Liu, M., & Bowling, N.A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100(3), 828-845. <http://dx.doi.org/10.1037/a0038510>
- Jackson, D.N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55(4), 243-252. <http://dx.doi.org/10.1037/h0045996>
- Kam, C.C.S. (2018). Testing the assumption of population homogeneity in the measurement of dispositional optimism: Factor mixture modeling analysis. *Journal of Personality Assessment*, 102(2), 238-249.
<https://doi.org/10.1080/00223891.2018.1502194>
- Kam, C.C.S. (2020). Bifactor model is not the best-fitting model for self-esteem: Investigation with a novel technique. *Assessment*, 1-14.
<https://doi.org/10.1177/1073191120949916>
- Kam, C.C.S., & Fan, X. (2018). Investigating response heterogeneity in the context of positively and negatively worded items by using factor mixture modeling. *Organizational Research Methods*, 1-12. <https://doi.org/10.1177/1094428118790371>
- Kam, C.C.S., & Meyer, J.P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, 18(3), 512-541.
<https://doi.org/10.1177/1094428115571894>

- Kamoen, N., Holleman, B., Mak, P., Sanders, T., & van den Bergh, H. (2011). Agree or disagree? Cognitive processes in answering contrastive survey questions. *Discourse Processes*, 48(5), 355-385. <http://dx.doi.org/10.1080/0163853X.2011.578910>
- Kamoen, N., Holleman, B., Nouwen, R., Sanders, T., & van den Bergh, H. (2011). Absolutely relative or relatively absolute? The linguistic behavior of gradable adjectives and degree modifiers. *Journal of Pragmatics*, 43(13), 3139-3151. <https://doi.org/10.1016/j.pragma.2011.05.017>
- Kamoen, N., Holleman, B., & van den Bergh, H. (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, 7(3), 181-189. <https://doi.org/10.18148/srm/2013.v7i3.5034>
- Lanza, S.T., Flaherty, B.P., & Collins, L.M. (2003). Latent class and latent transition analysis. In J.A. Schinka, & W.F. Velicer (Eds.). *Handbook of Psychology. Research Methods in Psychology* (Vol. 2, pp. 663-685). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.
- Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, 94(2), 196-204. <http://dx.doi.org/10.1080/00223891.2011.645936>
- Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778. <http://doi.org/10.1093/biomet/88.3.767>
- Lubke, G.H., & Muthén, B.O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 32-39. <https://doi.org/10.1037/1082-989X.10.1.21>
- Lubke, G.H., & Muthén, B.O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class specific parameters. *Structural Equation Modeling*, 14(1), 26-47. http://doi.org/10.1207/s15328007sem1401_2

- Maniaci, M.R., & Rogge, R.D. (2014). Caring about careless: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61-83.
<https://doi.org/10.1016/j.jrp.2013.09.008>
- Marsh, H.W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, 13(4), 335-361.
<https://doi.org/10.1177/014662168901300402>
- Marsh, H.W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, 70(4), 810-819. <https://doi.org/10.1037/0022-3514.70.4.810>
- Marsh, H.W., & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. *Structural Equation Modeling*, 1(4), 317–359. <https://doi.org/10.1080/10705519409539984>
- Marsh, H.W., & Grayson, D. (1995). Latent variable models of multitrait–multimethod data. In R.H. Hoyle (Ed.), *Structural equation modeling: Concept, issues, and applications* (pp. 177–198). Thousand Oaks, CA: Sage Publications, Inc.
- Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler’s (1999) findings. *Structural Equation Modeling*, 11(3), 320-341. https://doi.org/10.1207/s15328007sem1103_2
- Marsh, H.W., Lüdtke, O., Trautwein, U., & Morin, A.J.S. (2009). Classic latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, 16(2), 191-225. <https://doi.org/10.1080/10705510902751010>
- Marsh, H.W., Scalas, L.F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment*, 22(2), 366-381.
<https://doi.org/10.1037/a0019225>
- McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.

- Michaelides, M.P., Koutsogiorgi, C., & Panayiotou, G. (2016). Method effects on an adaptation of the Rosenberg self-esteem scale in Greek and the role of personality traits. *Journal of Personality Assessment*, 98(2), 178-188.
<http://dx.doi.org/10.1080/00223891.2015.1089248>
- Michaelides, M.P., Zenger, M., Koutsogiorgi, C., Brähler, E., Stöbel-Richter, Y., & Berth, H. (2016). Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Personality and Individual Differences*, 97, 13-18. <https://doi.org/10.1016/j.paid.2016.03.011>
- Mislevy, R.J., & Verhelst, N. (1990). Modeling items responses when different subjects employ different solution strategies. *Psychometrika*, 55(2), 195-215.
<https://doi.org/10.1007/BF02295283>
- Motl, R.W., & Conroy, D.E. (2000). Validity and factorial invariance of the Social Physique Anxiety Scale. *Medicine and Science in Sports Exercise*, 32, 1007–1017.
- Motl, R.W., Conroy, D.E., & Horan, P.M. (2000). The social physique anxiety scale: An example of the potential consequences of negatively worded items in factorial validity studies. *Journal of Applied Measurement*, 1(4), 327–345.
- Motl, R.W., Dishman, R.K., Felton, G., & Pate, R.R. (2003). Self-motivation and physical activity among black and white adolescent girls. *Medicine and Science in Sports and Exercise*, 35, 128-136.
<https://doi.org/10.1249/01.MSS.0000043478.38705.CD>
- Motl, R.W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and methods effects associated with negatively worded items. *Structural Equation Modeling*, 9(4), 562-578. https://doi.org/10.1207/S15328007SEM0904_6
- Muthén, L.K., & Muthén, B.O. (1998-2014). *Mplus User's Guide* (7th Ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, B., & Muthén, L.K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24(6), 882–891. <http://doi.org/10.1111/j.1530-0277.2000.tb02070.x>

- Neely-Barnes, S. (2010). Latent class models in social work. *Social Work Research*, 34(2), 114-121. <http://doi.org/10.1093/swr/34.2.114>
- Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- Osborne, D., & Sibley, C.G. (2017). Identifying “types” of ideologies and intergroup biases: Advancing a person-centred approach to social psychology. *European Review of Social Psychology*, 28(1), 288-332. <http://dx.doi.org/10.1080/10463283.2017.1379265>
- Pastor, D.A., Barron, K.E., Miller, B.J., & Davis, S.L. (2007). A latent profile analysis of college students’ achievement goal orientation. *Contemporary Educational Psychology*, 32(1), 8-47. <https://doi.org/10.1016/j.cedpsych.2006.10.003>
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. <http://dx.doi.org/10.1037/0021-9010.88.5.879>
- Porcu, M., & Giambona, F. (2017). Introduction to latent class analysis with applications. *Journal of Early Adolescence*, 37(1), 129-158. <https://doi.org/10.1177/0272431616648452>
- Quilty, L.C., Oakman, J.M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling*, 13(1), 99-117. http://dx.doi.org/10.1207/s15328007sem1301_5
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69, 167-190. <https://doi.org/10.1007/BF02295939>
- Reise, S.P., Kim, D.S., Mansolf, M., & Widaman, K.F. (2016). Is the bifactor model a better model or is it just better a modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg self-esteem scale. *Multivariate*

- Behavioral Research*, 51(6), 818-838.
<http://dx.doi.org/10.1080/00273171.2016.1243461>
- Reise, S.P., & Widaman, K.F. (1999). Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. *Psychological Methods*, 4(1), 3-21. <http://doi.org/10.1037/1082-989x.4.1.3>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Scherpenzeel, A. (2011). Data collection in a probability-based Internet panel: How the LISS Panel was built and how it can be used. *Bulletin de Méthodologie Sociologique*, 109, 56-61. <http://dx.doi.org/10.1177/0759106310387713>
- Schriesheim, C.A., Eisenbach, R.J., & Hill, K.D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, 51(1), 67-78. <https://doi.org/10.1177/0013164491511005>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464. <http://dx.doi.org/10.1214/aos/1176344136>
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
- Tomás, J.M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling*, 6(1), 84-98.
<https://doi.org/10.1080/10705519909540120>
- Tomás, J.M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling*, 20(2), 299-313.
<https://doi.org/10.1080/10705511.2013.769394>

- Vermunt, J.K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis*. Cambridge, England: Cambridge University Press.
- Weems, G.H., Onwuegbuzie, A.J., & Collins, K.M.T. (2006). The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation & Research in Education*, 19(1), 3-20.
<https://doi.org/10.1080/09500790608668322>
- Wood, D., Harms, P.D., Lowman, G.H., & DeSimone, J.A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychology and Personality Science*, 8(4), 454-464.
<https://doi.org/10.1177/1948550617703168>
- Yamamoto, K. Y. (1989). *HYBRID model of IRT and latent class models*. (ETS Research Report No.RR-89-41). Princeton, NJ: Educational Testing Service.
- Yamamoto, K. Y., & Everson, H. T. (1995). *Estimating the effects of test length and test time on parameter estimation using the HYBRID model*. (TOEFL Technical Report No.TOEFL-TR-10). Princeton, NJ: Educational Testing Service.
- Yuan, K.-H., & Bentler, P.M. (2000). Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika*, 65(1), 43-58.
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123-133.
<http://doi.org/10.1007/s11336-003-0974-7>

Tabla 1. Estadísticos descriptivos de los ítems de la RSES por muestra.

Ítems de la RSES	Muestra 1 (n=3381)				Muestra 2 (n=3381)			
	Media	DE	Asim	Curt	Media	DE	Asim	Curt
1. Siento que soy una persona digna de aprecio, al menos en igual medida que los demás.	5.976	1.145	-1.408	2.415	5.991	1.139	-1.356	2.050
2. Creo que tengo un buen número de cualidades.	6.012	0.956	-1.257	2.937	6.014	0.938	-1.150	2.171
3. En general me inclino a pensar que soy un/a fracasado/a.*	5.716	1.496	-1.189	0.601	5.753	1.473	-1.186	0.568
4. Soy capaz de hacer las cosas tan bien como la mayoría de la gente.	5.669	1.159	-1.004	1.153	5.636	1.148	-1.011	1.266
5. Siento que no tengo muchos motivos para sentirme orgulloso/a de mi.*	5.667	1.531	-1.188	0.494	5.688	1.513	-1.205	0.629
6. Tengo una actitud positiva hacia mi mismo/a.	5.466	1.239	-0.901	0.614	5.449	1.283	-0.900	0.520
7. En general, estoy satisfecho/a conmigo mismo/a.	5.601	1.128	-1.111	1.608	5.581	1.166	-1.145	1.639
8. Desearía valorarme más a mi mismo/a.*	4.985	1.832	-0.613	-0.836	4.978	1.851	-0.580	-0.898
9. A veces me siento verdaderamente inútil.*	5.586	1.592	-1.103	0.276	5.588	1.603	-1.114	0.334
10. A veces pienso que no soy bueno/a para nada.*	5.915	1.469	-1.459	1.508	5.919	1.464	-1.495	1.531

* Ítems redactados negativamente.

Tabla 2. Índices de ajuste para soluciones confirmatorias entre dos y seis perfiles.

Muestras		AIC	BIC	adjBIC	VLMR	LMR	Entropía	m%
Muestra 1	2-perfiles	104350.252	104540.156	104441.655	<.001	<.001	.929	26.56%
	3-perfiles	102063.305	102320.594	102187.140	.005	.005	.852	15.26%
	4-perfiles	100527.275	100851.949	100683.544	.002	.002	.873	8.73%
	5-perfiles	99484.808	99876.867	99673.509	.001	.001	.862	3.73%
	6-perfiles	98651.787	99111.231	98872.921	.455	.459	.875	3.55%
Muestra 2	2-perfiles	104512.929	104702.833	104604.332	<.001	<.001	.916	29.93%
	3-perfiles	101930.070	102187.359	102053.905	.001	.001	.865	13.81%
	4-perfiles	100429.235	100753.909	100585.503	<.001	<.001	.887	8.58%
	5-perfiles	99359.505	99751.564	99548.207	.021	.023	.884	4.14%
	6-perfiles	98550.070	99009.515	98771.205	.239	.244	.880	4.38%

Note. AIC=Criterio de Información de Akaike; BIC=Criterio de Información bayesiano; adjBIC=BIC ajustado por tamaño de muestra; VLMR=Test de Vuong-Lo-Mendell-Rubin; LMR=Test de Lo-Mendell-Rubin; m%=Proporción de participantes de la clase con menor prevalencia.

Tabla 3. Probabilidad media de clasificación al perfil latente (ALPCP) y prevalencia por muestra.

Muestras	Perfiles	ALPCP				Prevalencia	
		p1	p2	p3	p4	n	%
Muestra 1	p1. AE _{ALTA}	0.947	0.052	0.000	0.001	1763	52.14
	p2. AE _{MED}	0.079	0.895	0.009	0.016	980	28.99
	p3. AE _{BAJA}	0.000	0.023	0.939	0.038	343	10.15
	p4. AE _{MÉTODO}	0.004	0.049	0.038	0.909	295	8.73
Muestra 2	p1. AE _{ALTA}	0.954	0.046	0.000	0.001	1766	52.23
	p2. AE _{MED}	0.064	0.908	0.010	0.018	997	29.45
	p3. AE _{BAJA}	0.000	0.024	0.942	0.034	328	9.70
	p4. AE _{MÉTODO}	0.004	0.057	0.034	0.905	290	8.58

Tabla 4. Índices de ajuste para soluciones confirmatorias por condición.

Muestras	Condición	Modelo	Índices de ajuste					
			χ^2	gl	CFI	TLI	RMSEA[IC _{90%}]	SRMR
Muestra 1	MC (n=3381)	Un Factor	1,179.140***	33	0.878	0.834	0.101[0.096-0.106]	0.067
		Bifactor-FI	456.056***	28	0.955	0.927	0.067[0.062-0.073]	0.037
		Bifactor-FID	146.357***	23	0.987	0.974	0.040[0.034-0.046]	0.018
	MC-R (n=3086)	Un Factor	1,063.210***	33	0.879	0.835	0.101[0.095-0.106]	0.066
		Bifactor-FI	429.065***	28	0.953	0.924	0.068[0.063-0.074]	0.037
		Bifactor-FID	141.241***	23	0.986	0.973	0.041[0.034-0.047]	0.018
	MC-M (n=3086)	Un Factor	596.568***	33	0.933	0.909	0.074[0.069-0.080]	0.041
		Bifactor-FI	322.042***	28	0.965	0.944	0.058[0.053-0.064]	0.030
		Bifactor-FID			No converge			
Muestra 2	MC (n=3381)	Un Factor	1,008.263***	33	0.893	0.855	0.093[0.089-0.099]	0.061
		Bifactor-FI	349.497***	28	0.965	0.944	0.058[0.053-0.064]	0.031
		Bifactor-FID	163.068***	23	0.985	0.970	0.042[0.036-0.049]	0.017
	MC-R (n=3091)	Un Factor	896.360***	33	0.895	0.857	0.092[0.087-0.097]	0.061
		Bifactor-FI	304.111***	28	0.966	0.946	0.056[0.051-0.062]	0.031
		Bifactor-FID	145.233***	23	0.985	0.971	0.041[0.035-0.048]	0.017
	MC-M (n=3091)	Un Factor	626.686***	33	0.933	0.909	0.076[0.071-0.082]	0.040
		Bifactor-FI	321.279***	28	0.967	0.947	0.058[0.053-0.064]	0.026
		Bifactor-FID			No converge			
Ambas muestras	MC (n=6762)	Un Factor	2,442.182***	33	0.870	0.828	0.102[0.099-0.106]	0.184
		Bifactor-FI	778.602***	28	0.959	0.935	0.063[0.059-0.067]	0.034
		Bifactor-FID	284.372***	23	0.986	0.972	0.041[0.037-0.045]	0.017
	MC-R (n=6177)	Un Factor	2,212.422***	33	0.872	0.831	0.102[0.098-0.105]	0.183
		Bifactor-FI	700.240***	28	0.961	0.937	0.062[0.058-0.066]	0.033
		Bifactor-FID	265.360***	23	0.986	0.972	0.041[0.037-0.046]	0.017
	MC-M (n=6177)	Un Factor	1,292.359***	33	0.927	0.903	0.077[0.074-0.081]	0.120
		Bifactor-FI	610.087***	28	0.966	0.946	0.058[0.054-0.062]	0.028
		Bifactor-FID			No converge			

Nota. χ^2 =Test de chi-cuadrado de ajuste del modelo; CFI=Índice de ajuste comparativo; TLI=Índice de Tucker-Lewis; RMSEA=Raíz del error cuadrático medio de aproximación; SRMR=Raíz cuadrática media residual estandarizada.

* p < 0.05; ** p < 0.01; *** p < 0.001

Figura 1. Representación conceptual de las fuentes de varianza asumiendo homogeneidad poblacional.

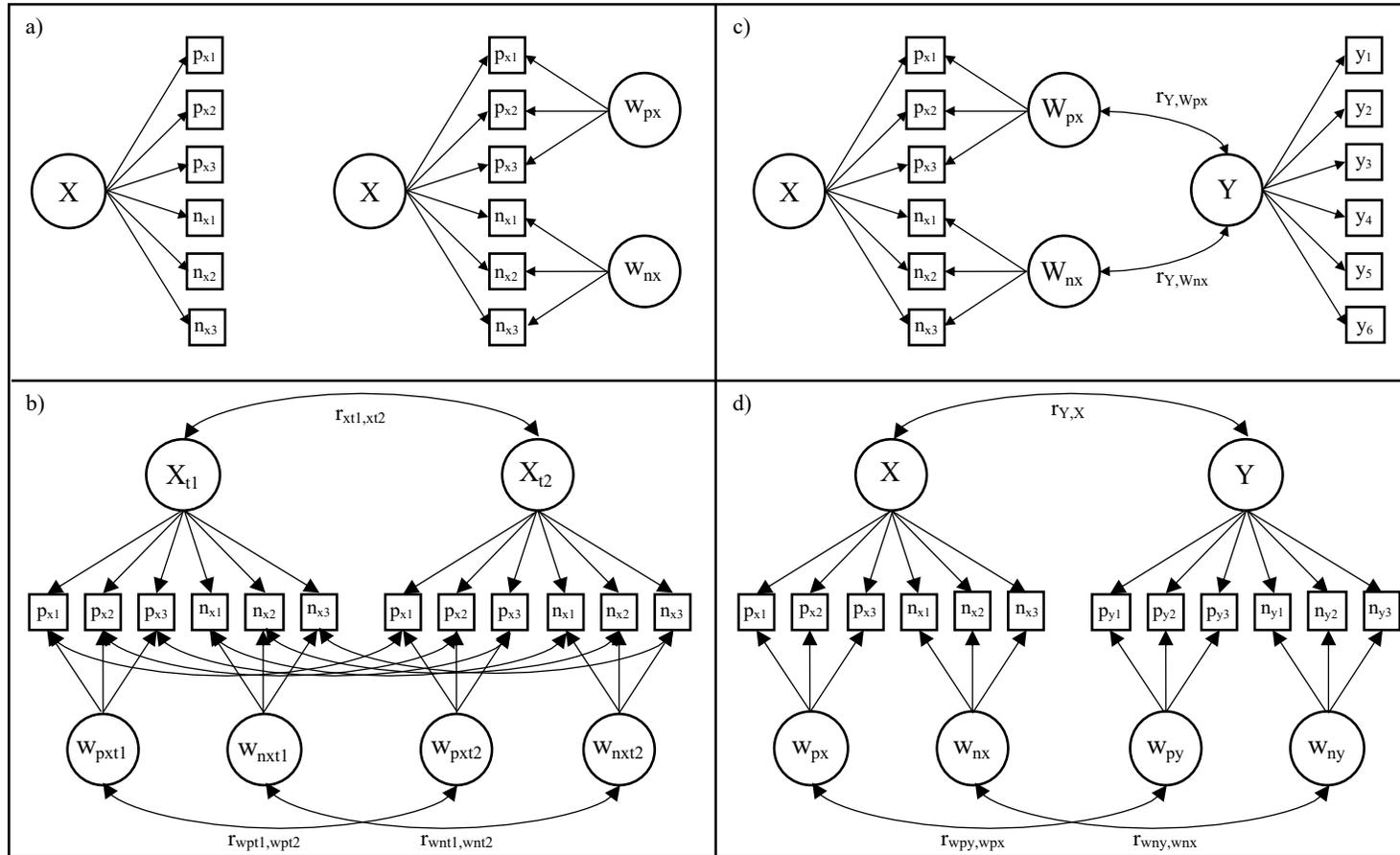


Figura 2. Medias estimadas para los ítems positivos y negativos de la RSES por perfiles identificados en ambas muestras.

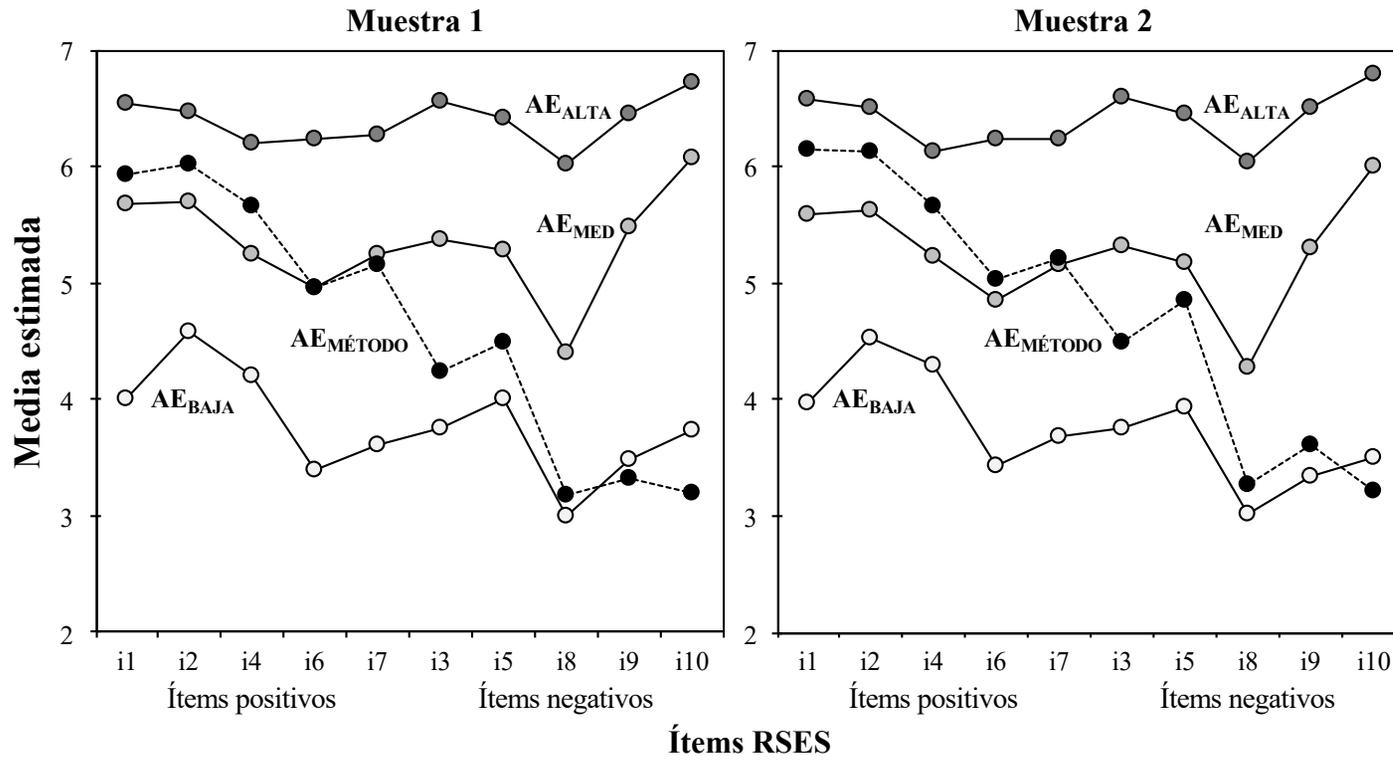
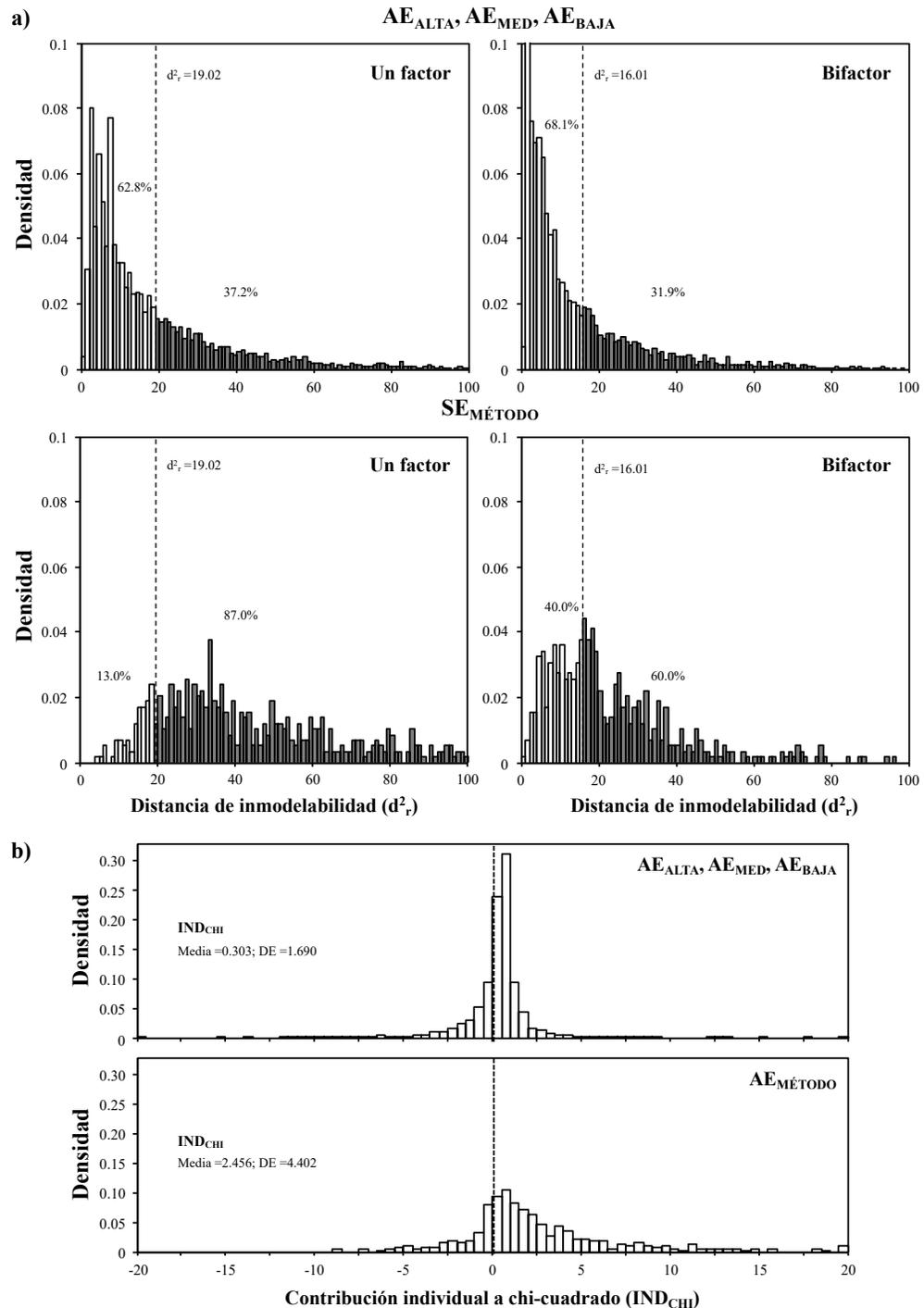


Figura 3. Panel a. Distribución de los valores de distancia de inmodelabilidad (d_r^2) con los modelos de un factor y bifactor para los perfiles ordenados y el perfil $AE_{MÉTOD}$. Panel b. Distribución de la contribución individual a los valores de chi-cuadrado (IND_{CHI}) para los perfiles ordenados y el perfil $AE_{MÉTOD}$.



Material Suplementario

Efectos de redacción en evaluación: Cuando el bosque no deja ver los árboles

En esta sección, presentamos los análisis suplementarios respecto de los resultados presentados en el Estudio 2. El primer grupo de tablas muestra las cargas factoriales estandarizadas extraídas de los análisis CFA en todas las condiciones muestrales (MC, MC-R, y MC-M) para ambas muestras 1 y 2 (Tablas S1 y S2, respectivamente) y para la muestra completa (Tabla S3). Todos los modelos fueron estimados utilizando Robust Maximum Likelihood (MLR) como método de estimación.

Luego examinamos la robustez de la falta de convergencia reportada para el modelo Bifactor-FID en la condición MC-M a través de examinar la frecuencia de este resultado luego de variar tanto los parámetros de estimación (*i.e.*, número de iteraciones y valores de criterio de convergencia) como las condiciones muestrales (ver Tablas S4 y S5). A su vez, utilizamos análisis paralelo (Horn, 1965; Tablas S6 y S7) para explorar el número hipotético de componentes (factores) en las condiciones MC-R y MC-M ($n=6177$), y en la muestra de participantes identificados con patrones de respuesta asimétricas para ítems redactados positiva y negativamente (perfil AEMÉTODO; $n=585$).

Cargas factoriales estandarizadas

A partir de los perfiles identificados en el Estudio 1, comparamos los resultados de ajuste de las diferentes estructuras factoriales (modelos de un factor, Bifactor-FI, y Bifactor-FID) mientras: (a) incluimos los casos asociados con la clase con patrones de respuesta asimétricas (condición MC), (b) excluimos esos casos (condición MC-M), y (c) excluimos una muestra aleatoria de casos (condición MC-R). Esos nueve modelos fueron estimados en cada muestra generada aleatoriamente y utilizada en el Estudio 1 ($n=3381$ cada uno) y con la muestra completa ($n=6762$).

Todos los modelos fueron estimados utilizando Robust Maximum Likelihood (MLR) como método de estimación (Bondades de ajuste fueron reportadas en la Tabla 4 en el manuscrito. Las cargas factoriales estandarizadas para cada modelo se reportaron en las Tablas S1-S3.

Tabla Suplementaria S1. Cargas factoriales estandarizadas y error estándar (en paréntesis) en la muestra 1 por condición.

Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
MC (n=3381)	Item1	.588(.019)***	.686(.018)***		.556(.021)***		.496(.035)***
	Item2	.540(.020)***	.667(.018)***		.489(.021)***		.660(.033)***
	Item3*	.731(.014)***	.632(.018)***	.329(.027)***	.746(.015)***	.080(.037)*	
	Item4	.501(.021)***	.611(.020)***		.448(.021)***		.510(.028)***
	Item5*	.614(.016)***	.554(.018)***	.231(.028)***	.626(.018)***	.028(.036)	
	Item6	.701(.016)***	.785(.014)***		.698(.018)***		.267(.030)***
	Item7	.700(.015)***	.770(.015)***		.694(.017)***		.244(.029)***
	Item8*	.673(.013)***	.574(.018)***	.321(.026)***	.687(.016)***	.085(.037)*	
	Item9*	.756(.016)***	.548(.020)***	.646(.019)***	.672(.024)***	.603(.094)***	
	Item10*	.786(.014)***	.595(.019)***	.615(.022)***	.717(.022)***	.412(.083)***	
	θ _{1,2}	.544(.022)***	.425(.028)***		.234(.069)**		
	θ _{6,7}	.561(.023)***	.435(.034)***		.504(.026)***		
	ECV	-	.798	.202	.717	.096	.187
	ω	.886	.906		.915		
ω _h	-	.815		.793			
Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
MC-R (n=3086)	Item1	.582(.020)***	.681(.019)***		.546(.022)***		.506(.036)***
	Item2	.544(.021)***	.668(.019)***		.491(.022)***		.654(.035)***
	Item3*	.734(.014)***	.629(.019)***	.340(.029)***	.747(.016)***	.081(.039)*	
	Item4	.500(.022)***	.611(.021)***		.444(.022)***		.514(.030)***
	Item5*	.613(.017)***	.555(.019)***	.227(.030)***	.630(.018)***	.010(.035)	
	Item6	.695(.017)***	.777(.015)***		.688(.019)***		.269(.031)***
	Item7	.699(.016)***	.767(.015)***		.690(.017)***		.243(.030)***
	Item8*	.669(.014)***	.574(.019)***	.312(.028)***	.688(.016)***	.065(.035)	
	Item9*	.755(.017)***	.550(.021)***	.641(.020)***	.678(.025)***	.605(.111)***	
	Item10*	.784(.015)***	.594(.020)***	.611(.023)***	.721(.021)***	.394(.088)***	
	θ _{1,2}	.547(.023)***	.431(.029)***		.237(.073)**		
	θ _{6,7}	.557(.025)***	.435(.035)***		.504(.027)***		
	ECV	-	.799	.201	.718	.093	.189
	ω	.885	.905		.915		
ω _h	-	.815		.793			

Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
MC-M (n=3086)	Item1	.686(.017)***	.717(.017)***				
	Item2	.646(.017)***	.692(.017)***				
	Item3*	.708(.016)***	.670(.018)***	.201(.032)***			
	Item4	.599(.019)***	.636(.019)***				
	Item5*	.598(.017)***	.585(.019)***	.098(.030)**			
	Item6	.762(.013)***	.790(.014)***				
	Item7	.762(.013)***	.784(.014)***				
	Item8*	.628(.015)***	.596(.018)***	.173(.030)***		No converge	
	Item9*	.708(.018)***	.604(.021)***	.595(.035)***			
	Item10*	.790(.014)***	.718(.016)***	.393(.029)***			
	$\theta_{1,2}$.465(.026)***	.409(.029)***				
	$\theta_{6,7}$.487(.026)***	.434(.032)***				
	ECV	-	.888	.112			
ω	.901	.910					
ω_h	-	.870					

Nota. Los ítems inversos están marcados con un asterisco.

* $p < ,05$; ** $p < ,01$; *** $p < ,001$

Tabla Suplementaria S2. Cargas factoriales estandarizadas y error estándar (en paréntesis) en la muestra 2 por condición.

Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
MC (n=3381)	Item1	.614(.018)***	.707(.017)***		.595(.022)***		.427(.038)***
	Item2	.580(.018)***	.693(.016)***		.547(.020)***		.550(.033)***
	Item3*	.730(.014)***	.628(.018)***	.341(.026)***	.739(.020)***	.102(.052)	
	Item4	.485(.020)***	.591(.020)***		.432(.022)***		.521(.030)***
	Item5*	.626(.016)***	.577(.017)***	.220(.025)***	.644(.020)***	.020(.051)	
	Item6	.696(.015)***	.772(.015)***		.691(.020)***		.269(.031)***
	Item7	.689(.015)***	.762(.015)***		.684(.020)***		.258(.032)***
	Item8*	.676(.013)***	.570(.017)***	.342(.024)***	.676(.018)***	.126(.047)**	
	Item9*	.760(.015)***	.567(.018)***	.613(.021)***	.683(.027)***	.479(.119)***	
	Item10*	.787(.013)***	.602(.018)***	.606(.021)***	.713(.024)***	.486(.107)***	
	θ1,2	.531(.021)***	.407(.026)***		.320(.043)***		
	θ6,7	.522(.024)***	.395(.034)***		.457(.029)***		
	ECV	-	.805	.195	.751	.089	.161
ω	.886	.907		.913			
ωh	-	.819		.804			
Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
MC-R (n=3091)	Item1	.609(.019)***	.701(.018)***		.597(.026)***		.407(.046)***
	Item2	.583(.018)***	.697(.017)***		.559(.023)***		.532(.039)***
	Item3*	.727(.015)***	.623(.018)***	.346(.027)***	.722(.020)***	.141(.049)**	
	Item4	.486(.022)***	.594(.021)***		.441(.024)***		.521(.034)***
	Item5*	.641(.016)***	.590(.018)***	.227(.026)***	.654(.021)***	.042(.061)	
	Item6	.688(.016)***	.766(.016)***		.692(.024)***		.258(.039)***
	Item7	.687(.016)***	.756(.016)***		.693(.024)***		.231(.040)***
	Item8*	.677(.013)***	.572(.018)***	.341(.026)***	.673(.022)***	.138(.061)*	
	Item9*	.759(.016)***	.566(.019)***	.607(.022)***	.677(.034)***	.453(.103)***	
	Item10*	.783(.014)***	.594(.019)***	.614(.022)***	.695(.027)***	.543(.070)***	
	θ1,2	.526(.022)***	.402(.028)***		.327(.044)***		
	θ6,7	.519(.026)***	.397(.035)***		.450(.032)***		
	ECV	-	.804	.197	.751	.098	.152
ω	.889	.907		.913			
ωh	-	.818		.804			

Condición	Ítem	Un-Factor		Bifactor-FI		Bifactor-FID		
		G		G	FI	G	FI	FD
FS-M (n=3091)	Item1	.715(.015)***		.755(.016)***				
	Item2	.683(.014)***		.737(.014)***				
	Item3*	.725(.015)***		.670(.018)***	.248(.033)***			
	Item4	.558(.019)***		.608(.019)***				
	Item5*	.630(.017)***		.600(.018)***	.173(.028)***			
	Item6	.746(.014)***		.780(.014)***				
	Item7	.737(.014)***		.772(.015)***				
	Item8*	.638(.014)***		.593(.018)***	.218(.029)***			No converge
	Item9*	.737(.016)***		.620(.019)***	.556(.028)***			
	Item10*	.830(.010)***		.747(.013)***	.419(.025)***			
	01,2	.430(.025)***		.341(.029)***				
	06,7	.475(.027)***		.406(.033)***				
	ECV	-		.885	.115			
ω	.907		.916					
ωh	-		.868					

Nota. Los ítems inversos están marcados con un asterisco.

* p < ,05; ** p < ,01; *** p < ,001

Tabla Suplementaria TS3. Cargas factoriales estandarizadas y error estándar (en paréntesis) en ambas muestras por condición.

Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
FS (n=6762)	Item1	.739(.006)***	.697(.012)***		.575(.015)***		.461(.025)***
	Item2	.669(.009)***	.680(.012)***		.517(.014)***		.604(.023)***
	Item3*	.765(.009)***	.629(.013)***	.335(.019)***	.744(.012)***	.084(.034)*	
	Item4	.545(.014)***	.602(.014)***		.439(.015)***		.516(.021)***
	Item5*	.661(.011)***	.564(.013)***	.227(.019)***	.633(.013)***	.028(.028)	
	Item6	.743(.014)***	.779(.010)***		.694(.014)***		.269(.021)***
	Item7	.739(.009)***	.766(.011)***		.688(.013)***		.252(.021)***
	Item8*	.711(.009)***	.572(.012)***	.332(.018)***	.681(.011)***	.107(.025)***	
	Item9*	.783(.011)***	.557(.014)***	.630(.014)***	.678(.017)***	.535(.064)***	
	Item10*	.810(.009)***	.598(.013)***	.611(.015)***	.716(.016)***	.448(.061)***	
	θ _{1,2}	.530(.016)***	.416(.019)***		.285(.038)***		
	θ _{6,7}	.532(.017)***	.414(.024)***		.480(.019)***		
	ECV	-	.801	.199	.736	.090	.174
	ω	.914	.907		.914		
ω _h	-	.817		.798			
Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
FS-R (n=6177)	Item1	.743(.006)***	.705(.013)***		.580(.015)***		.467(.026)***
	Item2	.672(.009)***	.686(.013)***		.521(.015)***		.607(.024)***
	Item3*	.764(.009)***	.626(.013)***	.343(.019)***	.743(.013)***	.083(.037)*	
	Item4	.549(.015)***	.607(.015)***		.445(.016)***		.508(.021)***
	Item5*	.660(.011)***	.560(.013)***	.236(.020)***	.632(.013)***	.031(.028)	
	Item6	.741(.010)***	.774(.011)***		.690(.014)***		.270(.022)***
	Item7	.742(.010)***	.769(.011)***		.688(.014)***		.259(.022)***
	Item8*	.712(.009)***	.570(.013)***	.340(.018)***	.682(.012)***	.109(.027)***	
	Item9*	.787(.011)***	.561(.014)***	.631(.015)***	.684(.018)***	.563(.068)***	
	Item10*	.811(.009)***	.601(.014)***	.600(.016)***	.722(.017)***	.409(.064)***	
	θ _{1,2}	.524(.017)***	.404(.021)***		.265(.041)***		
	θ _{6,7}	.537(.018)***	.422(.025)***		.487(.020)***		
	ECV	-	.802	.198	.737	.089	.174
	ω	.915	.908		.915		
ω _h	-	.817		.799			

Condición	Ítem	Un-Factor	Bifactor-FI		Bifactor-FID		
		G	G	FI	G	FI	FD
FS-M (n=6177)	Item1	.771(.006)***	.736(.011)**				
	Item2	.716(.009)***	.714(.011)***				
	Item3*	.739(.011)***	.669(.013)***	.225(.023)***			
	Item4	.610(.013)***	.622(.014)***				
	Item5*	.641(.012)***	.592(.013)***	.138(.021)***			
	Item6	.778(.009)***	.786(.010)***				
	Item7	.774(.009)***	.778(.010)***				
	Item8*	.659(.010)***	.595(.012)***	.197(.020)***		No converge	
	Item9*	.744(.012)***	.612(.014)***	.571(.021)***			
	Item10*	.828(.008)***	.732(.010)***	.410(.019)***			
	θ1,2	.439(.018)***	.378(.021)***				
	θ6,7	.479(.019)***	.418(.023)***				
	ECV	-	.887	.113			
ω	.918	.913					
ωh	-	.869					

Nota. Los ítems inversos están marcados con un asterisco.

* p < ,05; ** p < ,01; *** p < ,001

Falta de convergencia en la condición MC-M

Con el propósito de examinar la falta de convergencia encontrada con el modelo Bifactor-FID en la condición MC-M, evaluamos la robustez de este hallazgo, variando el número de iteraciones, el criterio de convergencia, y la proporción de participantes del grupo $AE_{MÉTODO}$ removidos de la muestra. En primer lugar, incrementamos el número de iteraciones en tres condiciones espaciadas logarítmicamente (IT=1000, IT=10000, y IT=100000). En segundo lugar, incrementamos el criterio de convergencia del algoritmo Quasi-Newton, usado a menudo para variables continuas, en saltos de 0,005 en cinco condiciones (CONV=.00005, CONV=.0001, CONV=.0005, CONV=.001, CONV=.005). Finalmente, la proporción de participantes del grupo $AE_{MÉTODO}$ (n=585) fue aleatoriamente seleccionada e incluida en saltos de 20% (n=177).

Para controlar los efectos de tamaño muestral, todas las estimaciones fueron llevadas a cabo emparejando el tamaño muestral de la condición MC-M (n=6177). La Tabla S4 muestra la distribución de frecuencia de participantes del grupo $AE_{MÉTODO}$ removidos en cada condición muestral, y la Tabla S5 ilustra aquellos casos en que se alcanzó una solución convergente.

Tabla Suplementaria S4. Distribución de frecuencias para los perfiles identificados por condición muestral.

Perfiles	Proporción de participantes del grupo $AE_{MÉTODO}$ removidos de la muestra					
	100%	80%	60%	40%	20%	0%
AE_{ALTA} ; AE_{MED} ; AE_{BAJA}	6177	6060	5943	5826	5709	5592
$AE_{MÉTODO}$	0	177	234	351	468	585
Total	6177	6177	6177	6177	6177	6177

Nota. La primera columna (100%) es la condición MC-M utilizada en el Estudio 2 y la última columna (0%) es similar a la condición MC-R.

Tabla Suplementaria S5. Estimaciones de convergencia a partir del modelo Bifactor-FID por condición muestral, número de iteraciones y criterios de convergencia.

Número de iteraciones = 1000						
Criterio de convergencia	Proporción de participantes del grupo AEMÉTODO removidos de la muestra					
	100%	80%	60%	40%	20%	0%
CONV = 0.00005	nConv	nConv	nConv	Conv	Conv	Conv
CONV = 0.0001	nConv	nConv	nConv	Conv	Conv	Conv
CONV = 0.0005	nConv	nConv	nConv ^a	Conv	Conv	Conv
CONV = 0.001	nConv ^a	nConv ^a	nConv ^a	Conv	Conv	Conv
CONV = 0.005	Conv	Conv	Conv	Conv	Conv	Conv
Número de iteraciones = 10000						
Criterio de convergencia	Proporción de participantes del grupo AEMÉTODO removidos de la muestra					
	100%	80%	60%	40%	20%	0%
CONV = 0.00005	nConv	nConv	nConv	Conv	Conv	Conv
CONV = 0.0001	nConv	nConv	nConv ^a	Conv	Conv	Conv
CONV = 0.0005	nConv ^a	nConv ^a	nConv ^a	Conv	Conv	Conv
CONV = 0.001	nConv ^a	nConv ^a	nConv ^a	Conv	Conv	Conv
CONV = 0.005	Conv	Conv	Conv	Conv	Conv	Conv
Número de iteraciones = 100000						
Criterio de convergencia	Proporción de participantes del grupo AEMÉTODO removidos de la muestra					
	100%	80%	60%	40%	20%	0%
CONV = 0.00005	nConv	nConv	nConv	Conv	Conv	Conv
CONV = 0.0001	nConv	nConv	nConv ^a	Conv	Conv	Conv
CONV = 0.0005	nConv ^a	nConv ^a	nConv ^a	Conv	Conv	Conv
CONV = 0.001	nConv ^a	nConv ^a	nConv ^a	Conv	Conv	Conv
CONV = 0.005	Conv	Conv	Conv	Conv	Conv	Conv

Nota. nConv =Modelo no convergente; Conv =Modelo convergente.

^a La matriz de covarianza de la variable latente fue definitiva no positiva.

Creemos que los patrones de respuesta identificados en el subgrupo AEMÉTODO contribuyen a la varianza multidimensional que es capitalizada por el modelo bifactor en las condiciones MC y MC-R. Luego de remover este subgrupo (condición MC-M), el conjunto de datos adquiriría una estructura preferentemente unidimensional. En este escenario, la estimación del modelo Bifactor-FID bajo la condición representa una instancia de sobrefactorización (Barendse, Oort, & Timmerman, 2015; Garrido, Abad, & Ponsoda, 2016), forzando a los datos unidimensionales a ajustar a un modelo con tres factores ortogonales.

Para explorar esta posibilidad, utilizamos análisis paralelo basado en análisis de componentes principales (Horn, 1965; Tabla S6) para explorar el número hipotético de

componentes subyacentes (factores) en las condiciones MC-R y MC-M (n=6177), y el perfil AEMÉTODO (n=585). Utilizamos *Unweighted Least Squares* (ULS) como método de estimación y oblimin directo como método de rotación para lograr la simplicidad factorial y estimar las cargas factoriales (Lorenzo-Seva, 2000; Tabla S7). Todos los análisis se realizaron con el software FACTOR v.9.2 (Lorenzo-Seva & Ferrando, 2006).

En la condición MC-R, los resultados del análisis paralelo sugieren retener dos factores. Este hallazgo es consistente con aquellos provenientes de hallazgos previos utilizando métodos factoriales exploratorios (Huang & Dong, 2012). Los ítems positivamente redactados tuvieron todas cargas factoriales sobre .45 en el primer factor y cargas negligibles en el segundo factor, mientras que este patrón se invirtió para los ítems redactados negativamente (ver Tabla S7). La correlación entre ambos factores fue también alta ($r = 0,597$). Para la condición MC-M, al contrario de la condición MC-R, el análisis sugirió una estructura unidimensional. Las cargas factoriales fueron todas sobre .59.

Para el subgrupo AEMÉTODO, el análisis paralelo sugiere una solución multidimensional (modelo de tres factores) compuesto por dos factores separados para los ítems positivos y un factor adicional para cuatro de los cinco ítems negativos. El ítem i10 no arrojó una alta carga en ninguno de estos factores. Este resultado es consistente con la noción de que el patrón de respuestas exhibida por este subgrupo contribuye a la varianza multidimensional, independiente del factor general, que podría ser capturado por el modelo bifactor.

Referencias

- Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling*, 22, 87-101. <http://dx.doi.org/10.1080/10705511.2014.934850>
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, 21(1), 93-111. <https://doi.org/10.1037/met0000064>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. <https://doi.org/10.1007/BF02289447>
- Huang, C., & Dong, N. (2012). Factor structures of the Rosenberg Self-Esteem Scale: A meta-analysis of pattern matrices. *European Journal of Psychological Assessment*, 28(2), 132-138. <https://doi.org/10.1027/1015-5759/a000101>
- Lorenzo-Seva, U. (2000). The weighted oblimin rotation. *Psychometrika*, 65(3), 301-318. <https://doi.org/10.1007/BF02296148>
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavioral Research Methods, Instruments and Computers*, 38(1), 88-91. <https://doi.org/10.3758/BF03192753>

Tabla Suplementaria S6. Valores propios observados y simulados extraídos del análisis paralelo (Horn, 1965).

Factores	MC-R (n =6177)			MC-M (n =6177)			AEMÉTODO (n =585)		
	Datos	Media	Per95	Datos	Media	Per95	Datos	Media	Per95
1	5.171*	1.059	1.074	5.473*	1.062	1.079	2.024*	1.209	1.268
2	1.224*	1.042	1.054	0.948	1.045	1.057	1.736*	1.444	1.185
3	0.701	1.029	1.039	0.661	1.030	1.040	1.525*	1.097	1.131
4	0.603	1.017	1.025	0.612	1.018	1.027	1.015	1.055	1.087
5	0.547	1.006	1.013	0.527	1.006	1.015	0.844	1.017	1.045
6	0.496	0.994	1.003	0.513	0.994	1.002	0.720	0.977	1.008
7	0.459	0.983	0.992	0.472	0.982	0.991	0.679	0.939	0.967
8	0.297	0.971	0.980	0.295	0.969	0.978	0.590	0.890	0.931
9	0.271	0.958	0.968	0.280	0.956	0.965	0.493	0.857	0.888
10	0.230	0.941	0.955	0.220	0.938	0.951	0.374	0.806	0.847

Nota. * Factores cuyos valores propios o porcentaje de varianza excede los valores aleatorios.

Tabla Suplementaria S7. Matriz de cargas rotadas y correlación entre factores.

Items	MC-R		MC-M	AEMÉTODO		
	Factor 1	Factor 2	Factor 1	Factor 1	Factor 2	Factor 3
Item1	.747	.052	.737	.586	.005	.078
Item2	.877	.085	.711	.727	-.022	.016
Item3*	.129	.643	.699	.093	.099	.491
Item4	.626	.012	.595	.453	.088	-.166
Item5*	.214	.457	.608	.178	-.075	.505
Item6	.490	.372	.798	.021	.875	-.030
Item7	.474	.380	.791	-.054	.654	.101
Item8*	.047	.658	.619	-.077	.098	.516
Item9*	-.080	.850	.684	-.108	-.034	.469
Item10*	-.020	.829	.785	-.191	-.209	.213
Factor1	-	.597	-	-	.081	.091
Factor2	-	-	-	-	-	.131
Factor3	-	-	-	-	-	-

Nota. Los ítems inversos están marcados con un asterisco. Las cargas factoriales mayores a 0.40 aparecen en negrita.

3.3. MANUSCRITO 3:

La naturaleza efímera del efecto de redacción

Fernando P. Ponce¹, David Torres Iribarra¹, Alvaro Vergés¹, & Víctor B. Arias²

¹ Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile.

² Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Facultad de Psicología, Universidad de Salamanca

Resumen

Este artículo examina el análisis e interpretación de los efectos de redacción asociados al uso de ítems directos e inversos en instrumentos de medición. Investigaciones previas han sugerido, utilizando modelos bifactor, una naturaleza sustantiva del efecto de redacción. Con el propósito de superar debilidades recientemente señaladas del enfoque de modelamiento bifactor, el presente estudio emplea modelamiento de mixturas para examinar la naturaleza subyacente del efecto de redacción.

En los Estudios 1 y 2, examinamos la presencia de un subgrupo de participantes que exhibe efectos de redacción y evaluamos su impacto sobre la dimensionalidad de las escalas de autoestima (RSES) y optimismo disposicional (LOT-R), confirmando la ubicuidad del efecto de redacción en escalas compuestas por ítems directos e inversos. Con los datos de ambas escalas (n=5953), se encontró que, en presencia de una asociación significativa entre los factores de redacción (Estudio 3a), se identificó una baja proporción de participantes que exhibieron respuestas asimétricas en ambas escalas simultáneamente (Estudio 3b). Del mismo modo, utilizando tres olas de los datos longitudinales del panel LISS (n=3712), se examinó la estabilidad temporal e invarianza longitudinal de este efecto (Estudio 4a). De nuevo, pese a encontrar estabilidad temporal del efecto, hubo una baja proporción de participantes consistentemente identificados con respuestas asimétricas a lo largo del tiempo (Estudio 4b), reflejado en parámetros de transición inferiores en comparación con los restantes patrones de perfiles de interés. Estos hallazgos respaldan la noción de una naturaleza efímera subyacente al efecto de redacción.

Palabras clave: Efectos de redacción, efectos de método, modelo bifactor, modelamiento de mixturas.

Introducción

La evaluación de constructos psicológicos es uno de los aspectos que más estudio y atención ha concitado dentro de los investigadores durante el desarrollo de la disciplina, con un particular interés sobre aquellos atributos que permiten dar cuenta de las diferencias individuales y que muestran cierta estabilidad en el tiempo (Cronbach, 1956; Sackett et al., 2017; Thorndike, 1918). Es decir, aquellas variables que, junto con permitir examinar la variabilidad interindividual, representan rasgos o atributos relativamente estables para hacer descripciones y predicciones sobre la conducta del individuo (Geukes et al., 2017).

En términos de estrategias metodológicas para evaluar este tipo de atributos, y con excepción de la inteligencia, la escala de auto-reporte ha predominado como técnica de recolección de datos por sobre otro tipo de registros (Baumeister et al., 2007; Haefffel & Howard, 2010). Para la construcción de este tipo de dispositivos, diversos autores han sugerido incorporar ítems directos e inversos para medir el mismo constructo con el propósito de contrarrestar determinados sesgos de respuesta (*e.g.*, aquiescencia; Nunnally, 1978; Paulhus, 1991), actuando como “obstáculos cognitivos” que promueven que los encuestados se involucren con un procesamiento cognitivo más controlado al momento de responder (Podsakoff et al., 2003) y evitan una tendencia a responder estereotipadamente (Likert, 1932). Adicionalmente, la idea de incorporar ítems inversos descansa en el supuesto de que un atributo puede ser conceptualizado como un conjunto de observables (Cureton, 1951, citado en Markus & Borsboom, 2013), donde tanto ítems directos como inversos son indicadores psicométricamente equivalentes del constructo de interés y cuya única diferencia reside en la dirección de la afirmación.

Sin embargo, diversas investigaciones han constatado que existen diferencias entre las puntuaciones de los ítems directos y los inversos recodificados (también conocido como efecto de redacción). Este efecto ha sido descrito como una tendencia a responder diferencialmente a ambos subsets, produciendo alteraciones ubicuas en la calidad de los datos que pueden inducir: (a) relaciones espurias entre variables que de otro modo no estarían correlacionadas (Huang, Liu, & Bowling, 2015), (b) incremento/decremento de

la consistencia interna (Roszkowski & Soven, 2010; Wood et al., 2017), (c) desajuste en modelos unidimensionales o la emergencia de factores de método asociados a ítems directos y/o inversos (Greenberger et al., 2003; Kam & Meyer, 2015; Woods, 2006), (d) alteraciones en la capacidad discriminativa (Rodebaugh et al., 2004; Wu et al., 2013), o (e) un sesgo sistemático que socava la efectividad de manipulaciones experimentales (Maniaci & Rogge, 2014), por mencionar algunos ejemplos. En este sentido, la interpretación y el control estadístico de dichos elementos sistemáticos implican desafíos tanto conceptuales como analíticos para la calidad de la evaluación con medidas de auto-reporte.

Efectos de redacción y su naturaleza sustantiva versus efímera

Los efectos de redacción, que emergen debido al uso de ítems inversos, representan varianza sistemática atribuible al método de medición, independiente del constructo teórico que se desea medir (Podsakoff et al., 2003; Podsakoff, MacKenzie, & Podsakoff, 2012). Esta definición es fundamentalmente empírica y no conlleva una atribución causal de la naturaleza subyacente de este sesgo (Arias & Arias, 2017), por lo que existe un debate sobre el significado de los efectos de redacción.

Algunos investigadores han considerado la varianza asociada a la dirección de los ítems como varianza sistemática de error que debe ser incorporada en el modelo de medición para minimizar su efecto sobre la varianza asociada al rasgo (*e.g.*, Cronbach, 1946; Greenberger et al., 2003; Marsh, 1996; Weijters et al., 2013). Weijters et al. (2013) propusieron que las propiedades del instrumento (*e.g.*, dirección del primer ítem, distribución de los ítems inversos) en interacción con mecanismos cognitivos como el sesgo confirmatorio permiten elicitar el efecto de redacción.

Por su parte, otros autores han propuesto que estos efectos pueden considerarse como un estilo de respuesta estable, asociado a características individuales específicas y duraderas (*i.e.*, como un rasgo de personalidad; DiStefano & Motl, 2006; Marsh et al., 2010; Quilty et al., 2006). Siguiendo la definición de un estilo de respuesta propuesta en la literatura (Bentler et al., 1971; Jackson & Messick, 1958; Tomás & Oliver, 1999), estos autores han reportado evidencia de estabilidad del efecto de redacción entre instrumentos

(DiStefano & Motl, 2006; Horan et al., 2003), a lo largo del tiempo (Gana et al., 2013; Marsh et al., 2010; Motl & DiStefano, 2002) y su relación con otros atributos criterio tales como rasgos de personalidad (Arias et al., 2018; Donnellan, Ackerman, & Brecheen, 2016; Michaelides, Koutsogiorgi, & Panayiotou, 2016), habilidades verbales (Dunbar et al., 2000; Gnambs & Schroeders, 2020; Marsh, 1996; Weems et al., 2006) o tendencias afectivas y motivacionales (Alessandri et al., 2015; Arias & Arias, 2017; DiStefano & Motl, 2009; Lindwall et al., 2012; Quilty et al., 2006; Tomás et al., 2013).

Si bien existe abundante evidencia que respalda la noción del efecto de redacción como un estilo de respuesta, la mayoría corresponde a la relación del factor de redacción con un criterio. Marsh et al. (2010) señalaron que la distinción entre efectos de redacción como artefactos metodológicos efímeros o estilos de respuesta estables “no puede ser adecuadamente examinada con datos transversales” (p.371). Coincidimos con estos autores, aunque consideramos que tanto la estabilidad temporal como también la consistencia del efecto de redacción entre instrumentos son las piezas de evidencia más relevantes para demostrar la naturaleza sustantiva de estos efectos. Desafortunadamente, pocos estudios han reportado ambos tipos de evidencia y, en algunos casos, la evidencia reportada es contradictoria (Kam, 2018a).

Toda la evidencia descrita anteriormente, ha sido examinada utilizando el modelo bifactor como estrategia analítica para separar la varianza asociada al rasgo (factor de rasgo) de aquella asociada a la dirección de los ítems (factor de redacción). A partir de este tipo de modelos, se emplea el factor de redacción para examinar su relación estructural con otras variables latentes criterio. El uso del modelo bifactor como estrategia analítica es controversial debido a que este tipo de modelos tiende a acomodar patrones de respuestas implausibles (Reise et al., 2016), permitiendo incluso ajustar set de datos que no representan una estructura bifactor (Bonifay & Cai, 2017; Greene, et al., 2019; Morgan et al., 2015; Murray & Johnson, 2013) o contienen heterogeneidad latente (Raykov et al., 2019). Esta tendencia al sobreajuste por parte del modelo bifactor se debe a su forma funcional y complejidad (Bonifay & Cai, 2017), dando como resultado excelentes índices de ajuste que han sido interpretados como la estructura subyacente para

escalas balanceadas. El mayor riesgo de esto es que los factores específicos de redacción no tengan una adecuada interpretación dado que pueden ser estimados a partir de respuestas improbables e implausibles que introducen multidimensionalidad capitalizada por los factores específicos (Ponce et al., 2021).

Aproximaciones basadas en heterogeneidad poblacional

Las críticas al uso e interpretación del modelo bifactor en el contexto de los efectos de redacción han conducido a otros investigadores a proponer estrategias analíticas alternativas para examinar los efectos de redacción en escalas balanceadas. Estas estrategias apuntan a la identificación de subgrupos latentes de participantes que no se ajustan a una estructura factorial en particular (*e.g.*, un-factor, dos-factores, bifactor). Por ejemplo, Reise et al. (2016), utilizando mínimos cuadrados no ponderados (IRLS; Yuan & Bentler, 2000), encontraron que 86% de los participantes exhibieron un patrón de respuestas consistente con el modelo unidimensional para la escala de autoestima de Rosenberg (RSES, Rosenberg, 1965), concluyendo que la selección del modelo bifactor como modelo óptimo para la muestra completa descansa en una baja proporción de individuos con patrones de respuesta consistente con dicha estructura (3%) o directamente no modelables (11%). Por su parte, Kam (2018b; 2021), utilizando modelamiento de factores mezclados (FMM; Lubke & Muthén, 2005) para las escalas RSES y de optimismo disposicional (LOT-R; Scheier et al., 1994), identificó dos subgrupos latentes derivados de un modelo de dos-factores correlacionados, con un grupo de participantes que exhibieron mayor variabilidad en sus respuestas a los ítems inversos. Luego de excluir este grupo específico de participantes del análisis, encontraron una solución esencialmente unidimensional para ambos atributos.

Ponce et al. (2021) notaron que con ambas técnicas, una vez que son detectados los casos que contribuyen a la heterogeneidad, no hay una forma clara de caracterizarlos sustantivamente o reportar los resultados entre los grupos de una manera comparable, debido a la ausencia de parámetros comunes que permitan equiparar las escalas en las que se expresaron los resultados de las clases no-invariantes resultantes. Estos autores propusieron el análisis de perfiles latentes como técnica más simple para modelar

directamente la heterogeneidad sin asumir ningún modelo factorial en particular, permitiendo diferenciar grupos de individuos con patrones de respuestas sensibles a los efectos de redacción, y produciendo una taxonomía de individuos sustantivamente interpretable. Para ilustrar su punto, identificaron y removieron el subgrupo de participantes que exhibió un patrón diferencial de respuestas entre ítems directos e inversos y encontraron, al comparar con una condición muestral control, una contribución sustantiva al deterioro del ajuste estadístico del modelo unidimensional de la RSES consistente con el impacto del efecto de redacción reportado en la literatura. Más importante aún, señalaron que esta estrategia analítica permitiría examinar las definiciones y recomendaciones propuestas en la literatura para examinar el efecto de redacción como un estilo de respuesta (para una descripción más detallada, ver Ponce et al., 2021).

Pese a las posibilidades que ofrecen estas alternativas analíticas, hasta donde sabemos ningún estudio ha examinado la naturaleza sustantiva o efímera del efecto de redacción, con excepción de un trabajo no publicado por Bandalos et al. (2015, citado por Reise et al., 2016) quienes identificaron, en dos muestras independientes, una solución de tres clases con la RSES, con una clase (entre 12% y 17% de las muestras) que exhibió respuestas diferenciales a los ítems inversos. Junto con ello, los autores reportaron que la pertenencia a esta clase se asoció con bajos puntajes en escrupulosidad y habilidad verbal (en la primera y segunda muestra, respectivamente). Estos resultados, basados en la relación entre el efecto de redacción y un atributo, sugieren una naturaleza tipo-rasgo del efecto de redacción.

El presente estudio

El presente trabajo tuvo como propósito examinar la aparente naturaleza tipo-rasgo del efecto de redacción, utilizando la estrategia analítica propuesta por Ponce et al. (2021). Como se señaló anteriormente, la consistencia del efecto de redacción entre escalas y la estabilidad temporal son las piezas de evidencia más relevantes para demostrar la naturaleza sustantiva de estos efectos, por lo que en este estudio nos enfocaremos en examinar ambas.

En los Estudios 1 y 2, identificamos un subgrupo de participantes que exhibieron un patrón de respuesta diferencial entre ítems directos e inversos (un subgrupo sesgado) de las escalas de autoestima de Rosenberg (RSES; Rosenberg, 1965) y optimismo disposicional (LOT-R; Scheier et al., 1994). De acuerdo a lo reportado por Reise et al. (2016) y Ponce et al. (2021), esperamos que la proporción de individuos identificados con este tipo de respuestas fluctúe entre 5% y 15%. Una vez identificado este subgrupo de participantes en ambas escalas, estimamos los modelos de medida utilizando la estrategia de remoción descrita por Ponce et al. (2021) y examinamos el impacto del subgrupo identificado sobre el ajuste estadístico del modelo unidimensional y bifactor. A partir de los resultados reportados por estos autores esperamos encontrar un incremento de la unidimensionalidad, reflejado a través de la mejora en el ajuste estadístico del modelo unidimensional y un incremento en la proporción de varianza capturada por el factor asociado a rasgo en el modelo bifactor. Junto con ello se espera, si el efecto de redacción es estable entre instrumentos, que el efecto de remoción sobre la unidimensionalidad sea generalizado entre escalas.

En los Estudios 3a y 3b examinamos la consistencia del efecto de redacción entre escalas. En primer lugar, estimamos la relación estructural entre los factores de redacción de la RSES y LOT-R, con el propósito de replicar conceptualmente la evidencia de asociación previamente reportada (Horan et al., 2003; DiStefano & Motl, 2006). Posteriormente, examinamos la solución conjunta de clases para ambas escalas con el propósito de identificar los individuos con patrones de respuestas asimétricas en cada escala y el patrón de perfiles consistentemente asimétricos en la RSES y LOT-R. A partir de los subgrupos identificados, estimamos la relación de los factores de método utilizando la estrategia de remoción muestral. Si el efecto de redacción es estable entre instrumentos, se espera encontrar una alta proporción de individuos con patrones de respuesta consistentemente asimétricas respecto del número de individuos identificados en cada escala y que la remoción de estos participantes tendrá un impacto significativo sobre la relación entre los factores de redacción.

Finalmente, en los Estudios 4a y 4b examinamos la evidencia de estabilidad temporal del efecto de redacción en la RSES en tres olas de medición. En primer lugar, evaluamos la invarianza longitudinal del modelo bifactor y la relación estructural entre los factores de redacción a lo largo del tiempo, con el propósito de replicar hallazgos previos (Gana et al., 2013; Marsh et al., 2010; Motl & DiStefano, 2002). Posteriormente, examinamos las trayectorias de los subgrupos identificados con respuestas simétricas y asimétricas, comparamos los parámetros de transición entre estos subgrupos y examinamos la proporción de individuos con patrones de respuesta consistentemente asimétricos a lo largo del tiempo. A partir de los patrones de perfiles identificados, estimamos la relación longitudinal de los factores de redacción utilizando la estrategia de remoción muestral. Si el efecto de redacción es estable en el tiempo, se espera encontrar una probabilidad de transición entre los subgrupos con respuestas asimétricas similar a la observable entre los subgrupos con respuestas simétricas, y una alta proporción de individuos con patrones de respuesta consistentemente asimétricas respecto del número de individuos identificados en cada ola y que la remoción de este subgrupo tenga un impacto significativo sobre la relación entre los factores de redacción.

Estudio 1. Identificación de respuestas asimétricas en distintas escalas

Datos y participantes

Utilizamos los datos de la quinta ola del panel LISS (Longitudinal Internet Studies in the Social Sciences) del CentERdata de la Universidad de Tilburg. Estos datos son distintos a los utilizados en Ponce et al. (2021; donde se emplearon datos de la primera ola). Los participantes (n=5953) fueron seleccionados aleatoriamente de los registros municipales y posteriormente invitados a completar un cuestionario online (una descripción detallada del estudio puede encontrarse en Scherpenzeel, 2011).

Considerando el tamaño de la muestra y para facilitar el análisis, descartamos 26 participantes (0,44%) con datos de respuesta perdidos. Este procedimiento dejó el tamaño de la muestra de análisis en 5927 participantes (53,47% mujeres; 0,07% no informó su sexo) cuyas edades oscilaron entre 16 y 92 años (Me=49,62; DE=17,43; 0,07% no informó su edad).

Medidas

Escala de Autoestima de Rosenberg (RSES; Rosenberg, 1965). La RSES es una escala de 10 ítems diseñada para evaluar el nivel de autoestima general, con 5 ítems directos y 5 ítems inversos. Los ítems fueron puntuados en una escala tipo-Likert de 7 puntos (1=Totalmente en desacuerdo; 7=Totalmente de acuerdo). Los ítems inversos fueron recodificados para que altas puntuaciones en la escala correspondan a alta autoestima.

Escala de Optimismo Disposicional (LOT-R; Scheier et al., 1994). La escala LOT-R se compone de 6 ítems que miden el grado de optimismo (3 ítems directos y 3 ítems inversos) y 4 ítems adicionales que sirvieron de relleno. Los ítems fueron puntuados en una escala tipo-Likert de 5 puntos (1=Muy en desacuerdo; 5=Muy de acuerdo). Los ítems inversos fueron recodificados para que altas puntuaciones en la escala reflejen mayor nivel de optimismo.

Plan de análisis

Siguiendo la estrategia de Ponce et al. (2021), utilizamos análisis de perfiles latentes (LPA) para identificar grupos caracterizados por distintos patrones de respuesta. Para evaluar la consistencia de las soluciones estimadas (Collins & Lanza, 2010), la muestra fue aleatoriamente dividida en dos submuestras (n=2963 y n=2964). Los análisis fueron llevados a cabo utilizando Mplus v.7.4 (Muthén & Muthén, 1998-2014).

Con el propósito de seleccionar la solución de perfiles que mejor se ajustara a los datos observados, se emplearon consideraciones teóricas y estadísticas (Muthén & Muthén, 2000). Los modelos examinados fueron comparados utilizando los criterios de información de Akaike (AIC; Akaike, 1974) y bayesiano (BIC; Schwarz, 1978), además de los test de razón de verosimilitud de Lo-Mendell-Rubin y Vuong-Lo-Mendell-Rubin (LMR y VLMR, respectivamente; Lo, Mendell, & Rubin, 2001). Valores inferiores en AIC y BIC indican un mejor ajuste del modelo, y un valor-p >0,05 de LMR y VLMR indican que el modelo con k clases no tiene un mejor ajuste a los datos que el modelo con k-1 clases. Junto con estos criterios, consideramos también como criterio de selección la

interpretación parsimoniosa de la solución, basado en una prevalencia mínima del 5% de cada clase identificada (Osborne & Sibley, 2017).

Dada una solución óptima identificada, examinamos la precisión en cada muestra utilizando el estadístico de entropía y la probabilidad media de clasificación del perfil latente (PMC). Los valores de entropía varían entre 0 y 1, y altos valores indican una mayor separación entre las clases (Celeux & Soromenho, 1996; Lubke & Muthén, 2007). Los valores de PMC varían entre 0 y 1, y altos valores indican niveles bajos de clasificaciones erróneas (Vermunt & Magidson, 2002; Neely-Barnes, 2010).

Resultados

Estadísticos Descriptivos. La Tabla 1 muestra los estadísticos descriptivos para las dos muestras utilizadas en este estudio. Como se puede esperar, dada la estrategia de aleatorización empleada para generar ambas muestras, no hubo diferencias estadísticamente significativas entre los valores de los ítems entre las muestras ($|t_s|(5925) < 1,601$; $p_s > 0,109$).

[Insertar Tabla 1 aproximadamente aquí]

Perfiles latentes. La Tabla 2 muestra los índices de ajuste para los modelos de dos a cinco perfiles para la escala RSES y de dos a seis perfiles para la escala LOT-R en cada muestra. Los índices AIC y BIC no permitieron seleccionar una solución en particular, mientras que los valores de VLMR y LMR fueron inconsistentes en señalar una solución entre ambas muestras. Por su parte, el criterio de parsimonia arrojó consistentemente en ambas muestras una solución de cuatro perfiles para la escala de autoestima y cinco para la de optimismo disposicional. Esta solución está parcialmente respaldada por los criterios de VLMR y LMR. Por lo tanto, basados en los criterios estadísticos de VLMR/LMR y el criterio de parsimonia, seleccionamos la solución de cuatro perfiles para la RSES y cinco perfiles para la LOT-R.

[Insertar Tabla 2 aproximadamente aquí]

En relación al grado de certidumbre respecto a la asignación de los participantes a los perfiles latentes, los valores de entropía (Tabla 2) reflejan un aceptable nivel de clasificación de los participantes (Clark, 2010), con valores superiores al 90% de

asignaciones correctas (Lubke & Muthén, 2007). Consistentemente, los valores de probabilidad media de clasificación al perfil fueron iguales o superiores a 0,884 para la RSES y 0,717 para la LOT-R (Tabla 3).

[Insertar Tabla 3 aproximadamente aquí]

La Figura 1 muestra las medias estimadas de los ítems directos e inversos de ambas escalas en ambas muestras. En el caso de la RSES, ambas muestras presentan similares patrones de respuesta con similares prevalencias (Tabla 3), donde tres perfiles mostraron un patrón ordenado que puede ser interpretado como de alta autoestima (AE_A), autoestima moderada (AE_M), y un perfil de autoestima baja (AE_B). El cuarto y último perfil, sin embargo, muestra un patrón asimétrico de respuestas en función de la dirección de los ítems, cercano al de alta autoestima para los ítems directos y al de baja autoestima con ítems inversos (AE_{MET}). Por su parte, en el caso de la LOT-R, ambas muestras exhibieron patrones de respuesta similares entre sí con prevalencias similares (Tabla 3), donde cuatro perfiles muestran un patrón de ordenamiento que puede interpretarse como de alto optimismo (OP_A), de optimismo medio-alto (OP_{MA}), un perfil de optimismo moderado (OP_M), y uno de bajo optimismo (OP_B). Tal como con la RSES, el último y quinto perfil arrojó un patrón de respuestas asimétrico de acuerdo a la dirección de los ítems (OP_{MET}).

[Insertar Figura 1 aproximadamente aquí]

El patrón de respuestas observado de los perfiles AE_{MET} y OP_{MET} fue claramente distinto del observado en los restantes perfiles los cuales, más allá de ciertas diferencias entre ítems, son consistentes con la estructura unidimensional originalmente propuesta para ambos atributos. Por su parte, el patrón asimétrico de respuestas detectado en ambas escalas sugiere que los participantes identificados responderían diferencialmente a ítems inversos, expresando una opinión menos favorable en comparación a la expresada con ítems directos, consistente con lo propuesto por otros autores (Kamoen, Holleman, Mak et al., 2011; Kamoen, Holleman, Nouwen et al., 2011; Kamoen et al., 2013; Quilty et al., 2006; Schmitt & Allik, 2005).

Esta asimetría observada en el patrón de respuestas produce un detrimento en la estructura unidimensional teóricamente propuesta para ambas medidas, induciendo la

presencia de variables latentes adicionales en un modelo factorial en orden de dar cuenta del patrón de respuesta diferencial observado en subconjuntos específicos de ítems. El siguiente estudio examinó el impacto de los subgrupos identificados en los modelos de medida de ambas escalas.

Estudio 2. Impacto específico de las respuestas asimétricas sobre modelos de medida

Los resultados reportados en el Estudio 1 son consistentes con los descritos por Ponce et al (2021) con la escala de autoestima y permiten extender la detección del perfil con respuestas asimétricas a la escala de optimismo disposicional. Estos autores reportaron que los participantes con patrones asimétricos de respuesta a ítems directos e inversos contribuyen al desajuste obtenido del modelo de un factor de la RSES (Ponce et al., 2021; Estudio 2). De hecho, al remover este subgrupo identificado, el modelo de un factor alcanzó criterios de ajuste aceptables y, dependiendo del número de factores específicos, el modelo bifactor incrementó su grado de unidimensionalidad (expresado en los valores de varianza común explicada) o fue incapaz de alcanzar convergencia debido a la sobre factorización de una matriz de datos esencialmente unidimensional.

Considerando la noción propuesta de que el sesgo de método es consistente entre escalas balanceadas (DiStefano & Motl, 2006; Horan, DiStefano, & Motl, 2003), sería esperable que el efecto de remoción, observado sobre la estructura factorial de una escala, sea transferible a otra escala balanceada. Con el propósito de examinar esta posibilidad, el Estudio 2 examinó el potencial impacto generalizado del subgrupo identificado en una escala sobre su propia estructura factorial (buscando replicar los hallazgos reportados por Ponce et al., 2021) y sobre la estructura factorial de una escala balanceada complementaria.

Datos y participantes

Los datos empleados en este estudio fueron los mismos que los utilizados en el Estudio 1.

Plan de análisis

Para examinar la estructura factorial subyacente a ambas escalas, comparamos el ajuste estadístico de cuatro modelos utilizando Análisis Factorial Confirmatorio (Figura 2). El primer modelo (M1) no tuvo en cuenta los efectos de redacción, especificando un solo factor asociado al rasgo. Los restantes modelos consistieron en tres modelos bifactor, en los cuales, junto con el factor de rasgo, fueron especificados factores de redacción ortogonales para capturar la varianza de los ítems redactados en la misma dirección: un factor de método para los ítems inversos (Bifactor-FI; M2), para los ítems directos (Bifactor-FD; M3), o para ambos subconjuntos de ítems (Bifactor-FID; M4)⁹. Optamos por examinar los efectos de redacción utilizando estos modelos dadas las ventajas comparativas que tienen por sobre los modelos con unicidades correlacionadas (UCs) para examinar la relación de los efectos de redacción entre distintas escalas o con datos longitudinales (Marsh et al., 2010). Todos los modelos fueron estimados utilizando Máxima Verosimilitud con una estimación robusta de los errores estándar (MLR) como método de estimación.

[Insertar Figura 2 aproximadamente aquí]

Con el propósito de examinar el impacto del subgrupo identificado, los modelos examinados fueron estimados en ambas escalas bajo las siguientes cinco condiciones: (a) Muestra completa (Condición MC; $n=5927$); (b) Muestra completa menos el subgrupo identificado con respuestas asimétricas en la RSES (AE_{MET} , $n=694$; Condición MC- M_{AE} , $n=5233$), (c) Muestra completa menos un número de participantes, aleatoriamente seleccionados, equivalente al tamaño muestral de la condición MC- M_{AE} (Condición MC- R_{AE}), (d) Muestra completa menos el subgrupo identificado con respuestas asimétricas en la LOT-R (OP_{MET} , $n=595$; Condición MC- M_{OP} , $n=5332$), y (e) Muestra completa menos

⁹ La ortogonalidad entre los factores de método sigue el razonamiento propuesto por otros autores de que ambos factores representan fuentes independientes de varianza asociada a la dirección de los ítems (Alessandri et al., 2015; Quilty et al., 2006). Estos autores reportaron una asociación no significativa entre ambos factores específicos ($r=0,05$; $p=0,21$; Alessandri et al., 2015), así como un detrimento no significativo en el ajuste una vez esta correlación fue fijada a cero ($SB\Delta\chi^2_{(1)}=0,74$, $p=0,39$; Alessandri et al., 2015). Una inspección preliminar con nuestros datos confirmó ambos hallazgos ($r=-0,09$; $p=0,34$; $SB\Delta\chi^2_{(1)}=0,68$, $p=0,41$).

un número de participantes, aleatoriamente seleccionados, equivalente al tamaño muestra de la condición MC-M_{OP} (Condición MC-R_{OP}).

La bondad de ajuste de los modelos estimados fue evaluada utilizando el índice de ajuste comparativo (CFI), el índice de ajuste de Tucker-Lewis (TLI), el error cuadrático medio de aproximación (RMSEA) y la raíz cuadrada media estandarizada residual (SRMR). Para los índices CFI y TLI, valores estimados por sobre 0,90 y 0,95 indican un nivel de ajuste aceptable y bueno, respectivamente (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). Para los índices RMSEA y SRMR, valores iguales o inferiores a 0,05 y 0,08 se consideran buenos y aceptables, respectivamente (Browne & Cudeck, 1992; Hu & Bentler, 1999). Los análisis fueron llevados a cabo utilizando Mplus v.7.4 (Muthén & Muthén, 1998-2014).

A partir de los modelos bifactor estimados, calculamos la proporción de varianza explicada en cada condición (ECV; Brunner, Nagy, & Wilhelm, 2012), tanto para el factor general del rasgo como para los factores específicos asociados al subconjunto de ítems directos y/o inversos.

Resultados

La Tabla 4 muestra los índices de ajuste de los modelos estimados junto con la proporción de varianza capturada por los factores especificados en cada modelo bifactor. Con respecto a los modelos estimados con la muestra completa de participantes (Condición MC), los modelos Bifactor FID ($\chi^2_{(23)}=246,291$; $p<0,001$; CFI=0,987; TLI=0,974; RMSEA=0,040 [IC 90% =0,036; 0,045]; SRMR=0,015), y FD ($\chi^2_{(6)}=53,307$; $p<0,001$; CFI=0,990; TLI=0,974; RMSEA=0,036 [IC 90% =0,028; 0,046]; SRMR=0,014) fueron los que arrojaron el mejor ajuste y una solución esencialmente unidimensional (ECV \geq .70; Rodríguez, Reise, & Haviland, 2016a) para las escalas de autoestima y optimismo, respectivamente. Estos resultados son consistentes con las últimas rendiciones propuestas para ambas escalas (Alessandri et al., 2015; Gnamb, Scharl, & Schroeders, 2018). Por su parte, los modelos especificados con la muestra completa menos un número aleatorio de participantes (condiciones MC-R_{AE} y MC-R_{OP}) arrojaron índices de ajuste y de varianza explicada consistentes con los obtenidos con la

muestra completa (MC). En este sentido, ambas representan condiciones control comparables a las condiciones MC-M_{AE} y MC-M_{OP}.

Finalmente, los modelos especificados con la muestra completa menos los participantes identificados con respuestas asimétricas (condiciones MC-M_{AE} y MC-M_{OP}) evidenciaron un incremento en la unidimensionalidad de la solución en ambas escalas. Para la RSES, la condición MC-M_{AE} arrojó una mejora en el ajuste del modelo de un factor ($\chi^2_{(33)}=955,366$; $p<0,001$; CFI=0,939; TLI=0,917; RMSEA=0,073 [IC 90% =0,069; 0,077]; SRMR=0,040) en comparación al ajuste observado con la condición MC-R_{AE} ($\chi^2_{(33)}=1422,259$; $p<0,001$; CFI=0,905; TLI=0,871; RMSEA=0,090 [IC 90% =0,086; 0,094]; SRMR=0,062). Del mismo modo, la varianza explicada por el factor general de los modelos bifactor en la condición MC-M_{AE} aumentó respecto de MC-R_{AE} entre un 7,7% (Bifactor FD) y 8,5% (Bifactor FI). Estos efectos específicos son consistentes con los reportados por Ponce et al. (2021), incluyendo la falta de convergencia del modelo Bifactor FID en la condición MC-M_{AE}. Sin embargo, esta mejora en el ajuste y varianza explicada de la RSES no se evidenció en la condición MC-M_{OP} al compararla con MC-R_{OP}.

Esta mejora específica según con qué escala los individuos con respuestas asimétricas fueron identificados también fue encontrada con los modelos estimados para la LOT-R. La condición MC-M_{OP} arrojó una mejora en el ajuste del modelo de un factor ($\chi^2_{(9)}=275,060$; $p<0,001$; CFI=0,939; TLI=0,899; RMSEA=0,074 [IC 90% =0,067; 0,082]; SRMR=0,036) en comparación al ajuste observado con la condición MC-R_{OP} ($\chi^2_{(9)}=843,506$; $p<0,001$; CFI=0,795; TLI=0,659; RMSEA=0,132 [IC 90% =0,124; 0,139]; SRMR=0,069), así como una mejora en la proporción de varianza capturada en la condición MC-M_{OP} respecto de MC-R_{OP} que fluctuó entre un 11,7% (Bifactor FD) y 18,8% (Bifactor FI). Sin embargo, este efecto no se evidenció al comparar los valores de ECV de los modelos de medida del LOT-R en las condiciones MC-M_{AE} y MC-R_{AE}.

[Insertar Tabla 4 aproximadamente aquí]

Tomados en conjunto, estos resultados muestran el impacto de los participantes con respuestas asimétricas sobre los modelos de medida de escalas balanceadas. Estos

resultados son consistentes con los hallazgos reportados por Ponce et al. (2021) con la RSES y extienden sus hallazgos a la LOT-R. Sin embargo, aquí podemos apreciar que la mejora en el ajuste y el incremento de la unidimensionalidad en una escala es específica a la condición de remoción de los participantes detectados con respuestas asimétricas en dicha escala y no es generalizada entre instrumentos.

Este último hallazgo contradice, en principio, la noción propuesta del efecto de redacción como un estilo de respuesta a partir de la relación estructural reportada entre los factores de redacción de escalas balanceadas (DiStefano & Motl, 2006; Horan et al., 2003). El siguiente estudio examinará directamente la noción de la consistencia del sesgo en distintas escalas, como evidencia de un potencial estilo de respuesta.

Estudio 3a. Evidencia del sesgo de método en distintas escalas

Luego de establecer la presencia de efectos de redacción en escalas balanceadas que evaluaban distintos contenidos sustantivos, Horan et al (2003) reportaron que los factores asociados al sesgo de redacción en dichas escalas compartían varianza común, sugiriendo la hipótesis del estilo de respuesta como un rasgo de personalidad.

El Estudio 2 examinó parcialmente esta afirmación a través del efecto de remoción de respuestas asimétricas identificadas descrito por Ponce et al. (2021). Sin embargo, en lugar de encontrar un efecto generalizado entre escalas, los resultados sugieren un impacto específico en oposición a la idea de un estilo de respuesta que se despliega masivamente al momento de responder escalas balanceadas.

Para examinar la noción de un estilo de respuesta con más detalle, comenzaremos identificando la relación entre factores de método entre escalas, siguiendo la estrategia analítica empleada en investigaciones previas (ver, Horan et al., 2003; DiStefano & Motl, 2006).

Datos y participantes

Los datos empleados en este estudio fueron los mismos utilizados en los estudios anteriores.

Plan de análisis

A partir de los modelos estimados en el Estudio 2 para ambas escalas, se estimaron tres modelos estructurales (Figura 3). El primer modelo (M5) no tuvo en cuenta los efectos de redacción, especificando un solo factor asociado al rasgo en ambas escalas relacionados entre sí. Los restantes modelos consistieron en dos modelos bifactor (Bifactor FID para la RSES y Bifactor FD para la LOT-R, de acuerdo a los resultados del Estudio 2) especificando sólo la correlación entre los rasgos (M6) o permitiendo la correlación entre los factores de redacción (M7)¹⁰. Todos los modelos fueron estimados utilizando MLR como método de estimación.

[Insertar Figura 3 aproximadamente aquí]

Tal como en el estudio anterior, calculamos la proporción de varianza explicada del factor general y los factores específicos asociados al subconjunto de ítems directos y/o inversos de los modelos bifactor estimados (Brunner et al., 2012).

Resultados

La Tabla 5 muestra los índices de ajuste de los modelos estimados, la proporción de varianza capturada por los factores específicos asociados a los ítems directos e inversos, y la correlación entre los factores de rasgo (Modelos M5-M7) y entre los factores de redacción (Modelo M7). Con la muestra completa de participantes (Condición MC), el Modelo M5 exhibió un ajuste inferior al nivel óptimo, con la mayoría de los índices por debajo del nivel aceptable ($\chi^2_{(101)}=3313,305$; $p<0,001$; CFI=0,882; TLI=0,860; RMSEA=0,073 [IC 90% =0,071; 0,075]; SRMR=0,060). El Modelo M6, que contenía los factores de rasgo correlacionados y los factores de redacción sin correlacionar, ofreció una mejora en el ajuste alcanzando criterios aceptables ($\chi^2_{(88)}=1103,799$; $p<0,001$; CFI=0,963;

¹⁰ Cabe destacar que el Modelo M7 permite la correlación entre el factor FD de la escala de optimismo y el factor FI de la escala de autoestima. Esta asociación entre factores asociados a ítems positivos y negativos contradice la ortogonalidad descrita y reportada entre ambos tipos de factores al interior del modelo de la escala de Rosenberg. Nuestros análisis preliminares sugieren que esta correlación, además de ser significativa (vid., Tabla 5), contribuye a una mejora significativa en el ajuste del Modelo M7 ($S\Delta\chi^2_{(1)}=65,905$, $p<0,001$). A partir de la mejora significativa observada en el ajuste una vez liberado este parámetro, hemos optado por reportar nuestros hallazgos empleando esta especificación. Sin embargo, vale la pena mencionar que los hallazgos reportados en este estudio fueron consistentes empleando ambas especificaciones.

TLI=0,949; RMSEA=0,044 [IC 90% =0,042; 0,046]; SRMR=0,051), con una solución esencialmente unidimensional para las escalas de autoestima (ECV=0,756) y optimismo (ECV=0,726). El Modelo M7, con las variables de rasgo y redacción correlacionadas entre sí, fue el que arrojó el mejor ajuste entre los modelos propuestos ($\chi^2_{(86)}=1044,025$; $p<0,001$; CFI=0,965; TLI=0,951; RMSEA=0,043 [IC 90% =0,041; 0,046]; SRMR=0,043) y una solución esencialmente unidimensional en ambas escalas.

[Insertar Tabla 5 aproximadamente aquí]

Por su parte, las correlaciones presentadas en el Modelo M7 entre los factores de redacción reflejan una covarianza positiva entre los factores de la escala de autoestima y de optimismo (Tabla 5). Estas asociaciones siguen, como lo propusieron DiStefano y Motl (2006), la dirección de la asociación observada entre los factores asociados al rasgo. Estos resultados, empleando los mismos datos utilizados en los Estudios 1 y 2, son consistentes con la noción de consistencia de sesgo de redacción propuesta por Horan et al. (2003), basada en la relación significativa entre los factores de redacción de dos escalas balanceadas.

A partir de esto, sin embargo, emerge una aparente contradicción entre los resultados de este estudio, que proponen que el sesgo de método representa un estilo de respuesta estable entre escalas y los resultados del Estudio 2 que indican que la remoción de individuos con respuestas asimétricas impacta de manera específica en el ajuste estadístico y la unidimensionalidad en función de la escala balanceada con que fueron identificados. El siguiente estudio pretende explicar esta aparente contradicción.

Estudio 3b. Convergencia de las respuestas asimétricas en distintas escalas

Datos y participantes

Los datos empleados en este estudio fueron los mismos utilizados en los estudios anteriores.

Plan de análisis

A partir del número de perfiles estimados en el Estudio 1 para ambas escalas, se estimó un modelo con dos variables latentes categóricas correlacionadas utilizando Análisis Confirmatorio de Perfiles Latentes (Finch & Bronk, 2011). La precisión de la

solución, así como de las asignaciones a los patrones de perfiles latentes (*e.g.*, AE_{MET} - OP_{MET}) fueron examinadas utilizando el estadístico de entropía y la probabilidad media de clasificación al patrón de perfil latente (PMC). Los criterios de interpretación de los valores de PMC de un patrón de perfil latente son idénticos a los que se utilizaron para interpretar los valores de PMC para perfiles latentes (Vermunt & Magidson, 2002; Neely-Barnes, 2010).

Una vez identificados los perfiles AE_{MET} y OP_{MET} en ambas escalas, así como el patrón de perfiles AE_{MET} - OP_{MET} que agrupa a los participantes que coincidieron con ambos perfiles, estimamos los modelos M5-M7 siguiendo la estrategia de remoción descrita en el Estudio 2, examinando ocho condiciones muestrales a partir de la remoción de: (a) el subgrupo de participantes identificados con el perfil AE_{MET} (MC- M_{AE}) o con el perfil OP_{MET} (MC- M_{OP}), y un número equivalente de participantes aleatoriamente seleccionados (MC- R_{AE} y MC- R_{OP} , respectivamente), (b) aquellos participantes identificados con respuestas asimétricas en ambas escalas (*i.e.*, patrón de perfiles AE_{MET} - OP_{MET} ; MC- M_{2EA}) y un número idéntico de participantes, seleccionados aleatoriamente (MC- R_{2EA}), y (c) el subgrupo de participantes identificados con respuestas asimétricas en al menos una de las dos escalas (MC- M_{3OV}) y un número idéntico seleccionado aleatoriamente (MC- RM_{3OV}). Todos los modelos factoriales fueron estimados siguiendo las especificaciones descritas en el Estudio 3a.

Resultados

El modelo estimado arrojó un nivel de clasificación aceptable (Entropía=0,796; Clark, 2010; Lubke & Muthén, 2007). Tanto los patrones de respuestas exhibidos como las prevalencias observadas en cada perfil de ambas escalas fueron consistentes con los valores reportados en el Estudio 2. Con respecto a la precisión de las asignaciones a los patrones de perfiles latentes, los valores de PMC fueron altos para todos los patrones examinados (Tabla 6), siendo el patrón AE_{MET} - OP_M el que obtuvo el valor más bajo (PMC=0,682) y cuyo segundo patrón más probable en dicha clasificación arrojó un valor considerablemente inferior (AE_M - OP_M ; PMC=0,060).

[Insertar Tabla 6 aproximadamente aquí]

Con respecto a los patrones de perfiles observados entre las clases identificadas en ambas escalas, la Tabla 6 muestra la proporción de participantes por patrón de perfiles. Los perfiles que representan niveles de autoestima y optimismo exhibieron una distribución de frecuencia en los patrones de perfiles que confirman la relación propuesta entre ambos constructos (Brissette, Scheier, & Carver, 2002; Scheier et al., 1994). Por su parte, los participantes identificados con el patrón de perfiles AE_{MET} - OP_{MET} ($n=101$) representaron una baja proporción respecto de los participantes identificados con un perfil AE_{MET} y OP_{MET} (15,3% y 20,8%, respectivamente), sugiriendo una baja relación entre ambos perfiles asimétricos identificados en distintos constructos. Estos resultados son consistentes con la idea de una naturaleza efímera del sesgo de redacción y respaldan la evidencia del impacto específico sobre el ajuste, grado de unidimensionalidad y convergencia del modelo bifactor, encontrada en el Estudio 2.

En este punto emerge la inquietud acerca de cómo es posible obtener, utilizando el mismo set de datos, resultados que respaldan conclusiones opuestas. Para examinar esto, utilizamos la estrategia de remoción descrita en el plan de análisis y estimamos los modelos M5-M7 del Estudio 3a. La Tabla 5 muestra los índices de ajuste, la proporción de varianza explicada por los factores de ambas escalas y la correlación entre los factores de rasgo y redacción.

En términos de bondad de ajuste y varianza explicada, todos los modelos estimados bajo las condiciones de remoción de individuos identificados con respuestas asimétricas en al menos una de las dos escalas ($MC-M_{AE}$ y $MC-M_{OP}$) evidenciaron una mejora sustantiva en el ajuste estadístico en comparación al observado en sus respectivas condiciones control ($MC-R_{AE}$ y $MC-R_{OP}$), junto con una falta de convergencia de los modelos bifactor en la condición $MC-M_{AE}$ y $MC-M_{2EV}$. Del mismo modo, se observó un incremento específico de la varianza capturada por el factor de rasgo en la condición $MC-M_{OP}$. Más importante, las asociaciones entre los factores de redacción de ambas escalas observada en la condición $MC-R_{OP}$ (similar a la reportada en la condición MC en el Estudio 2a; Tabla 5), disminuyeron sustancialmente en la condición $MC-M_{OP}$, alcanzando ausencia de significancia estadística ($ps>0,575$).

Por su parte, en la condición MC-M_{2EA}, que descartó a los individuos identificados con respuestas asimétricas simultáneamente en ambas escalas, el efecto fue variable al arrojar tanto una disminución en la asociación entre FD_{OP} y FD_{AE} ($r=0,183$; $p<0,001$) como un incremento en la magnitud entre FD_{OP} y FI_{AE} ($r=0,181$; $p<0,001$) respecto de los valores observados en la condición MC-R_{2EA}. Esto significa que cuando los individuos que consistentemente respondieron de manera asimétrica a ítems directos e inversos fueron removidos, se observó un mínimo incremento generalizado de la varianza explicada por los factores de rasgo, una leve mejora en la bondad de ajuste y una variación en la magnitud de la relación de los factores de redacción que no explica la relación en su conjunto. Esto último sugiere que la correlación observada entre los factores de redacción representa algo más que lo que propusieron Horan et al. (2003) y DiStefano y Motl (2006).

Una posibilidad es que la asociación entre los factores de redacción esté capitalizando la distribución conjunta de las puntuaciones factoriales de individuos que, paradójicamente, no exhiben patrones de respuesta asimétricos en una de las dos escalas. En este caso, la distribución conjunta de las puntuaciones de los individuos que fueron identificados con respuestas asimétricas en la LOT-R y simétricas en la RSES. Esto es consistente con la noción propuesta por algunos autores acerca de la capacidad del modelo bifactor para acomodar cualquier patrón de respuesta (Bonifay & Cai, 2017; Reise et al., 2016). Para ilustrar esta posibilidad, examinamos la distribución conjunta de las puntuaciones factoriales que representaron la relación entre los factores de redacción asociados a ítems directos (FD_{OP} y FD_{AE}; Figura 4), estimada en el modelo M7 en la condición MC¹¹.

La Figura 4 muestra la dispersión de las puntuaciones factoriales de la asociación entre FD_{OP} y FD_{AE} en la condición MC. Los participantes con un patrón de perfiles AE_{MET}-OP_{MET} (puntos blancos), que exhibieron un patrón de respuestas asimétrico en ambas escalas, mostraron una distribución de puntuaciones factoriales que favorece la

¹¹ Optamos por examinar esta relación con estas condiciones muestrales debido a que fue la asociación más alta observada entre factores de redacción (entre $r=0,279$ en MC y $r=0,294$ en MC-M_{OP}) y dado que ningún modelo bifactor alcanzó criterios de convergencia en la condición MC-M_{AE}.

asociación observada entre ambos factores de redacción. Sin embargo, la asociación observada entre FD_{OP} y FD_{AE} en la condición MC- M_{2EA} , que descartó este grupo de individuos ($r=0,183$; $p<0,001$; Tabla 5), sugiere que la relación entre ambos factores excede la noción de responder consistentemente de manera sesgada en ambas escalas. Por su parte, el desvanecimiento de esta correlación en la condición MC- M_{OP} ($r=-0,005$; $p=0,965$; *vid.*, Tabla 5) reveló que la asociación entre ambos factores de redacción contiene también a aquellos participantes que exhibieron un patrón de respuestas simétrico en la escala de autoestima y asimétrico en la de optimismo, y cuya dispersión de puntuaciones factoriales (puntos negros en la Figura 4) fue consistente con este resultado.

[Insertar Figura 4 aproximadamente aquí]

Tomados en conjunto, los resultados de los Estudios 3a y 3b sugieren que hay una baja consistencia por parte de los participantes en exhibir el patrón de respuesta asimétrico asociado a la dirección de los ítems entre escalas, reflejada en correlaciones moderadas entre los factores de redacción. Similares magnitudes de asociación han sido interpretadas tanto como consistentes con la hipótesis del estilo de respuesta (DiStefano & Motl, 2006; Horan et al., 2003) como con la hipótesis de que el efecto de redacción es específico de la escala (Kam, 2018a).

Ambas posibilidades fueron examinadas estimando la proporción conjunta de exhibir el patrón asimétrico en dos escalas, encontrando una baja proporción de individuos identificados simultáneamente con dicho patrón, incrementos específicos de la varianza explicada por el factor asociado al atributo en función de la escala con que fueron los participantes identificados y correlaciones entre factores de redacción que toman ventaja de individuos que no fueron identificados con respuestas asimétricas en ambas escalas. En este punto, nuestros resultados no respaldan la noción del estilo de respuesta generalizado entre escalas, sino más bien la idea de que el efecto de redacción sería específico de la escala o de naturaleza inespecífica (también referida como efímera). La siguiente serie de estudios tuvieron como objetivo examinar la evidencia asociada a la estabilidad temporal con el propósito de distinguir entre las distintas nociones sobre la naturaleza subyacente del efecto de redacción.

Estudio 4a. Evidencia longitudinal del efecto de redacción

Otro hallazgo reportado como evidencia concluyente de la aparente naturaleza sustantiva del sesgo de método, corresponde a la consistencia temporal del efecto de redacción. Siguiendo la estrategia propuesta en los estudios 3a y 3b, comenzamos examinando la relación longitudinal de los factores de redacción comparando el ajuste estadístico de tres modelos estructurales. A partir del modelo identificado como óptimo, examinamos la invarianza longitudinal del modelo de medición subyacente a lo largo de tres olas medidas en intervalos de un año.

Datos y participantes

Utilizamos los datos de las primeras tres olas del panel LISS (Longitudinal Internet Studies in the Social Sciences) del CentERdata de la Universidad de Tilburg. Luego de descartar 12 participantes (0,32% de la muestra inicial), la muestra de análisis quedó compuesta por 3712 participantes (54,15% mujeres; 0,02% no informó su sexo) que respondieron todos los ítems de la escala de Rosenberg en las tres olas y cuyas edades, al momento de aplicar la primera ola, oscilaron entre 16 y 95 años (Me=48,31; DE=15,76; 0,02% no informó su edad).

Plan de análisis

En primer lugar, utilizamos Análisis Factorial Confirmatorio (CFA) para examinar el ajuste de los modelos M1, M2 y M4 (Figura 2) para cada ola por separado. Del mismo modo, con el propósito de seleccionar el modelo que mejor se ajusta a los datos longitudinales, comparamos el ajuste de un modelo que no tuvo en cuenta los efectos de redacción (M8 en la Figura 5) con dos modelos bifactor FID alternativos que incorporan los factores asociados a redacción y la relación longitudinal entre ellos (M9 y M10 en la Figura 5, respectivamente). La bondad de ajuste fue evaluada utilizando el índice de ajuste comparativo (CFI), el índice de ajuste de Tucker-Lewis (TLI), el error cuadrático medio de aproximación (RMSEA) y la raíz cuadrada media estandarizada residual (SRMR). Utilizamos los mismos puntos de corte descritos anteriormente para interpretar el ajuste de los modelos examinados.

En segundo lugar, examinamos la invarianza de los parámetros estimados de la solución identificada como óptima, siguiendo una secuencia de especificación que incorporó modelos sucesivamente más restrictivos (Widaman, Ferrer, & Conger, 2010) para representar los diferentes niveles de equivalencia entre los parámetros del modelo en las tres mediciones: (1) invarianza configural como modelo de línea base, donde todos los parámetros fueron estimados libremente para examinar si la estructura factorial permanece constante a lo largo del tiempo, (2) invarianza métrica (incorporando restricciones a las cargas factoriales) para examinar si las variables latentes están relacionadas con los ítems de la misma forma a lo largo del tiempo, (3) invarianza escalar (agregando restricciones a los interceptos), e (4) invarianza residual (incorporando restricciones a los residuos de los indicadores). Adicionalmente, examinamos la invarianza estructural del modelo a nivel de la varianza de los factores y la covarianza entre los factores a lo largo de las tres olas evaluadas.

[Insertar Figura 5 aproximadamente aquí]

El procedimiento empleado para examinar la invarianza longitudinal, así como la estrategia de especificación, siguió los lineamientos propuestos por Hoffman (2015). Todos los modelos fueron estimados utilizando MLR como método de estimación y fueron llevados a cabo utilizando Mplus v.7.4. (Muthén & Muthén, 1998-2014). Los modelos de invarianza fueron comparados utilizando el test de chi-cuadrado basado en parámetros de no centralidad (pnc) para pequeñas diferencias en el ajuste (especificando una diferencia de RMSEA igual a 0,001; MacCallum, Browne, & Cai, 2006) e introduciendo la corrección de Steiger (1998) basada en el número de olas. Adicionalmente, se examinaron las diferencias de los indicadores de ajuste relativo (*i.e.*, CFI, TLI, RMSEA, SRMR) siguiendo los criterios definidos por Chen (2007) para tamaños muestrales grandes, donde valores de $CFI/TLI \leq 0,010$, $RMSEA \leq 0,015$ y $SRMR \leq 0,030$ (para métrica) o $\leq 0,010$ (para escalar y residual) indican invarianza entre los grupos.

Resultados

Estadísticos descriptivos. La Tabla 7 muestra los estadísticos descriptivos para las tres olas examinadas en este estudio. A través de un ANOVA de medidas repetidas, examinamos las diferencias de medias entre las olas evaluadas por ítem, utilizando la corrección de Bonferroni para contrastar dichas diferencias. En términos generales, se puede indicar que los ítems directos muestran un efecto pequeño y significativo del factor intra-sujeto ($F_{S(2,7422)} \geq 25,051$; $p_s < 0,001$; $\eta^2 p \leq 0,012$), cuyas comparaciones post-hoc sugieren que la media de la ola O3 fue inferior a las olas O1 y O2 (excepto en el ítem 6 donde todas difieren significativamente). En el caso de los ítems inversos, los ítems 5 y 10 arrojaron un efecto consistente al reportado con ítems directos ($F_{S(2,7422)} \geq 4,485$; $p_s < 0,001$) y un patrón idéntico de diferencias post-hoc, aunque con un tamaño de efecto aún más pequeño ($\eta^2 p = 0,001$), mientras que los restantes ítems inversos no exhibieron un efecto significativo ($F_{S(2,7422)} \leq 1,567$).

[Insertar Tabla 7 aproximadamente aquí]

Modelos confirmatorios. La Tabla 8 muestra los índices de ajuste de los modelos estimados. El modelo Bifactor FID arrojó excelentes índices de ajuste en la primera ($\chi^2_{(23)} = 187,258$; $p < 0,001$; CFI=0,984; TLI=0,968; RMSEA=0,044 [IC 90% =0,038; 0,050]; SRMR=0,019), segunda ($\chi^2_{(23)} = 106,829$; $p < 0,001$; CFI=0,992; TLI=0,984; RMSEA=0,031 [IC 90% =0,025; 0,037]; SRMR=0,019), y tercera ola ($\chi^2_{(23)} = 160,209$; $p < 0,001$; CFI=0,987; TLI=0,975; RMSEA=0,040 [IC 90% =0,034; 0,046]; SRMR=0,016). Todas las cargas factoriales fueron significativas. Consistente con este hallazgo, el modelo Bifactor FID que incorporó la relación temporal de los factores de redacción (M10) fue el que arrojó el mejor ajuste estadístico entre los modelos longitudinales examinados ($\chi^2_{(330)} = 971,678$; $p < 0,001$; CFI=0,986; TLI=0,982; RMSEA=0,023 [IC 90% =0,021; 0,025]; SRMR=0,022). Una inspección de las correlaciones entre los factores a lo largo del tiempo proporcionó evidencia de la estabilidad de los factores de redacción, exhibiendo asociaciones con un tamaño de efecto que fluctuó entre moderado y grande, tanto para los factores asociados a ítems negativos ($r_{O1O2} = 0,458$, $r_{O2O3} = 0,483$, y $r_{O1O3} = 0,424$; $p_s < 0,001$) como para los factores asociados a

ítems positivos ($r_{0102}=0,613$, $r_{0203}=0,598$, y $r_{0103}=0,508$; $ps<0,001$), aunque inferiores a las magnitudes observadas entre los factores de rasgo ($r_{0102}=0,826$, $r_{0203}=0,788$, y $r_{0103}=0,752$; $ps <0,001$).

[Insertar Tabla 8 aproximadamente aquí]

Invarianza longitudinal. A partir de la solución estructural identificada como óptima (M10), junto con la solución propuesta originalmente por Motl y DiStefano (utilizando el modelo Bifactor-FI; 2002)¹², examinamos la invarianza longitudinal entre las tres olas de medición. Los índices de ajuste para cada modelo de invarianza se muestran en la Tabla 9.

El modelo de invarianza configural arrojó que la estructura del modelo Bifactor-FID se mantiene a lo largo del tiempo ($\chi^2_{(330)}=971,678$; $p<0,001$; CFI=0,986; TLI=0,982; RMSEA=0,023 [IC 90% =0,021; 0,025]; SRMR=0,022). El modelo de invarianza métrica exhibió una variación no significativa del índice absoluto ($\chi^2_{(34)}=36,623$; $pnc=40,618$; $p=0,998$) y pequeñas diferencias en los índices relativos ($\Delta s<0,002$). Para la invarianza escalar de ambas medidas, los resultados de los índices de ajuste fueron similares a los obtenidos en el modelo métrico (CFI=0,985; TLI=0,983; RMSEA=0,022 [IC 90%=0,020; 0,024]; SRMR=0,024), con una diferencia no significativa entre ambos modelos ($\chi^2_{(14)}=45,218$; $pnc=29,423$; $p=0,411$). El modelo de invarianza residual exhibió ligeras diferencias respecto del modelo escalar (CFI=0,984; TLI=0,983; RMSEA=0,022 [IC 90%=0,021; 0,024]; SRMR=0,029), pero la diferencia entre ambos modelos no fue estadísticamente significativa ($\chi^2_{(20)}=67,154$; $pnc=34,129$; $p=0,162$), sugiriendo una solución invariante del modelo de medición. Por su parte, a nivel estructural, el modelo de invarianza de la varianza de los factores arrojó índices de ajuste relativos similares a los del modelo residual (CFI=0,984; TLI=0,983; RMSEA=0,022 [IC 90%=0,021; 0,024]; SRMR=0,033) y diferencias no significativas ($\chi^2_{(6)}=29,388$; $pnc=26,081$; $p=0,565$), siendo este resultado también encontrado al examinar el modelo de invarianza de las

¹² Optamos por incorporar el análisis de invarianza del modelo longitudinal propuesto por Motl y DiStefano (2002) con propósitos ilustrativos respecto a la replicabilidad de nuestros hallazgos respecto de aquellos previamente propuestos. Dado el hecho de que ambos modelos fueron invariantes, aquí describimos los resultados del análisis de invarianza del modelo Bifactor FID.

covarianzas de los factores (ver Tabla 9). Estos resultados, en términos de invarianza del modelo de medición y estructural fueron también encontrados con el análisis de invarianza longitudinal del modelo Bifactor-FI.

[Insertar Tabla 9 aproximadamente aquí]

Tomados en conjunto, estos resultados proveen evidencia de la estabilidad longitudinal del sesgo de método, reflejado a través de la magnitud y dirección de las covarianzas entre los factores de redacción, así como la consistencia temporal de la estructura factorial propuesta para separar la varianza de rasgo y de método. Estos hallazgos son consistentes con los reportados en investigaciones previas (Gana et al., 2014; Horan et al., 2003; Motl & DiStefano, 2002). El sexto y último estudio tuvo como objetivo verificar este patrón de estabilidad temporal en respuestas a ítems directos e inversos, utilizando transiciones latentes como aproximación basada en la heterogeneidad poblacional.

Estudio 4b. Transiciones temporales de respuestas asimétricas

La noción de estabilidad temporal, examinada a través de la invarianza longitudinal y la relación de los factores de redacción, ha sido propuesta como requisito esencial para proponer el sesgo de redacción como un estilo de respuesta (Marsh et al., 2010). El razonamiento supone que: (a) el mismo constructo (y sus respectivos factores de redacción) está siendo medido a lo largo del tiempo, y que (b) el orden relativo de los individuos en el constructo permanece constante a través del tiempo (Pitts, West, & Tein, 1996).

Otra posibilidad, hasta donde sabemos no explorada, implica suponer que, así como un individuo puede tener un rasgo relativamente constante a lo largo del tiempo, un estilo de respuesta asociado a responder diferencialmente a ítems directos e inversos implica que los individuos sean consistentemente identificados con dicho patrón de respuesta a lo largo del tiempo.

Datos y participantes

Los datos empleados en este estudio fueron los mismos que los utilizados en el Estudio 4a.

Plan de análisis

En primer lugar, utilizamos análisis de perfiles latentes (LPA) con los datos transversales de las olas por separado para examinar el número óptimo de perfiles en cada medición. Para identificar la solución óptima de número de clases, se empleó la misma estrategia y criterios de selección utilizados en el Estudio 1. Luego, a partir de las soluciones identificadas en cada ola de medición, utilizamos Análisis de Transiciones Latentes (LTA) para examinar los cambios en el tiempo del grupo de individuos que exhibe un patrón de respuestas diferencial a los ítems directos e inversos. Para abordar esto, examinamos dos especificaciones LTA alternativas al modelo no-invariante tradicional: (a) modelo invariante (*i.e.*, todos los parámetros de medición son iguales a lo largo del tiempo, permitiendo una interpretación más sencilla de los parámetros de transición, y (b) modelo invariante y estacionario (*i.e.*, incorporando restricciones a los parámetros de transición entre las olas O₁-O₂ y O₂-O₃) para asegurar que las probabilidades de transición entre perfiles equivalentes (*e.g.*, AE_A-AE_A) sea la misma a lo largo de todos los puntos de transición. Los modelos examinados fueron comparados utilizando la prueba de razón de verosimilitud. Sin embargo, debido a los inconvenientes que tiene este test para estimar apropiadamente las diferencias entre modelos cuando el tamaño muestral es suficientemente grande y hay un elevado número de celdas con bajas frecuencias (Muthén & Asparouhov, 2020; Nylund, Asparouhov, & Muthén, 2007), incluimos los índices criterio de información bayesianos (BIC, adjBIC) que incorporan una corrección al valor de verosimilitud basada en el número de parámetros y tamaño muestral (Schwartz, 1978; Sclove, 1987). Nylund et al. (2007) encontraron que estos índices funcionan mejor para modelos con tamaños muestrales, distribución de prevalencias y número de ítems equivalentes a los empleados en el presente estudio.

A partir de la solución de transiciones identificada como óptima, se estimaron las probabilidades de pertenencia a cada estado latente en las tres mediciones, las cuales representan la proporción que se espera que pertenezca a cada estado latente, así como las probabilidades de transición (τ_{ijk}) entre estados latentes que reflejan la probabilidad de que un individuo i transite al estado j en el tiempo t desde el estado k en el tiempo $t-1$.

Por su parte, tanto en el análisis de perfiles como en el de transiciones latentes, examinamos la precisión de la clasificación obtenida utilizando tanto el estadístico de entropía como la media de probabilidad de pertenencia a los perfiles (en el caso de LPA) o a los patrones de perfiles de transición (en el caso de LTA).

Una vez identificados los perfiles AE_{MET} en cada ola evaluada, examinamos los modelos M8-M10 siguiendo la estrategia de remoción utilizada en los Estudios 2 y 3b, examinando diez condiciones muestrales luego de remover: (a) los subgrupos de participantes identificados con el perfil AE_{MET} en la primera (MC-M₀₁), segunda (MC-M₀₂) y tercera ola (MC-M₀₃), (b) un número equivalente de participantes, aleatoriamente seleccionados a las condiciones MC-M₀₁, MC-M₀₂, y MC-M₀₃ (MC-R₀₁, MC-R₀₂, y MC-R₀₃), (c) el subgrupo de participantes identificados con el patrón de perfiles AE_{MET} en las tres olas conjuntamente (MC-M_{30Λ}), (d) un número idéntico de participantes a la condición MC-M_{30Λ}, seleccionados aleatoriamente (MC-R_{30Λ}), (e) los participantes que fueron identificados en al menos una de las olas examinadas (MC-M_{30V}), y (f) un número de participantes, seleccionados aleatoriamente, idéntico a la condición MC-M_{30V} (MC-R_{30V}). Todos los modelos fueron estimados siguiendo las especificaciones descritas en el Estudio 4a.

Resultados

Identificación de perfiles latentes. La Tabla 10 muestra los índices de ajuste para los modelos de dos a cinco perfiles en cada ola. Si bien los índices AIC y BIC no permitieron seleccionar un modelo en particular, los valores de VLMR y LMR respaldaron una solución de cuatro perfiles en la segunda y tercera ola, mientras que la proporción de individuos del perfil menos prevalente en la solución de cinco perfiles fue inferior al 5% en las tres mediciones. Por lo tanto, basados en los estadísticos de VLMR y LMR y el criterio de parsimonia, optamos por la solución de cuatro perfiles.

[Insertar Tabla 10 aproximadamente aquí]

Con respecto a la precisión en la asignación de los participantes a los perfiles latentes, los valores de entropía (*vid.*, Tabla 10) indican un buen nivel de clasificación de los participantes (Clark, 2010; Lubke & Muthén, 2007). Del mismo modo, los valores de

media de probabilidad de pertenencia al perfil fueron iguales o superiores a 0,88 en todas las mediciones (*vid.*, Tabla 11).

[Insertar Tabla 11 aproximadamente aquí]

Los cuatro perfiles identificados en cada ola exhibieron patrones de respuestas (*vid.*, panel superior de la Figura 6) y prevalencias consistentes con los reportados en los Estudios 1 y 2b, confirmando la presencia de respuestas asimétricas frente a ítems directos e inversos en cada ola evaluada. A partir del número de clases identificadas y el nivel de clasificación observado en cada ola, optamos por examinar los modelos de transiciones latentes siguiendo una solución de cuatro perfiles en cada ola.

Transiciones latentes. La Tabla 12 muestra los índices de ajuste de los modelos de transiciones examinados. El modelo LTA invariante arrojó un mejor grado de ajuste ($\text{LogL}=-156965,775$; $\text{BIC}=314778,141$; $\text{adjBIC}=314450,857$) en comparación con el obtenido con el modelo libre o no-invariante ($\text{LogL}=-159961,136$; $\text{BIC}=321377,093$; $\text{adjBIC}=320814,673$), sugiriendo que las clases identificadas en las tres olas son invariantes. Por su parte, el modelo de transiciones invariante con restricciones en los parámetros de transición exhibió índices de ajuste más altos a los obtenidos con el modelo invariante pero ligeramente inferiores a los observados con el modelo no-invariante ($\text{LogL}=-160203,243$; $\text{BIC}=321055,812$; $\text{adjBIC}=320804,788$). A partir de estos resultados, optamos por examinar los parámetros estimados del modelo LTA invariante, identificado como la solución óptima para este set de datos. El patrón de respuesta de los perfiles identificados de la solución LTA (*vid.*, panel inferior de la Figura 6) fue consistente con el observado tanto en las soluciones de perfiles estimadas en cada ola separadamente como en la solución no-invariante (paneles superior e intermedio en la Figura 6, respectivamente).

[Insertar Tabla 12 aproximadamente aquí]

En términos de la calidad de la clasificación, los valores de entropía de las tres soluciones estimadas fueron casi idénticas (*vid.*, Tabla 12). Con respecto a la precisión en la asignación a los patrones de perfiles, los valores de PMC fueron altos para todos los patrones de perfiles examinados (sección superior de la Tabla 13), siendo el patrón AE_A-

AE_M-AE_{MET} el que obtuvo el valor más bajo ($PMC=0,481$) y cuyo segundo patrón más probable en dicha clasificación arrojó un valor notoriamente inferior ($AE_M-AE_M-AE_{MET}$; $PMC=0,183$).

[Insertar Figura 6 aproximadamente aquí]

La sección intermedia de la Tabla 13 muestra las probabilidades de transición entre los perfiles identificados en las olas examinadas. Entre los perfiles identificados con respuestas simétricas (AE_A , AE_M , y AE_B), hubo una alta estabilidad en las membresías a lo largo del tiempo, con las probabilidades de transición notablemente más altas de permanecer en el mismo estado y valores similares de τ (sugiriendo estacionalidad entre los puntos de transición). Por su parte, en términos de cambio, los valores de transiciones más altos a un estado diferente fueron desde AE_B a AE_M ($\tau =0,18$ en ambos puntos de transición) y desde AE_M a AE_A ($\tau =0,15$ en ambos puntos de transición). Estos patrones de resultados sugieren un alto nivel de estabilidad dentro de los estados, y cuyas variaciones sugieren un incremento en el nivel de autoestima de los participantes entre las olas examinadas. Por su parte, el perfil identificado con respuestas asimétricas (AE_{MET}) exhibió una menor estabilidad en la membresía a lo largo del tiempo, con probabilidades de transición para aquellos individuos que permanecieron en el estado AE_{MET} que disminuyeron a lo largo del tiempo ($\tau_{O1O2}=0,53$; $\tau_{O1O2}=0,41$). En términos de cambio, las transiciones más frecuentes a un estado diferente variaron según los puntos de transición, siendo las transiciones $AE_{MET}-AE_M$ ($\tau=0,19$) y $AE_{MET}-AE_A$ ($\tau=0,17$) las más frecuentes entre O_1-O_2 , y $AE_{MET}-AE_M$ ($\tau=0,25$) y $AE_{MET}-AE_B$ ($\tau=0,20$) las más frecuentes entre O_2-O_3 .

[Insertar Tabla 13 aproximadamente aquí]

Finalmente, la sección inferior de la Tabla 13 muestra la distribución de frecuencias de los participantes asignados a cada patrón de perfiles obtenida a partir del modelo LTA invariante. En el caso de los perfiles con respuestas simétricas, la proporción de individuos identificados con el perfil AE_A en las tres olas ($n=1478$) fluctuó entre 74,5% y 81,9% del número de individuos identificados en cada ola examinada, mientras que en los participantes identificados consistentemente con el perfil AE_M ($n=492$), la proporción

fluctuó entre 45,6% y 51,6% y con el perfil AE_B ($n=216$) entre 42,5% y 55,4%. Estos valores reflejan una mayor consistencia en la identificación y asignación de los individuos al mismo patrón de respuestas en lugar de una combinación de perfiles o una asignación aislada en una ola específica.

Por su parte, la proporción de individuos consistentemente identificados con respuestas asimétricas (AE_{MET} ; $n=111$) fue considerablemente inferior, fluctuando entre 26,7% y 32,6% del número de individuos identificados con este perfil en las tres olas examinadas, sugiriendo una baja estabilidad a lo largo del tiempo. De hecho, una inspección más detallada de estos valores reveló que este número fue inferior a la cifra de participantes asignados únicamente al perfil AE_{MET} en las olas O_1 ($n=136$; b en la Tabla 13), O_2 ($n=122$; c en la Tabla 13), u O_3 ($n=129$; d en la Tabla 13). En este sentido, la probabilidad de responder asimétricamente sólo en una ola en particular es mayor a la de responder consistentemente de forma asimétrica en las tres olas examinadas.

Estos resultados, en contraposición a los expuestos en el Estudio 4a, sugieren una baja estabilidad temporal del efecto de redacción asociado a la dirección de los ítems, expresada en una menor e inestable probabilidad de transición entre perfiles AE_{MET} a lo largo del tiempo comparado con los restantes perfiles, así como en una baja proporción de individuos con el patrón de perfiles AE_{MET} en las tres olas.

Modelo longitudinal. Tal como en el Estudio 3b, examinamos los modelos longitudinales utilizando la estrategia de remoción descrita en el plan de análisis y estimamos los modelos M8-M10. La Tabla 14 muestra los índices de ajuste y la proporción de varianza explicada por los factores en las tres olas por condición muestral examinada.

Los modelos estimados bajo las condiciones de remoción de los individuos identificados con respuestas asimétricas por ola ($MC-M_{O1}$, $MC-M_{O2}$, y $MC-M_{O3}$) arrojaron una mejora en el ajuste observado en el modelo de un factor en comparación al obtenido en las respectivas condiciones muestrales control ($MC-R_{O1}$, $MC-R_{O2}$, y $MC-R_{O3}$), mientras que la mejora en los modelos Bifactor FID y FIDc fue marginal ($\Delta \leq 0,010$).

Más importante, se observó un incremento específico de la varianza explicada por el factor de rasgo de la ola desde donde los sujetos removidos fueron identificados (ver Tabla 14).

Por su parte, la condición MC-M_{30Λ}, que excluyó a los individuos que exhibieron respuestas asimétricas en las tres olas, arrojó una mejora marginal respecto al obtenido en la condición MC-R_{30Λ} en todos los modelos examinados y un leve incremento generalizado en la proporción de varianza capturada por los factores de rasgo a lo largo de las tres olas. Finalmente, la condición MC-M_{30V}, donde se descartaron los participantes identificados con el perfil AE_{MET} en al menos una ola, exhibió la mayor mejora en el ajuste en el modelo de un factor en comparación al de la condición MC-R_{30V}, así como un gran incremento generalizado en la proporción de varianza explicada por los factores de rasgo de cada ola.

En relación a la asociación longitudinal de los factores de rasgo (FR) y de redacción (FD y FI), la Tabla 15 muestra las correlaciones estimadas entre los factores especificados por modelo (M8-M10) en las condiciones muestrales examinadas. En el caso de las asociaciones longitudinales de los factores FR y FD, la variación de las magnitudes de las condiciones muestrales y sus respectivos controles (*i.e.*, $r_{MC-M} - r_{MC-R}$) fueron marginales (Me=0,004; DE=0,015), mientras que las observadas con el factor FI exhibieron un patrón de disminución específico a la condición de remoción empleada. Por ejemplo, en las condiciones de remoción basados en una ola específica (*i.e.*, MC-M₀₁, MC-M₀₂, y MC-M₀₃) el decremento observado fue mayor en aquellas asociaciones que involucraban la ola utilizada para la identificación y remoción (Me=-0,088; DE=0,018) en comparación a las restantes asociaciones observadas (Me=-0,041; DE=0,004). Por su parte, en la condición MC-M_{30Λ}, los decrementos observados fueron generalizados (Me=-0,109; DE=0,008) mientras que las variaciones en la condición MC-M_{30V} fueron marginales (Me=0,006; DE=0,044).

Pese a las disminuciones observadas en la magnitud de las asociaciones FI estimadas, esta persistió incluso en las condiciones MC-M_{30Λ} y MC-M_{30V}, en donde fueron removidos los participantes identificados con respuestas asimétricamente en todas las olas o en al menos una de ellas, respectivamente. Esto indica, tal como lo señalamos

en el Estudio 3b, que la asociación entre los factores de reacción del modelo bifactor estaría capitalizando la distribución conjunta de las puntuaciones factoriales de individuos que no exhiben patrones de respuestas asimétricos.

Estos resultados sugieren que la asociación longitudinal de los factores de redacción representa algo más (o quizás distinto) que la aparente estabilidad del efecto de redacción. Para ilustrar esto, examinamos la distribución conjunta de las puntuaciones factoriales que representaron la relación entre los factores de redacción asociados a ítems inversos en la primera y segunda ola (FI_{T1} y FI_{T2} ; *vid.*, Figura 7), estimadas con el modelo M10 en las condiciones MC y MC-M_{3OV}¹³.

[Insertar Figura 7 aproximadamente aquí]

La Figura 7 muestra la distribución conjunta de las puntuaciones factoriales de la asociación entre FI_{T1} y FI_{T2} en la condición MC, identificando los individuos con patrones de respuesta consistentes en las tres olas. Una inspección detallada de estas distribuciones reveló que los individuos con patrones de perfiles simétricos en las tres olas (AE_A , AE_M , y AE_B ; $n=2176$) exhibieron una distribución de puntuaciones alineadas con la asociación r_{0102} estimada con la muestra completa de participantes ($n=3712$). Junto con ello, las distribuciones de las puntuaciones factoriales en ambas olas de estos patrones de perfil (*vid.*, panel superior de la Figura 8) fueron simétricas, con medias cercanas a 0 y con una alta proporción de participantes con puntuaciones entre -1 y +1. Por su parte, el patrón de perfiles AE_{MET} en las tres olas ($n=111$) exhibió una asociación consistente con la observada con la muestra completa aunque con una distribución de puntuaciones en ambas olas mayormente dispersa. A partir de estos resultados podemos señalar que, pese a la evidente relación entre las puntuaciones FI_{T1} y FI_{T2} en el subgrupo AE_{MET} , la relación estimada con la muestra completa de participantes está dominada por la distribución conjunta de los individuos con respuestas simétricas y estables en un mismo nivel de autoestima en todas las olas examinadas ($n=2176$; 58,6% de la muestra).

[Insertar Figura 8 aproximadamente aquí]

¹³ Optamos por describir esta relación dado que las restantes asociaciones (r_{0103} y r_{0203}) exhibieron las mismas distribuciones conjuntas a la reportada utilizando r_{0102} .

Encontramos este patrón también en la condición MC-M_{3OV}, donde ningún participante identificado con respuestas asimétricas fue incluido en el análisis. La Figura 9 muestra la distribución conjunta de las puntuaciones factoriales de la asociación FI_{T1} y FI_{T2} en la condición MC-M_{3OV}, siguiendo la misma estrategia empleada en la Figura 7. El patrón de dispersión que exhibieron los participantes con perfiles consistentemente identificados como AE_A, AE_M, y AE_B en las tres olas fue similar al observado en la condición MC, aunque con algunas diferencias en la distribución global por ola respecto a la media y la proporción de individuos con puntuaciones entre -1 y +1 (*vid.*, panel inferior Figura 8), presumiblemente producto de la redistribución generada por la remoción de los individuos con patrones de perfiles AE_{MET} en las tres olas.

Discusión

El presente trabajo tuvo como propósito examinar la aparente naturaleza tipo-rasgo del efecto de redacción, utilizando la estrategia analítica propuesta por Ponce et al. (2021). En los Estudios 1 y 2, examinamos el valor del análisis de perfiles latentes para identificar un subgrupo de individuos que exhiben efectos de redacción en la escala RSES y LOT-R, y estimamos el impacto de su remoción en la unidimensionalidad subyacente de la estructura factorial de ambas escalas. Los resultados de ambos estudios revelaron la presencia de una proporción de individuos con un patrón de respuestas consistente con la noción de efectos de redacción, confirmando hallazgos previamente reportados (Ponce et al., 2021; Reise et al., 2016), junto con corroborar el valor de esta estrategia para detectar efectos de redacción.

Posteriormente examinamos las implicancias del uso de perfiles latentes y sus extensiones para examinar la naturaleza sustantiva o efímera del efecto de redacción, evaluando la consistencia de este efecto entre escalas balanceadas (Estudios 3a y 3b) y su estabilidad a lo largo del tiempo (Estudios 4a y 4b). Ambas son piezas de evidencia que han sido descritas como cruciales para respaldar la naturaleza tipo-rasgo del efecto de redacción. Los resultados de la serie de estudios revelaron que, en circunstancias que los modelos estructurales arrojaron resultados consistentes con la noción de un estilo de respuesta (Estudios 3a y 4a), los modelos basados en heterogeneidad poblacional

revelaron una baja proporción de participantes que exhibieron consistentemente efectos de redacción en distintas escalas balanceadas (Estudio 3b), o a lo largo del tiempo (Estudio 4b). Esto implica que, con el mismo set de datos, es posible obtener conclusiones opuestas dependiendo del tipo de técnica empleada.

Para abordar esta aparente contradicción, examinamos la distribución conjunta de las puntuaciones factoriales por patrón de perfiles identificados. En ambos tipos de evidencia, los resultados mostraron que la asociación estimada con modelos estructurales capitalizaba la distribución conjunta de individuos que respondieron simétricamente en al menos una de las dos escalas examinadas o en algunas de las olas evaluadas (e incluso de individuos que no fueron identificados con respuestas asimétricas). Esto requiere considerar con más detalle la utilidad del modelo bifactor, sus limitaciones y el valor de los factores específicos de redacción.

Modelos bifactor como estrategia para modelar efectos “intrascendentes”

El modelamiento bifactor (Holzinger & Harman, 1938; Holzinger & Swineford, 1937) permite asumir que la varianza común de un conjunto de ítems estaría dada tanto por la presencia de un factor general, que representa la comunalidad de todos los ítems, así como de una variedad de factores específicos, donde la varianza se comparte entre subconjuntos de ítems. Estas agrupaciones de ítems pueden representar subdominios específicos de un constructo en particular (como con medidas de personalidad y psicopatología; Reise, 2012) o que comparten la estrategia de medición (en este caso, la dirección con que los ítems son redactados). A partir de esta especificación, el modelo permitiría capturar la covarianza residual que no es capturada por el factor general (Reise, Bonifay, & Haviland, 2013; Reise, Moore, & Haviland, 2010).

El modelo bifactor, por lo tanto, parece idóneo para representar la multidimensionalidad que emerge de la presencia de respuestas asimétricas consistentes con la noción de efectos de redacción, permitiendo agrupar la varianza sistemática atribuible a la dirección de los ítems y utilizar los factores extraídos para examinar su relación estructural a lo largo del tiempo, con factores de redacción de otras escalas o con otros atributos, exhibiendo, en todos los casos, excelentes índices de ajuste. Sin embargo,

la selección e interpretabilidad del modelo bifactor a partir del ajuste estadístico obtenido ha sido extensamente cuestionada dada su propensión a acomodar todo tipo de patrones de respuesta (Bonifay & Cai, 2017; Reise et al., 2016), generando inestabilidad en las estimaciones de sus parámetros (Bonifay, 2015; Bonifay et al., 2017; Watts et al., 2019).

Frente a esta posibilidad, diversos autores han propuesto indicadores complementarios calculables a partir de una solución bifactor que permiten examinar la calidad del modelo más allá del ajuste estadístico (Rodríguez, Reise, & Haviland, 2016a; 2016b). Estos indicadores permiten inspeccionar: (a) el grado en que las puntuaciones totales y por subescala reflejan la contribución conjunta y única del factor general o los factores específicos (coeficientes ω , ω_h , ω_s , ω_{hs} ; McDonald, 1999; Revelle & Zinbarg, 2009; Zinbarg, Revelle, Yovel, & Li, 2005), (b) el grado en que el modelo refleja una medida esencialmente unidimensional (índice ECV; Sijtsma, 2009; Ten Berge & Socân, 2004), y (c) el grado en que una variable latente es adecuadamente reflejada a partir de un conjunto de ítems (índices H y FD; Grice, 2001; Hancock & Mueller, 2001), lo cual implica, en última instancia, su utilidad en la especificación en un modelo de medición en el contexto de modelamiento de ecuaciones estructurales (Rodríguez, Reise, & Haviland, 2016b).

Con respecto a los modelos bifactor aplicados al modelamiento de los efectos de redacción, estos indicadores han sido escasamente empleados para interpretar el valor informativo del modelo bifactor, así como de los factores específicos de redacción, los cuales, si bien no representan subdominios del constructo general (*e.g.*, autoestima, optimismo), corresponden a variables latentes que han sido utilizadas para proponer, utilizando modelamiento de ecuaciones estructurales, la naturaleza sustantiva del efecto de redacción. Ponce et al. (en preparación) revisaron diversos estudios que emplearon modelos bifactor para dar cuenta del efecto de redacción con la RSES y, a partir de las cargas factoriales estandarizadas, calcularon los indicadores descritos en el párrafo anterior. Los autores reportaron que, independiente de la multidimensionalidad subyacente, la estructura es esencialmente unidimensional y los factores específicos de redacción no se encuentran apropiadamente representados por sus respectivos indicadores

y, por ende, son inestables, difícilmente replicables e interpretables y, por lo tanto, no recomendables para su uso en modelamiento estructural.

Esto es crucial considerando que la mayor parte de la evidencia previamente reportada acerca de la naturaleza sustantiva o constructo-específica del efecto de redacción se sostiene a partir de la estimación de redes nomológicas para examinar la validez relacionada a diversos criterios (Cronbach & Meehl, 1955), utilizando el factor de redacción como una entidad distintiva del efecto de redacción y aplicando el ajuste estadístico como única prueba de su existencia y, por extensión, de las soluciones teóricas propuestas derivadas de las relaciones contenidas (pese a las limitaciones asociadas para distinguir entre estructuras de variables latentes; Molenaar & von Eye, 1994). Esta ha sido una práctica extensamente generalizada en la investigación psicológica post-conductista (Borsboom, Mellenbergh, & Heerden, 2003), ampliamente dominada por el auto-reporte (Baumeister et al., 2007).

Otra ventaja que ofrece el modelamiento bifactor de los efectos de redacción es que permitiría obtener una puntuación “refinada” del atributo, controlando el impacto del efecto de redacción sobre las asociaciones estimadas del rasgo con otras variables criterio. Esto es crucial si el efecto de redacción introduce un sesgo sustancial en los parámetros estructurales estimados. Sin embargo, Donnellan et al. (2016) y Michaelides, Koutsogiorgi, y Panayiotou (2017), examinaron la relación del factor de autoestima con diversas variables criterio (rasgos de personalidad, optimismo, satisfacción vital, ansiedad social, entre otras) utilizando tanto el modelo de un-factor como el modelo bifactor que tomaba en cuenta los efectos de redacción. En ambos casos, reportaron pequeñas diferencias en las correlaciones estimadas entre las estrategias de modelamiento empleadas, llevándolos a concluir que esta estrategia de control estadístico no tiene impacto sobre las asociaciones estimadas con variables criterio.

Consistente con estos hallazgos, Ponce et al. (en preparación), utilizando el sesgo de parámetro relativo promedio (ARPB; Reise et al., 2013) como indicador de la diferencia entre las cargas factoriales estimadas del factor general del modelo de un-factor y bifactor, concluyeron que la incorporación de los factores de redacción no influía

sustantivamente en la magnitud de los parámetros estimados. En nuestro caso, los resultados de los estudios 3b y 4b sugieren que el efecto de redacción fue omnipresente pero intrascendental sobre asociaciones longitudinales y transversales independiente de emplear tanto la estrategia de modelamiento de factores de redacción o la de exclusión de participantes con respuestas asimétricas.

Una naturaleza efímera

El enfoque más aceptado considera el efecto de redacción como una regularidad conductual consistente con la noción de estilo de respuesta (DiStefano & Motl, 2006; Horan et al., 2003). En este sentido, reflejaría una tendencia generalizada de respuesta que opera a través del tiempo y entre medidas de rasgos sustancialmente distintas. Pese a que numerosas investigaciones han reportado relaciones estructurales consistentes con esta noción de *efecto generalizado*, exhibiendo excelentes índices de ajuste, se debe notar que la mayoría de los estudios reporta tamaños de efecto pequeños a moderados (e.g., DiStefano & Motl, 2006; 2009; Donnellan et al., 2016; Lindwall et al., 2012; Quilty et al., 2006; por mencionar algunos), o exhibe baja replicabilidad de los hallazgos e inconsistencia de las relaciones reportadas entre factores de redacción y variables criterio (Kam, 2018a).

Este escenario ha llevado a algunos autores a considerar la idea de que la naturaleza de los factores de redacción es cualitativamente distinta entre escalas (*i.e.*, *efectos específicos de la escala*; Kam, 2018a), proponiendo que el grado con que los factores de redacción correlacionan con variables criterio depende de la escala empleada. Esta es una hipótesis empírica que introduce flexibilidad para explicar la baja relación observada entre factores de redacción, así como inconsistencias en la asociación con distintas variables criterio. Sin embargo, no explica resultados inconsistentes dentro de una misma escala. Por ejemplo, utilizando la RSES como escala balanceada y medidas de depresión (CESD-D; Radloff, 1977) y satisfacción vital (SWLS; Diener et al., 1985) como variables criterio, Alessandri et al. (2015) y Donnellan et al. (2016) encontraron relaciones inconsistentes de los factores de redacción a las originalmente reportadas por Lindwall et al. (2012).

Otra alternativa, a menudo descartada, considera los efectos de redacción como varianza sistemática de error de naturaleza efímera en cuanto a su estabilidad y consistencia (*i.e.*, *efectos inespecíficos*), lo cual implica la ausencia de cualquier tipo de tendencia generalizada en el tiempo o entre escalas (Marsh, 1996). La razón por la que se ha descartado la interpretación del efecto de redacción como un efecto inespecífico reside en que la asociación del factor de redacción con otras variables latentes criterio, independiente de la magnitud o consistencia reportada entre estudios, ha sido considerada evidencia suficiente para respaldar la idea de un estilo de respuesta generalizado. Sin embargo, considerando que este tipo de factores residualizados exhiben cargas factoriales congénicas bajas, tienden a ser factores con parámetros estimados inestables. Esto podría explicar el hecho de que la evidencia ha mostrado efectos débiles e inconsistentes entre estudios.

Nuestros hallazgos apoyan la interpretación de la naturaleza efímera de los efectos de redacción. A diferencia de lo propuesto por la noción de un estilo de respuesta generalizado, encontramos una baja proporción de individuos que exhibieron simultáneamente patrones de respuesta asimétrica en dos escalas distintas (Estudio 3b), aunque los factores de redacción de ambas escalas correlacionaron moderadamente (Estudio 3a). Del mismo modo, la baja proporción de individuos que exhibieron respuestas asimétricas en las tres olas examinadas (Estudio 4b) contradice tanto la hipótesis de efectos generalizados como la de efectos específicos de la escala.

Describir el efecto de redacción como de naturaleza efímera no significa que sea un efecto meramente aleatorio, sino que implica plantear que el patrón de respuestas asimétricas no emerge de procesos estables o determinísticos sino de procesos de respuesta transitorios de los participantes que interactúan con condiciones específicas del contexto de evaluación y que comprometerían la precisión de la medición de un atributo de interés. A continuación examinamos algunos de los mecanismos hipotetizados para explicar la emergencia de este efecto.

Algunos autores han propuesto que la respuesta diferencial en escalas balanceadas resulta de la interacción del nivel del atributo evaluado del participante y el nivel de

dificultad del ítem (Spector et al., 1997) o el tipo de ítem inverso empleado (Kam et al., 2021), produciendo diferentes distribuciones de respuesta. Esta noción deriva del *principio de punto ideal* (Andrich, 1988; Cliff et al., 1988; Thurstone, 1928) que propone que los sujetos tienden a estar de acuerdo con los ítems cercanos a su nivel de rasgo y en desacuerdo con los restantes, siguiendo una distribución curvilínea entre el grado de acuerdo y la posición individual sobre el nivel del atributo (Spector et al., 1997), independientemente de si el ítem refleja niveles muy altos o muy bajos del atributo¹⁴. De acuerdo con esta hipótesis, los participantes con niveles intermedios del atributo responderían empleando la misma categoría de respuesta a ítems positivos y negativos extremos¹⁵. Como explica Spector et al. (1997), esta combinación de patrones de respuesta resultaría, luego de recodificar los ítems inversos, en correlaciones más altas dentro de ambos subsets de ítems generando distorsiones en la estructura factorial.

Esta hipótesis ofrece diversas posibilidades para explicar nuestros hallazgos. En primer lugar, admite la posibilidad de encontrar diferencias en la expresión del efecto de redacción en escalas que evalúan diferentes constructos, donde tanto la variación del grado de dificultad de los ítems de ambas escalas o del nivel del individuo en ambos atributos permitirían explicar la inconsistencia en exhibir respuestas asimétricas. Por su parte, la falta de estabilidad longitudinal se explicaría por la variabilidad de los individuos a lo largo del tiempo para alcanzar o dejar el “punto ideal intermedio” del rasgo. De hecho, esta noción explicaría la presencia de individuos con respuestas asimétricas tanto estables como intermitentes.

Sin embargo, los hallazgos de McPherson y Mohr (2005) empleando la versión original de la escala de optimismo (LOT; Scheier & Carver, 1985) y de Nye et al. (2010) utilizando la escala de inteligencia emocional de Wong y Law (WLEIS; Wong & Law,

¹⁴ Esta noción difiere del *principio de dominancia* que subyace a los métodos tradicionales de construcción de escalas. Una discusión comparativa entre ambas nociones excede los márgenes del presente trabajo (para una discusión más detallada del trabajo en esta área, ver Chernyshenko et al., 2007; Drasgow et al., 2010; Reise, 2010).

¹⁵ Kam et al. (2021), empleando la estatura como atributo, ejemplificó cómo un participante de estatura promedio podría responder “en desacuerdo” tanto a un ítem *regular* (directo; “soy alto”) como a uno *polar opuesto* (inverso; “soy bajo”). Del mismo modo, respondería “de acuerdo” a un ítem *regular negado* (directo negado; “no soy alto”) como a uno *polar opuesto negado* (inverso negado; “no soy bajo”).

2002) sugieren que esta noción tiene un efecto marginal. En ambos estudios, los autores parafrasearon los ítems para crear una versión moderada de los ítems originalmente extremos y, con ello, disminuir la brecha de dichos ítems y los participantes moderados. El análisis factorial confirmatorio posterior reveló que ambas versiones no exhibieron diferencias sustantivas en la estructura factorial de ambas escalas. Los autores argumentaron que los participantes podrían “centrarse en la intención general del ítem, en lugar del sentido literal, y responder en consecuencia” (Nye et al., 2010, p. 807) de acuerdo a su propia interpretación del sentido de la pregunta. Este principio de reinterpretación de las afirmaciones extremas no se aplicaría tan fácilmente a atributos físicos (como a la “escala de estatura” empleada por Kam et al., 2021).

Por su parte, varios autores han propuesto que, en ocasiones, los participantes prestan menor esfuerzo o atención al contenido de los ítems y responden los ítems de una manera descuidada. Esto conlleva la emergencia de varianzas sistemáticas de error y, en consecuencia, factores espurios o artefactos, entre otras consecuencias (Arias et al., 2020). El concepto de patrón de respuesta descuidada (también referido como esfuerzo insuficiente para responder; Huang et al., 2012; 2015) ha sido empleado para describir diversos tipos de patrones de respuesta aleatorios¹⁶ o no aleatorios de una forma que no refleja con precisión el nivel del atributo (DeSimone et al., 2018; Meade & Craig, 2012), donde el individuo responde sin prestar suficiente atención al contenido ni a la dirección de los ítems.

Dos aspectos centrales de la noción de respuestas descuidadas relacionados con los hallazgos reportados en el presente estudio son la forma de respuesta y su naturaleza transitoria. Con respecto a la forma, los autores han sugerido que este tipo de respuestas exhiben un patrón predominantemente recto (*straightlining*, DeSimone et al., 2018), donde los participantes seleccionan categorías de respuesta que se concentran en el lado

¹⁶ La noción de patrones de respuesta aleatorios ha sido largamente cuestionada debido a que hay autores que han señalado que los individuos no son naturalmente capaces de generar números aleatorios (Figurska et al., 2008; Neuringer, 1986), y los respondedores desatentos tienden a exhibir diferentes niveles de sistematicidad, incluso si son directamente instruidos a responder aleatoriamente (ver, Huang et al., 2012). Otros autores describen el patrón aleatorio como una tendencia a utilizar todas las categorías de respuesta sin prestar atención al contenido de los ítems (DeSimone & Harms, 2018).

positivo o negativo de la escala de respuestas, independiente del contenido y la dirección de los ítems (Dunn et al., 2018). Luego de recodificar, este patrón resultará en una forma de respuesta asimétrica dependiente de la dirección de los ítems, como la que hemos identificado en el presente estudio. Con respecto al grado de intermitencia, los autores proponen que solo una proporción menor de individuos responde de manera descuidada de una manera generalizada (Meade & Craig, 2012). Esta característica podría explicar tanto la falta de consistencia entre escalas como la falta de estabilidad temporal del efecto de redacción, así como también que encontraríamos una baja proporción de individuos con patrones asimétricos generalizados.

Por otra parte, pese al grado de intermitencia del patrón de descuido, existe consenso en establecer una prevalencia de este patrón que oscila entre 5% y 15% de la muestra completa (Arias et al., 2020; Curran, 2016; Curran et al., 2010; DeRight & Jorgensen, 2015; Johnson, 2005; Kurtz & Parrish, 2001; Maniaci & Rogge, 2014; Meade & Craig, 2012) y cuyas variaciones dependerán de las condiciones de aplicación y los participantes. En nuestro estudio, la proporción de individuos con respuestas asimétricas osciló entre 7,3% y 13,2% (ver Tabla 11). Estos valores, aunque bajos, son suficientes para producir multidimensionalidad artefactual consistente con los efectos de redacción (Huang et al., 2015; Schmitt & Stults, 1985; Woods, 2006).

Esta noción ofrece una descripción más simple de los procesos subyacentes al efecto de redacción, proponiendo que la falta de atención y descuido de los encuestados actuaría como mecanismo principal e independiente del nivel de rasgo para explicar los patrones de respuesta asimétricos y, al mismo tiempo, actuaría transitoriamente alterando los procesos de respuesta de los encuestados, siendo su expresión moderada tanto por condiciones específicas de los participantes como del contexto de evaluación (Barnette, 1995; Edwards, 2019; Meade & Craig, 2012).

¿Invertir o no invertir ítems?

La inclusión de ítems inversos aborda diversos aspectos relacionados con los procesos de respuesta a ítems (Tourangeau et al., 2000), y cuyo propósito consiste en optimizar las condiciones de evaluación y evitar la estereotipación de los procesos de

respuesta producto de la emergencia de diversos tipos de sesgos. Por ejemplo, Weijters et al. (2013) sugirieron que los encuestados tienden a recordar episodios consistentes con el contenido del ítem (sesgo de confirmación). La inserción de ítems inversos permitiría recuperar episodios asociados tanto a la presencia del rasgo como a su ausencia, equilibrando la propensión a un determinado tipo de contenido. Análogamente, otros autores sugirieron que los ítems inversos permitirían controlar la tendencia de los participantes a responder utilizando una porción menor de la escala de respuesta debido a la iteración de ítems en una única dirección (sesgo de error de espacio; Likert, 1932) o categorías específicas de respuesta en escenarios de incertidumbre (sesgo de acquiescencia; Nunnally, 1978; Paulhus, 1991).

Pese a todas estas aparentes ventajas, diversos investigadores, basados en los hallazgos reportados de las propiedades métricas de escalas balanceadas, han cuestionado el grado en que las escalas balanceadas ofrecen una condición de medición más precisa que las versiones compuestas sólo por ítems directos o inversos. La estrategia más común para examinar esto consiste en comparar las estimaciones de validez y confiabilidad entre escalas compuestas únicamente por ítems directos, inversos o escalas que contienen simultáneamente ambos tipos de ítems, creando condiciones de medición del atributo (*e.g.*, Dueber et al., 2021; Greenberger et al., 2003; Suárez-Álvarez et al., 2018; Zhang et al., 2016). Estos estudios coinciden en señalar que la inclusión de ítems inversos perjudica el ajuste de modelos unidimensionales y altera la precisión del instrumento de medición. Para asociaciones con otras variables, sin embargo, la evidencia muestra que no hay diferencias significativas según condición de medición (Dueber et al., 2021). Tomados en conjunto, estos hallazgos han llevado a los autores a cuestionar la utilidad práctica de los ítems inversos en cuestionarios de auto-reporte.

Sin embargo, la descripción del efecto de redacción basada en la presencia de participantes con patrones de respuestas independientes del contenido del ítem permite reinterpretar estos hallazgos. Por ejemplo, considerando el hecho de que este tipo de patrones operan independientemente del contenido y dirección de los ítems, se espera que los participantes exhiban estos patrones de respuesta en cualquiera de las variaciones de

la escala, con consecuencias comunes sobre el grado de precisión de la medición a nivel individual pero diferenciales a nivel de escala según la condición de medición. En el caso de las escalas balanceadas, la recodificación de los ítems inversos permite identificar el patrón de respuestas independientes del contenido mientras que en el caso de las escalas compuestas exclusivamente por ítems directos o inversos, este patrón pasaría inadvertido e incluso contribuiría a la unidimensionalidad y la consistencia interna. Por su parte, la ausencia de diferencias en la asociación del rasgo objetivo con otras variables criterio según condición de medición se explicaría tanto por la baja proporción de individuos con respuestas aberrantes como por el hecho de que sus puntuaciones se mueven en torno a la media de la escala, con una menor capacidad de palanca sobre la relación dependiendo del grado de asimetría de la distribución de puntuaciones observada con la muestra completa.

Nuestros análisis de las escalas de autoestima y optimismo disposicional, utilizando modelamiento de mixturas, nos permitieron identificar este patrón de respuestas y examinar, utilizando estrategias de remoción, su impacto sobre las propiedades psicométricas del instrumento, entre otras características. Otros autores han propuesto estrategias alternativas para detectar respuestas aberrantes con escalas que contienen o no ítems inversos, incluyendo técnicas de evaluación atencional (ítems *bogus* o de respuesta instruida; Meade & Craig, 2012; Oppenheimer et al., 2009), evaluación de datos auxiliares o parados (latencias de respuesta o fijaciones visuales; Henninger & Plieninger, 2021; Koutsogiorgi & Michaelides, 2022; Zhang & Conrad, 2014) y detección de valores atípicos o funciones de respuesta (Curran et al., 2016). Sin embargo, la evidencia muestra una serie de inconvenientes asociados al uso e interpretación de estas técnicas, incluyendo la falta de efectividad y consistencia entre ellas debido a la arbitrariedad con que se definen sus puntos de corte, dependiendo del conjunto de datos utilizado (Curran, 2016; Nissen et al., 2016) y una baja sensibilidad y especificidad para la detección de patrones de respuestas no completamente aleatorias (Meade et al., 2017; Schroeders et al., 2022).

Si bien la inclusión de ítems inversos en escalas balanceadas tiene como propósito optimizar los procesos cognitivo-atencionales de los encuestados para obtener respuestas

más exhaustivas, una baja proporción de participantes exhibió un patrón de respuestas con efectos significativos sobre la dimensionalidad y confiabilidad de la escala. Algunos autores han considerado esto como un argumento suficiente para justificar el rechazo de esta estrategia. Considerando las limitaciones de las estrategias de detección descritas en el párrafo anterior, así como la reticencia de algunos investigadores a examinar la calidad de sus datos (Borsboom, 2006), consideramos necesario el uso y adecuado análisis de los ítems inversos como estrategia de detección y manejo de participantes inconsistentes (Patton et al., 2019; Ponce et al., 2021). Peor sería ignorar el problema o asumir que no tiene consecuencias relevantes, simplemente porque descartamos o no utilizamos correctamente las estrategias para su detección.

Limitaciones y direcciones para futuras investigaciones

Nuestros resultados confirman el valor del modelamiento de mixturas como estrategia analítica, no sólo para identificar aquellos participantes que exhiben efectos de redacción, sino también para examinar la evidencia crítica asociada a la naturaleza subyacente al patrón de respuesta basado en la dirección de los ítems.

Sin embargo, estos hallazgos deben ser corroborados y extendidos empleando otras escalas y muestras de encuestados. Adicionalmente, se requieren más estudios que contribuyan a comprender por qué ciertos participantes pueden dejar de prestar atención a los ítems o exhibir respuestas descuidadas, así como a identificar si esto se relaciona con características o condiciones especiales asociadas a los instrumentos de evaluación utilizados. La posibilidad de explorar las bases sustantivas de estos efectos y los procesos subyacentes que los desencadenan puede abrir interesantes líneas de investigación, incorporando otras disciplinas, como la neurociencia conductual, así como técnicas de recolección de información menos invasivas.

Referencias

- Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement, 12*(1), 33-51. <https://doi.org/10.1177/014662168801200105>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. <http://dx.doi.org/10.1109/TAC.1974.1100705>
- Alessandri, G., Vecchione, M., Eisenberg, N., & Laguna, M. (2015). On the factor structure of the Rosenberg (1965) General Self-Esteem Scale. *Psychological Assessment, 27*(2), 621-536. <https://doi.org/10.1037/pas0000073>
- Arias, V.B., & Arias, B. (2017). The negative wording factor of Core Self-Evaluations Scale (CSES): Methodological artifact, or substantive specific variance? *Personality and Individual Differences, 109*, 28-34. <https://doi.org/10.1016/j.paid.2016.12.038>
- Arias, V.B., Garrido, L.E., Jenaro, C., Martínez-Molina, A., Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods, 52*, 2489-2505. <https://doi.org/10.3758/s13428-020-01401-8>
- Arias, V.B., Jenaro, C., & Ponce, F.P. (2018). Testing the generality of the general factor of personality: An exploratory bifactor approach. *Personality and Individual Differences, 129*, 17-23. <https://doi.org/10.1016/j.paid.2018.02.042>
- Barnette, J.J. (1995, November 9). Responses that may indicate nonattending behaviors in three self-administered educational surveys [Conference Papers]. Annual Meeting of the Mid-South Educational Research Association, Biloxi, MS, United States. <https://eric.ed.gov/?id=ED393897>
- Baumeister, R.F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements. *Perspectives on Psychological Science, 2*(4), 396-403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>

- Bentler, P.M., Jackson, D.N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 76(3), 186-204. <https://doi.org/10.1037/h0031474>
- Bonifay, W. (2015). An integrative framework of model evaluation (Unpublished doctoral dissertation). University of California, Los Angeles.
- Bonifay, W., & Cai, L. (2017). On the complexity of Item Response Theory models. *Multivariate Behavioral Research*, 52(4), 465-484. <https://doi.org/10.1080/00273171.2017.1309262>
- Borsboom, D., Mellenbergh, G.J., van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203-219.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71, 425-440. <https://doi.org/10.1007/s11336-006-1447-6>
- Brissette, I., Scheier, M. F., & Carver, C. S. (2002). The role of optimism in social network development, coping, and psychological adjustment during a life transition. *Journal of Personality and Social Psychology*, 82(1), 102–111. <https://doi.org/10.1037/0022-3514.82.1.102>
- Browne, M.W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258. <http://dx.doi.org/10.1177/0049124192021002005>
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80(4), 796-846. <https://doi.org/10.1111/j.1467-6494.2011.00749.x>
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195-212. <http://doi.org/10.1007/BF01246098>
- Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>

- Chernyshenko, O.S., Stark, S., Drasgow, F., & Roberts, B.W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19*(1), 88-106. <https://doi.org/10.1037/1040-3590.19.1.88>
- Clark, S.L. (2010). Mixture-modeling with behavioral data (Doctoral dissertation). University of California, Los Angeles, Los Angeles, CA.
- Cliff, N., Collins, L.M., Zarkin, J., Gallipeau, D., & McCormick, D.J. (1988). An ordinal scaling method for questionnaire and other ordinal data. *Applied Psychological Measurement, 12*(1), 83-97. <https://doi.org/10.1177/014662168801200108>
- Collins, L.M., & Lanza, S.T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. New York: Wiley.
- Cronbach, L.J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*(4), 475-494. <https://doi.org/10.1177/001316444600600405>
- Cronbach, L.J., (1956). Assessment of individual differences. *Annual Review of Psychology, 7*, 173-196. <https://doi.org/10.1146/annurev.ps.07.020156.001133>
- Curran, P.G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19. <https://doi.org/10.1016/j.jesp.2015.07.006>
- Curran, P.G., Kotrba, L., & Denison, D. (2010, April 8-10). Careless responding in surveys: Applying traditional techniques to organizational settings [Poster Presentation]. 25th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA, United States.
- DeRight, J., & Jorgensen, R.S. (2015). I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *The Clinical Neuropsychologist, 29*, 101-117. <https://doi.org/10.1080/13854046.2014.989267>
- DeSimone, J.A., DeSimone, A.J., Harms, P.D., & Wood, D. (2018). The differential impacts of two forms of insufficient effort responding. *Applied Psychology, 67*(2), 309–338. <https://doi.org/10.1111/apps.12117>

- DeSimone, J.A., & Harms, P.D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology*, 33(5), 559-577. <https://doi.org/10.1007/s10869-017-9514-9>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, 49(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13
- DiStefano, C., & Motl, R.W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13(3), 440-464. https://doi.org/10.1207/s15328007sem1303_6
- DiStefano, C., & Motl, R.W. (2009). Personality correlates of method effects due to negatively worded items on the Rosenberg self-esteem scale. *Personality and Individual Differences*, 46(3), 309-313. <https://doi.org/10.1016/j.paid.2008.10.020>
- Donnellan, M.B., Ackerman, R.A., & Brecheen, C. (2016). Extending structural analyses of the Rosenberg self-esteem scale to consider criterion-related validity: Can composite self-esteem scores be good enough? *Journal of Personality Assessment*, 98(2), 169-177. <http://dx.doi.org/10.1080/00223891.2015.1058268>
- Drasgow, F., Chernyshenko, O.S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3(4), 465-476. <https://doi.org/10.1111/j.1754-9434.2010.01273.x>
- Dueber, D.M., Toland, M.D., Lingat, J.E., Love, A.M.A., Qiu, C., Wu, R., & Brown, A.V. (2021). To reverse item orientation or not to reverse item orientation, That is the question. *Assessment*. <https://doi.org/10.1177/10731911211017635>
- Dunbar, M., Ford, G., Hunt, K., & Der, G. (2000). Question wording effects in the assessment of global self-esteem. *European Journal of Psychological Assessment*, 16(1), 13-19. <https://doi.org/10.1027/1015-5759.16.1.13>
- Dunn, A.M., Heggstad, E.D., Shanock, L.R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, 33(1), 105-121. <https://doi.org/10.1007/s10869-016-9479-0>

- Edwards, J.R. (2019). Response invalidity in empirical research: Causes, detection, and remedies. *Journal of Operations Management*, 65(1), 62-76.
<https://doi.org/10.1016/j.jom.2018.12.002>
- Figurska, M., Stańczyk, M., & Kulesza, K. (2008). Humans cannot consciously generate random numbers sequences: Polemic study. *Medical Hypotheses*, 70(1), 182-185.
<https://doi.org/10.1016/j.mehy.2007.06.038>
- Finch, W.H., & Bronk, K.C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling*, 18(1), 132-151.
<https://doi.org/10.1080/10705511.2011.532732>
- Gana, K., Saada, Y., Bailly, N., Joulain, M., Hervé, C., & Alaphilippe, D. (2013). Longitudinal factorial invariance of the Rosenberg Self-Esteem Scale: Determining the nature of method effects due to item wording. *Journal of Research in Personality*, 47(4), 406-416. <https://doi.org/10.1016/j.jrp.2013.03.011>
- Geukes, K., Nestler, S., Hutteman, R., Küfner, A.C.P., & Back, M.D. (2017). Trait personality and state variability: Predicting individual differences in within- and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, 69, 124-138.
<https://doi.org/10.1016/j.jrp.2016.06.003>
- Gnams, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg self-esteem scale: A cross-cultural meta-analysis. *Zeitschrift für Psychologie*, 226(1), 14-29. <https://doi.org/10.1027/2151-2604/a000317>
- Gnams, T., & Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment*, 1-15.
<https://doi.org/10.1177/1073191117746503>
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S.P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem scale: Do they matter? *Personality and Individual Differences*, 35(6), 1241-1254. [https://doi.org/10.1016/S0191-8869\(02\)00331-8](https://doi.org/10.1016/S0191-8869(02)00331-8)

- Greene, A.L., Eaton, N.R., Li, K., Forbes, M.K., Krueger, R.F., Markon, K.E., Waldman, I.D., Cicero, D.C., Conway, C.C., Docherty, A.R., Fried, E.I., Ivanova, M.Y., Jonas, K.G., Latzman, R.D., Patrick, C.J., Reininghaus, U., Tackett, J.L., Wright, A.G.C., & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology, 128*(7), 740-764. <https://doi.org/10.1037/abn0000434>
- Grice, J.W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430-450. <https://doi.org/10.1037/1082-989X.6.4.430>
- Haefffel, G.J., & Howard, G.S. (2010). Self-report: Psychology's four-letter word. *The American Journal of Psychology, 123*(2), 181-188. <https://doi.org/10.5406/amerjpsyc.123.2.0181>
- Hancock, G.R., & Mueller, R.O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future*. Lincolnwood, IL: Scientific Software International.
- Henninger, M., & Plieninger, H. (2021). Different styles, different times: How response times can inform our knowledge about the response process in rating scale measurement. *Assessment, 28*(5), 1301-1319. <https://doi.org/10.1177/1073191119900003>
- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. NY: Routledge.
- Holzinger, K.J., & Harman, H.H. (1938). Comparison of two factorial analyses. *Psychometrika, 3*, 45-60. <http://dx.doi.org/10.1007/BF02287919>
- Holzinger, K.J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*, 41-54. <http://dx.doi.org/10.1007/BF02287965>
- Horan, P.M., DiStefano, C., & Motl, R.W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling, 10*(3), 435-455. https://doi.org/10.1207/S15328007SEM1003_6

- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. <http://dx.doi.org/10.1080/10705519909540118>
- Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M., & DeShon, R.P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27, 99-114. <https://doi.org/10.1007/s10869-011-9231-8>
- Huang, J.L., Liu, M., & Bowling, N.A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100(3), 828-845. <http://dx.doi.org/10.1037/a0038510>
- Jackson, D.N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55(4), 243-252. <http://dx.doi.org/10.1037/h0045996>
- Johnson, J.A. (2005). Ascertain the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, 39(1), 103-129. <https://doi.org/10.1016/j.jrp.2004.09.009>
- Kam, C.C.S., & Meyer, J.P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, 18(3), 512-541. <https://doi.org/10.1177/1094428115571894>
- Kam, C.C.S. (2018a). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociological Methods & Research*, 47(3), 574-597. <https://doi.org/10.1177/0049124115626177>
- Kam, C.C.S. (2018b). Testing the assumption of population homogeneity in the measurement of dispositional optimism: Factor mixture modeling analysis. *Journal of Personality Assessment*, 102(2), 238-249. <https://doi.org/10.1080/00223891.2018.1502194>
- Kam, C.C.S. (2021). Bifactor model is not the best-fitting model for self-esteem: Investigation with a novel technique. *Assessment*, 28(7), 1799-1812. <https://doi.org/10.1177/1073191120949916>

- Kam, C.C.S., Meyer, J.P., & Sun, S. (2021). Why do people agree with both regular and reversed items? A logical response perspective. *Assessment*, 28(4), 1110-1124.
<https://doi.org/10.1177/10731911211001931>
- Kamoen, N., Holleman, B., Mak, P., Sanders, T., & van den Bergh, H. (2011). Agree or disagree? Cognitive processes in answering contrastive survey questions. *Discourse Processes*, 48(5), 355-385. <http://dx.doi.org/10.1080/0163853X.2011.578910>
- Kamoen, N., Holleman, B., Nouwen, R., Sanders, T., & van den Bergh, H. (2011). Absolutely relative or relatively absolute? The linguistic behavior of gradable adjectives and degree modifiers. *Journal of Pragmatics*, 43(13), 3139-3151.
<https://doi.org/10.1016/j.pragma.2011.05.017>
- Kamoen, N., Holleman, B., van den Bergh, H. (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, 7(3), 181-189. <https://doi.org/10.18148/srm/2013.v7i3.5034>
- Koutsogiorgi, C.C., & Michaelides, M.P. (2022). Response tendencies due to item wording using eye-tracking methodology accounting for individual differences and item characteristics. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01719-x>
- Kurtz, J.E., & Parrish, C.L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment*, 76(2), 315-332.
https://doi.org/10.1207/S15327752JPA7602_12
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 5-55.
- Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, 94(2), 196-204.
<http://dx.doi.org/10.1080/00223891.2011.645936>
- Lo, Y., Mendell, N., & Rubin, D. (2001). Testing t 1 he number of components in a normal mixture. *Biometrika*, 88, 767-778. <http://doi.org/10.1093/biomet/88.3.767>

- Lubke, G.H., & Muthén, B.O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*(1), 32-39.
<https://doi.org/10.1037/1082-989X.10.1.21>
- Lubke, G.H., & Muthén, B.O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class specific parameters. *Structural Equation Modeling, 14*(1), 26-47. http://doi.org/10.1207/s15328007sem1401_2
- MacCallum, R.C., Browne, M.W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods, 11*(1), 19-35. <https://doi.org/10.1037/1082-989X.11.1.19>
- Maniaci, M.R., & Rogge, R.D. (2014). Caring about careless: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61-83.
<https://doi.org/10.1016/j.jrp.2013.09.008>
- Markus, K.A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.
- Marsh, H.W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*, 810-819. <https://doi.org/10.1037/0022-3514.70.4.810>
- Marsh, H.W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341. https://doi.org/10.1207/s15328007sem1103_2
- Marsh, H.W., Scalas, L.F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22*(2), 366-381.
<https://doi.org/10.1037/a0019225>
- McDonald, R.P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.
- McPherson, J., & Mohr, P. (2005). The role of item extremity in the emergence of keying-related factors: An exploration with the Life Orientation Test. *Psychological Methods, 10*(1), 120-131. <https://doi.org/10.1037/1082-989X.10.1.120>

- Meade, A.W., Allred, C.M., Pappalardo, G., & Stoughton, J.W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, 76, 417-430.
<https://doi.org/10.1016/j.chb.2017.06.032>
- Meade, A.W., & Craig, S.B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437-455. <https://doi.org/10.1037/a0028085>
- Michaelides, M.P., Koutsogiorgi, C., & Panayiotou, G. (2016). Method effects on an adaptation of the Rosenberg self-esteem scale in Greek and the role of personality traits. *Journal of Personality Assessment*, 98(2), 178-188.
<http://dx.doi.org/10.1080/00223891.2015.1089248>
- Michaelides, M.P., Koutsogiorgi, C., & Panayiotou, G. (2017). Method/group factors: Inconsequential but meaningful—A comment on Donnellan, Ackerman, and Brecheen (2016). *Journal of Personality Assessment*, 99(3), 334-335.
<https://doi.org/10.1080/00223891.2016.1233560>
- Molenaar, P.C.M., & von Eye, A. (1994). On the arbitrary nature of latent variables. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis*. Thousand Oaks: Sage.
- Morgan, G.B., Hodge, K.J., Wells, K.E., & Watkins, M.W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, 3, 2-20. <https://doi.org/10.3390/jintelligence3010002>
- Motl, R.W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*, 9(4), 562-578. http://dx.doi.org/10.1207/S15328007SEM0904_6
- Murray, A.L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, 41(5), 407-422. <https://doi.org/10.1016/j.intell.2013.06.004>
- Muthén, B., & Asparouhov, T. (2020). Latent transition analysis with random intercepts (RI-LTA). *Psychological Methods*. Advance online publication.
<https://doi.org/10.1037/met0000370>

- Muthén, B.O., & Muthén, L.K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24(6), 882–891. <http://doi.org/10.1111/j.1530-0277.2000.tb02070.x>
- Muthén, L.K., & Muthén, B.O. (1998-2014). *Mplus User's Guide* (7th Ed.). Los Angeles, CA: Muthén & Muthén.
- Neely-Barnes, S. (2010). Latent class models in social work. *Social Work Research*, 34(2), 114-121. <http://doi.org/10.1093/swr/34.2.114>
- Neuringer, A. (1986). Can people behave "randomly?" The role of feedback. *Journal of Experimental Psychology: General*, 115, 62-75. <https://doi.org/10.1037/0096-3445.115.1.62>
- Nissen, A.S.M., Meijer, R.R., & Tendeiro, J.N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality*, 63, 1-11. <https://doi.org/10.1016/j.jrp.2016.04.010>
- Nunnally, J.M. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- Nye, C.D., Newman, D.A., & Joseph, D.L. (2010). Never say “always”? Extreme item wording effects on scalar invariance and item response curves. *Organizational Research Methods*, 13(4), 806-830. <https://doi.org/10.1177/1094428109349512>
- Nylund, K.L., Asparouhov, T. & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- Oppenheimer, D.M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45 (4), 867-872. <https://doi.org/10.1016/j.jesp.2009.03.009>
- Osborne, D., & Sibley, C.G. (2017). Identifying “types” of ideologies and intergroup biases: Advancing a person-centred approach to social psychology. *European Review of Social Psychology*, 28(1), 288-332. <http://dx.doi.org/10.1080/10463283.2017.1379265>

- Patton, J.M., Cheng, Y., Hong, M., & Diao, Q. (2019). Detection and treatment of careless responses to improve item parameter estimation. *Journal of Educational and Behavioral Statistics, 44*(3), 309-341.
<https://doi.org/10.3102/1076998618825116>
- Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press.
- Pitts, S.C., West, S.G., & Tein, J. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning, 19*(4), 333-350. [https://doi.org/10.1016/S0149-7189\(96\)00027-4](https://doi.org/10.1016/S0149-7189(96)00027-4)
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
<https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P.M., MacKenzie, S.B., & Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Ponce, F.P., Gómez, F., Cumsille, P., & Arias, V.B. (en preparación). The use and abuse of wording factors in the RSES: A systematic review and examination of model quality.
- Ponce, F.P., Torres-Iribarra, D., Vergés, A.J., & Arias, V.B. (2021). Wording effects in assessment: Missing the trees for the forest. *Multivariate Behavioral Research*.
<http://dx.doi.org/10.1080/00273171.2021.1925075>
- Quilty, L.C., Oakman, J.M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling, 13*(1), 99-117.
http://dx.doi.org/10.1207/s15328007sem1301_5
- Radloff, L.S. (1977). The CES–D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401.
<https://doi.org/10.1177/014662167700100306>

- Raykov, T., Marcoulides, G.A., Menold, N., & Harrison, M. (2019). Revisiting the bifactor model: Can mixture modeling help assess its applicability? *Structural Equation Modeling*, 26(1), 110-118.
<https://doi.org/10.1080/10705511.2018.1436441>
- Reise, S.P. (2010). Thurstone might have been right about attitudes, but Drasgow, Chernyshenko, and Stark fail to make the case for personality. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3(4), 485-488.
<https://doi.org/10.1111/j.1754-9434.2010.01276.x>
- Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667-696.
<http://dx.doi.org/10.1080/00273171.2012.715555>
- Reise, S.P., Bonifay, W.E., & Haviland, M.G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129-140.
<https://doi.org/10.1080/00223891.2012.725437>
- Reise, S.P., Kim, D.S., Mansolf, M., & Widaman, K.F. (2016). Is the bifactor model a better model or is it just better a modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg self-esteem scale. *Multivariate Behavioral Research*, 51(6), 818-838.
<http://dx.doi.org/10.1080/00273171.2016.1243461>
- Reise, S.P., Moore, T.M., & Haviland, M.G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544-559.
<http://dx.doi.org/10.1080/00223891.2010.496477>
- Reise, S.P., Scheines, R., Widaman, K.F., & Haviland, M.G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73, 5-26.
<https://doi.org/10.1177/0013164412449831>

- Revelle, W., & Zinbarg, R.E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74, 145-154.
<http://dx.doi.org/10.1007/s11336-008-9102-z>
- Rodebaugh, T.L., Woods, C.M., Thissen, D.M., Heimberg, R.G., Chambless, D.L., & Rapee, R.M. (2004). More information from fewer questions: The factor structure and item properties of the original and brief fear of negative evaluation scale. *Psychological Assessment*, 16, 169-181. <https://doi.org/10.1037/1040-3590.16.2.169>
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016a). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137-150. <https://doi.org/10.1037/met0000045>
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016b). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223-237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: University Press.
- Roszkowski, M.J., & Soven, M. (2010). Shifting gears: consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113-130.
<https://doi.org/10.1080/02602930802618344>
- Sackett, P.R., Lievens, F., Van Iddekinge, C.H., & Kuncel, N.R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology*, 102, 254-273. <https://doi.org/10.1037/apl0000151>
- Scheier, M.F., & Carver, C.S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4(3), 219-247. <https://doi.org/10.1037/0278-6133.4.3.219>
- Scheier, M.F., Carver, C.S., & Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67(6), 1063-1078. <https://doi.org/10.1037/0022-3514.67.6.1063>

- Scherpenzeel, A. (2011). Data collection in a probability-based Internet panel: How the LISS Panel was built and how it can be used. *Bulletin de Méthodologie Sociologique*, 109, 56-61. <http://dx.doi.org/10.1177/0759106310387713>
- Schmitt, D.P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, 89, 623–642. <https://doi.org/10.1037/0022-3514.89.4.623>
- Schmitt, N., & Stults, D.M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9(4), 367-373. <https://doi.org/10.1177/014662168500900405>
- Schroeders, U., Schmidt, C., & Gnams, T. (2022). Detecting careless responding in survey data using stochastic gradient boosting. *Educational and Psychological Measurement*, 82(1), 29-56. <https://doi.org/10.1177/00131644211004708>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464. <http://dx.doi.org/10.1214/aos/1176344136>
- Sclove, S.L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343. <https://doi.org/10.1007/BF02294360>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107-120. <https://doi.org/10.1007/s11336-008-9101-0>
- Spector, P.E., Van Katwyk, P.T., Brannick M.T., & Chen P.Y. (1997). When two factors don't reflect two constructs: how item characteristics can produce artifactual factors. *Journal of Management*, 23(5), 659-677. [https://doi.org/10.1016/S0149-2063\(97\)90020-9](https://doi.org/10.1016/S0149-2063(97)90020-9)
- Steiger, J.H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, 5(4), 411-419. <https://doi.org/10.1080/10705519809540115>

- Suárez-Alvarez, J., Pedrosa, I., Lozano, L.M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158. DOI <https://doi.org/10.7334/psicothema2018.33>
- Ten Berge, J.M.F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69, 613-625. <https://doi.org/10.1007/BF02289858>
- Thorndike, E.L. (1918). Individual differences. *Psychological Bulletin*, 15(5), 148-159. <https://doi.org/10.1037/h0070314>
- Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554. <https://doi.org/10.1086/214483>
- Tomás, J.M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling*, 6(1), 84-98. <https://doi.org/10.1080/10705519909540120>
- Tomás, J.M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling*, 20(2), 299-313. <https://doi.org/10.1080/10705511.2013.769394>
- Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511819322>
- Vermunt, J.K., & Magidson, J. (2002). Latent class cluster analysis. In J.A. Hagenaars & A.L. McCutcheon (Eds.), *Applied latent class analysis*. Cambridge, England: Cambridge University Press.
- Watts, A.L., Poore, H.E., Waldman, I.D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 7(6), 1285-1303. <https://doi.org/10.1177/2167702619855035>
- Weems, G.H., Onwuegbuzie, A.J., & Collins, K.M.T. (2006). The role of reading comprehension in responses to positively and negatively worded items on rating

- scales. *Evaluation & Research in Education*, 19(1), 3-20.
<https://doi.org/10.1080/09500790608668322>
- Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, 18, 320-334.
<https://doi.org/10.1037/a0032121>
- Widaman, K.F., Ferrer, E., & Conger, R.D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10–18. <http://dx.doi.org/10.1111/j.1750-8606.2009.00110.x>
- Wong, C.-S., & Law, K.S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *Leadership Quarterly*, 13(3), 243-274. [https://doi.org/10.1016/S1048-9843\(02\)00099-1](https://doi.org/10.1016/S1048-9843(02)00099-1)
- Wood, D., Harms, P.D., Lowman, G.H., & DeSimone, J.A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychology and Personality Science*, 8(4), 454-464.
<https://doi.org/10.1177/1948550617703168>
- Woods, C.M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 189-194. <https://doi.org/10.1007/s10862-005-9004-7>
- Wu, S.M., Schuler, T.A., Edwards, M.C., Yang, H.C., & Brothers, B.M. (2013). Factor analytic and item response theory evaluation of the Penn State Worry Questionnaire in women with cancer. *Quality of Life Research*, 22, 1441-1449.
<https://doi.org/10.1007/s11136-012-0253-0>
- Yuan, K.-H., & Bentler, P.M. (2000). Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika*, 65(1), 43-58.
<https://doi.org/10.1007/BF02294185>
- Zhang, C. & Conrad, F.G. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2), 127-135. <https://doi.org/10.18148/srm/2014.v8i2.5453>

- Zhang, X., Noor, R., Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the Need for Cognition Scale. *PLoS ONE*, 11(6), e0157795. <https://doi.org/10.1371/journal.pone.0157795>
- Zinbarg, R.E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω_H : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70, 123-133. <http://dx.doi.org/10.1007/s11336-003-0974-7>

Tabla 1. Estadísticos descriptivos por muestra.

Muestra	Escala	Ítems	Descriptivos					
			Media (DE)	Min-Máx	Asimetría	Curtosis		
Muestra 1 n=2964	RSES	Item1. Siento que soy una persona Digna de aprecio, al menos en igual medida que los demás.	5,836(1,215)	1-7	-1,272	1,752		
		Item2. Creo que tengo un buen número de cualidades.	5,868(0,977)	1-7	-1,025	1,716		
		Item3. En general me inclino a pensar que soy un fracasado/a.*	5,655(1,457)	1-7	-1,060	0,324		
		Item4. Soy capaz de hacer las cosas tan bien como la mayoría de la gente.	5,497(1,163)	1-7	-0,893	0,987		
		Item5. Siento que no tengo muchos motivos para sentirme orgulloso de mi.*	5,568(1,468)	1-7	-1,055	0,401		
		Item6. Tengo una actitud positiva hacia mismo/a.	5,313(1,233)	1-7	-0,747	0,357		
		Item7. En general, estoy satisfecho/a conmigo mismo/a.	5,455(1,144)	1-7	-0,975	1,155		
		Item8. Desearía valorarme más a mi mismo/a.*	4,873(1,749)	1-7	-0,459	-0,906		
		Item9. A veces me siento verdaderamente inútil.*	5,487(1,555)	1-7	-0,915	-0,110		
		Item10. A veces pienso que no soy bueno/a para nada.*	5,815(1,437)	1-7	-1,275	0,921		
	LOT-R	Item1. En tiempos difíciles, suelo esperar lo mejor.	3,274(0,789)	1-5	-0,279	-0,231		
		Item3. Si algo malo me tiene que pasar, estoy seguro de que me pasará.*	3,333(0,886)	1-5	-0,398	-0,314		
		Item4. Siempre soy optimista en cuanto al futuro.	3,482(0,820)	1-5	-0,450	-0,098		
		Item7. Rara vez, espero que las cosa salgan a mi manera.*	3,280(0,894)	1-5	-0,331	-0,356		
		Item9. Casi nunca cuento con que me sucedan cosas buenas. *	3,438(0,899)	1-5	-0,317	-0,283		
		Item10. En general, espero que me ocurran más cosas buenas que malas.	3,617(0,738)	1-5	-0,469	0,453		
		Muestra 2 n=2963	RSES	Item1. Siento que soy una persona Digna de aprecio, al menos en igual medida que los demás.	5,840(1,193)	1-7	-1,232	1,734
				Item2. Creo que tengo un buen número de cualidades.	5,877(0,963)	1-7	-0,862	1,147
				Item3. En general me inclino a pensar que soy un fracasado/a.*	5,677(1,450)	1-7	-1,063	0,334
				Item4. Soy capaz de hacer las cosas tan bien como la mayoría de la gente.	5,505(1,163)	1-7	-0,850	0,876
Item5. Siento que no tengo muchos motivos para sentirme orgulloso de mi.*	5,618(1,430)			1-7	-1,039	0,355		
Item6. Tengo una actitud positiva hacia mismo/a.	5,333(1,257)			1-7	-0,765	0,311		
Item7. En general, estoy satisfecho/a conmigo mismo/a.	5,451(1,157)			1-7	-0,931	1,053		
Item8. Desearía valorarme más a mi mismo/a.*	4,953(1,771)			1-7	-0,512	-0,891		
Item9. A veces me siento verdaderamente inútil.*	5,551(1,521)			1-7	-0,948	0,025		
Item10. A veces pienso que no soy bueno/a para nada.*	5,825(1,437)			1-7	-1,225	0,694		
LOT-R	Item1. En tiempos difíciles, suelo esperar lo mejor.	3,274(0,802)	1-5	-0,378	-0,110			
	Item3. Si algo malo me tiene que pasar, estoy seguro de que me pasará.*	3,361(0,862)	1-5	-0,318	-0,325			
	Item4. Siempre soy optimista en cuanto al futuro.	3,498(0,830)	1-5	-0,439	-0,181			
	Item7. Rara vez, espero que las cosa salgan a mi manera.*	3,294(0,902)	1-5	-0,248	-0,485			
	Item9. Casi nunca cuento con que me sucedan cosas buenas. *	3,445(0,911)	1-5	-0,257	-0,378			
	Item10. En general, espero que me ocurran más cosas buenas que malas.	3,593(0,740)	1-5	-0,465	0,442			

Nota: RSES=Escala de autoestima de Rosenberg; LOT-R=Escala de optimismo disposicional. Los ítems inversos están marcados con un asterisco.

Tabla 2. Índices de ajuste de las soluciones de perfiles latentes para las escalas RSES y LOT-R.

Escala	Muestra	Perfiles	Índices de ajuste				Entropía	m%
			AIC	BIC	VLMR	LMR		
RSES	Muestra 1 n=2964	2-perfiles	90617,128	90802,951	<0,001	<0,001	0,919	30,43%
		3-perfiles	88381,215	88632,975	<0,001	<0,001	0,853	18,15%
		4-perfiles	87194,551	87512,248	<0,001	<0,001	0,849	12,08%
		5-perfiles	86383,644	86767,279	0,002	0,002	0,876	2,97%
	Muestra 2 n=2963	2-perfiles	90055,433	90241,246	<0,001	<0,001	0,923	32,13%
		3-perfiles	87806,272	88058,018	<0,001	<0,001	0,852	18,70%
		4-perfiles	86595,277	86912,957	0,001	0,001	0,869	10,63%
		5-perfiles	85750,990	86134,603	0,305	0,310	0,871	3,48%
LOT-R	Muestra 1 n=2964	2-perfiles	40987,405	41281,126	<0,001	<0,001	0,715	48,79%
		3-perfiles	40057,158	40500,736	<0,001	<0,001	0,769	9,85%
		4-perfiles	39566,661	40160,096	<0,001	<0,001	0,736	9,04%
		5-perfiles	39392,478	40135,771	0,020	0,020	0,754	7,76%
		6-perfiles	39247,375	40140,525	0,051	0,052	0,746	2,19%
	Muestra 2 n=2963	2-perfiles	41107,184	41400,887	<0,001	<0,001	0,712	47,76%
		3-perfiles	40233,176	40676,729	<0,001	<0,001	0,753	12,60%
		4-perfiles	39720,129	40313,531	<0,001	<0,001	0,738	10,56%
		5-perfiles	39532,755	40276,005	0,067	0,069	0,723	10,29%
		6-perfiles	39366,506	40259,605	0,005	0,006	0,733	1,79%

Nota. Índices de ajuste: AIC=Criterio de Información de Akaïke; BIC=Criterio de Información Bayesiano; VLMR=Test de Vuong-Lo-Mendell-Rubin; LMR=Test de Lo-Mendell-Rubin. m% =Proporción de participantes de la clase con menor prevalencia.

Tabla 3. Probabilidad media de clasificación del perfil latente (PMC) para la pertenencia al perfil latente más probable (fila) por perfil identificado (columna) para las escalas RSES y LOT-R.

Escala	Muestra	Perfiles	PMC (RSES)				PMC (LOT-R)					Prevalencia	
			AE _A	AE _M	AE _B	AE _{MET}	OP _A	OP _{MA}	OP _M	OP _B	OP _{MET}	n	%
RSES	Muestra 1 n=2964	AE _A	0,930	0,069	0,000	0,001						1369	46,19
		AE _M	0,077	0,883	0,008	0,031						858	28,95
		AE _B	0,000	0,016	0,941	0,043						358	12,08
		AE _{MET}	0,004	0,067	0,044	0,884						379	12,79
	Muestra 2 n=2963	AE _A	0,942	0,058	0,000	0,001						1390	46,91
		AE _M	0,063	0,903	0,009	0,025						877	29,60
		AE _B	0,000	0,017	0,939	0,044						381	12,86
		AE _{MET}	0,003	0,051	0,041	0,904						315	10,63
LOT-R	Muestra 1 n=2964	OP _A					0,903	0,085	0,000	0,000	0,011	257	8,67
		OP _{MA}					0,028	0,882	0,062	0,008	0,019	1364	46,02
		OP _M					0,000	0,090	0,808	0,046	0,055	843	28,44
		OP _B					0,001	0,043	0,078	0,818	0,061	230	7,76
		OP _{MET}					0,018	0,064	0,106	0,044	0,768	270	9,11
	Muestra 2 n=2963	OP _A					0,917	0,060	0,003	0,005	0,015	305	10,29
		OP _{MA}					0,040	0,842	0,059	0,006	0,052	1167	39,39
		OP _M					0,001	0,066	0,823	0,048	0,062	939	31,69
		OP _B					0,005	0,010	0,101	0,858	0,025	227	7,66
		OP _{MET}					0,028	0,111	0,110	0,034	0,717	325	10,97

Nota. Escalas: RSES=Escala de autoestima de Rosenberg; LOT-R=Escala de optimismo disposicional. **Perfiles:** AE_A=Autoestima alta; AE_M=Autoestima intermedia; AE_B=Autoestima baja; AE_{MET}=Participantes con respuestas asimétricas en la RSES; OP_A=Optimismo alto; OP_{MA}=Optimismo medio-alto; OP_B=Optimismo bajo; OP_{MET}=Participantes con respuestas asimétricas en la LOT-R.

Tabla 4. Análisis Factorial Confirmatorio de las escalas RSES y LOT-R por condición.

Escala	Condición	Modelo	Índices de ajuste						Varianza explicada		
			χ^2	gl	CFI	TLI	RMSEA [IC 90%]	SRMR	FR	FI	FD
RSES	Muestra completa (MC) n=5927	M1. Factor único	1546,238***	33	0,909	0,875	0,088 [0,084-0,092]	0,062			
		M2. Bifactor FI	609,111***	28	0,965	0,944	0,059 [0,055-0,063]	0,033	0,813	0,187	
		M3. Bifactor FD	510,360***	28	0,971	0,953	0,054 [0,050-0,058]	0,027	0,796		0,204
		M4. Bifactor FID	246,291***	23	0,987	0,974	0,040 [0,036-0,045]	0,015	0,756	0,092	0,152
	MC-RAE n=5233	M1. Factor único	1422,259***	33	0,905	0,871	0,090 [0,086-0,094]	0,062			
		M2. Bifactor FI	535,501***	28	0,965	0,944	0,059 [0,055-0,063]	0,033	0,813	0,187	
		M3. Bifactor FD	504,951***	28	0,967	0,948	0,057 [0,053-0,061]	0,029	0,796		0,204
		M4. Bifactor FID	235,796***	23	0,985	0,972	0,042 [0,037-0,047]	0,016	0,756	0,096	0,149
	MC-MAE n=5233	M1. Factor único	955,366***	33	0,939	0,917	0,073 [0,069-0,077]	0,040			
		M2. Bifactor FI	484,829***	28	0,970	0,951	0,056 [0,052-0,060]	0,027	0,898	0,102	
		M3. Bifactor FD	440,416***	28	0,973	0,956	0,053 [0,049-0,057]	0,023	0,873		0,127
		M4. Bifactor FID					No Converge				
	MC-ROP n=5332	M1. Factor único	1443,958***	33	0,906	0,872	0,090 [0,086-0,094]	0,062			
		M2. Bifactor FI	554,029***	28	0,965	0,944	0,059 [0,055-0,064]	0,033	0,813	0,187	
		M3. Bifactor FD	499,150***	28	0,969	0,949	0,056 [0,052-0,061]	0,028	0,796		0,204
		M4. Bifactor FID	227,838***	23	0,986	0,973	0,041 [0,036-0,046]	0,016	0,754	0,097	0,149
MC-MOP n=5332	M1. Factor único	1374,794***	33	0,913	0,881	0,087 [0,083-0,091]	0,059				
	M2. Bifactor FI	564,955***	28	0,965	0,944	0,060 [0,056-0,064]	0,033	0,827	0,173		
	M3. Bifactor FD	501,331***	28	0,969	0,951	0,056 [0,052-0,061]	0,029	0,809		0,191	
	M4. Bifactor FID	218,258***	23	0,987	0,975	0,040 [0,035-0,045]	0,016	0,768	0,095	0,137	

Nota. Índices de ajuste: CFI=Índice de ajuste comparativo; TLI=Índice de Tucker-Lewis; RMSEA=Error cuadrático medio de aproximación; SRMR=Raíz cuadrada media estandarizada residual. **Varianza explicada:** FR=Factor asociado al rasgo; FI=Factor asociado a los ítems inversos; FD=Factor asociado a los ítems directos.

Tabla 4. Análisis Factorial Confirmatorio de las escalas RSES y LOT-R por condición (continuación)

Escala	Condición	Modelo	Índices de ajuste							Varianza explicada		
			χ^2	gl	CFI	TLI	RMSEA [IC 90%]	SRMR	FR	FI	FD	
LOT-R	Muestra completa (MC) n=5927	M1. Factor único	906,381***	9	0,802	0,670	0,130 [0,123-0,137]	0,068				
		M2. Bifactor FI	195,393***	6	0,958	0,896	0,073 [0,064-0,082]	0,032	0,638	0,362		
		M3. Bifactor FD	53,307***	6	0,990	0,974	0,036 [0,028-0,046]	0,014	0,694		0,306	
MC-RAE n=5233		M1. Factor único	821,812***	9	0,796	0,660	0,131 [0,124-0,139]	0,069				
		M2. Bifactor FI	176,872***	6	0,957	0,893	0,074 [0,065-0,083]	0,032	0,634	0,366		
		M3. Bifactor FD	49,118***	6	0,989	0,973	0,037 [0,028-0,047]	0,015	0,692		0,308	
MC-MAE n=5233		M1. Factor único	788,762***	9	0,805	0,675	0,129 [0,121-0,136]	0,067				
		M2. Bifactor FI	172,133***	6	0,958	0,896	0,073 [0,064-0,082]	0,031	0,647	0,353		
		M3. Bifactor FD	49,331***	6	0,989	0,973	0,037 [0,028-0,047]	0,015	0,695		0,305	
MC-ROp n=5332		M1. Factor único	843,506***	9	0,795	0,659	0,132 [0,124-0,139]	0,069				
		M2. Bifactor FI	181,109***	6	0,957	0,893	0,074 [0,065-0,083]	0,032	0,632	0,368		
		M3. Bifactor FD	52,512***	6	0,989	0,971	0,038 [0,029-0,048]	0,015	0,691		0,309	
MC-MOp n=5332		M1. Factor único	275,060***	9	0,939	0,899	0,074 [0,067-0,082]	0,036				
		M2. Bifactor FI	114,484***	6	0,975	0,938	0,058 [0,049-0,068]	0,022	0,820	0,180		
		M3. Bifactor FD	31,423***	6	0,994	0,986	0,028 [0,019-0,038]	0,011	0,808		0,192	

Nota. Índices de ajuste: CFI=Índice de ajuste comparativo; TLI=Índice de Tucker-Lewis; RMSEA=Error cuadrático medio de aproximación; SRMR=Raíz cuadrada media estandarizada residual. **Varianza explicada:** FR=Factor asociado al rasgo; FI=Factor asociado a los ítems inversos; FD=Factor asociado a los ítems directos.

Tabla 5. Índices de ajuste de los modelos multirasgo-multimétodo.

Condición	Modelo	Índices de ajuste							Varianza explicada					Correlaciones				
		χ^2	gl	CFI	TLI	RMSEA	SRMR	RSES			LOT-R		RR	DD	ID			
								FR _{AE}	FD _{AE}	FI _{AE}	FR _{OP}	FD _{OP}						
MC n=5927	M5. Factor único	3313,305***	101	0,882	0,860	0,073	0,060											
	M6. Bifactor FID	1103,799***	88	0,963	0,949	0,044	0,051	0,756	0,150	0,094	0,726	0,274	0,574***					
	M7. Bifactor FIDc	1044,015***	86	0,965	0,951	0,043	0,043	0,750	0,166	0,085	0,718	0,282	0,552***	0,279***	0,140***			
MC-R _{AE} n=5266	M5. Factor único	2940,484***	101	0,882	0,860	0,073	0,059						0,581***					
	M6. Bifactor FID	969,993***	88	0,963	0,950	0,044	0,051	0,756	0,147	0,097	0,728	0,272	0,567***					
	M7. Bifactor FIDc	907,908***	86	0,966	0,952	0,043	0,042	0,749	0,163	0,088	0,719	0,281	0,555***	0,292***	0,159***			
MC-M _{AE} n=5266	M5. Factor único	2395,398***	101	0,909	0,891	0,066	0,049						0,582***					
	M6. Bifactor FID							No Converge										
	M7. Bifactor FIDc							No Converge										
MC-R _{OP} n=5442	M5. Factor único	3059,191***	101	0,881	0,859	0,073	0,060						0,578***					
	M6. Bifactor FID	1007,429***	88	0,963	0,950	0,044	0,051	0,754	0,150	0,096	0,727	0,273	0,564***					
	M7. Bifactor FIDc	942,595***	86	0,966	0,952	0,043	0,042	0,747	0,166	0,086	0,719	0,281	0,552***	0,294***	0,154***			
MC-M _{OP} n=5442	M5. Factor único	2376,955***	101	0,914	0,897	0,064	0,047						0,609***					
	M6. Bifactor FID	720,399***	88	0,976	0,967	0,036	0,027	0,762	0,124	0,114	0,861	0,139	0,647***					
	M7. Bifactor FIDc	743,062***	86	0,975	0,965	0,037	0,027	0,763	0,124	0,113	0,862	0,138	0,648***	-0,005	-0,029			
MC-R _{2EA} n=5826	M5. Factor único	3263,900***	101	0,882	0,859	0,073	0,060						0,575***					
	M6. Bifactor FID	1082,695***	88	0,963	0,949	0,044	0,051	0,755	0,150	0,095	0,726	0,274	0,560***					
	M7. Bifactor FIDc	1023,217***	86	0,965	0,951	0,043	0,043	0,748	0,166	0,086	0,718	0,282	0,548***	0,282***	0,142***			
MC-M _{2EA} n=5826	M5. Factor único	3093,492***	101	0,890	0,870	0,071	0,056						0,573***					
	M6. Bifactor FID	1067,925***	88	0,964	0,951	0,044	0,049	0,766	0,132	0,102	0,750	0,250	0,559***					
	M7. Bifactor FIDc	1033,570***	86	0,965	0,952	0,043	0,044	0,760	0,143	0,097	0,742	0,258	0,549***	0,183***	0,181***			
MC-R _{2EV} n=4882	M5. Factor único	2694,542***	101	0,884	0,862	0,073	0,060						0,556***					
	M6. Bifactor FID	946,267***	88	0,962	0,948	0,045	0,051	0,759	0,149	0,093	0,726	0,274	0,543***					
	M7. Bifactor FIDc	891,218***	86	0,964	0,950	0,044	0,042	0,754	0,166	0,081	0,719	0,281	0,529***	0,330***	0,138**			
MC-M _{2EV} n=4882	M5. Factor único	1676,940***	101	0,935	0,923	0,057	0,036						0,619***					
	M6. Bifactor FID							No Converge										
	M7. Bifactor FIDc							No Converge										

Nota. Índices de ajuste: CFI=Índice de ajuste comparativo; TLI=Índice de Tucker-Lewis; RMSEA=Error cuadrático medio de aproximación; SRMR=Raíz cuadrada media estandarizada residual. **Varianza explicada:** FR=Factor asociado al rasgo; FD=Factor asociado a los ítems directos; FI=Factor asociado a los ítems inversos. **Correlaciones:** RR=Correlación entre los factores de rasgo (FR_{AE} y FR_{OP}); DD=Correlación entre los factores de ítems directos (FD; RSES) e ítems directos (FD; LOT-R); ID=Correlación entre los factores de ítems inversos (FI; RSES) e ítems directos (FD; LOT-R).

* p < 0.05; ** p < 0.01; *** p < 0.001

Tabla 6. Probabilidad media de clasificación (PCM), frecuencia y proporción de participantes (FPP) por patrón de perfil latente identificado.

RSES		LOT-R				
		OP _A (620; 10,5%)	OP _{MA} (2624; 44,3%)	OP _M (1630; 27,5%)	OP _B (568; 9,6%)	OP _{MET} (485; 8,2%)
PMC	AE _A (2960; 49,9%)	0,884	0,824	0,703	0,747	0,704
	AE _M (1565; 26,4%)	0,723	0,746	0,714	0,691	0,641
	AE _B (741; 12,5%)	0,858	0,731	0,799	0,828	0,711
	AE _{MET} (661; 11,2%)	0,774	0,739	0,682	0,714	0,728
FPP	AE _A (2960; 49,9%)	554 (89,3%)	1686 (64,2%)	423 (26,0%)	53 (9,3%)	244 (50,3%)
	AE _M (1565; 26,4%)	39 (6,3%)	664 (25,3%)	643 (39,4%)	105 (18,5%)	114 (23,5%)
	AE _B (741; 12,5%)	8 (1,3%)	94 (3,6%)	329 (20,2%)	284 (50,0%)	26 (5,4%)
	AE _{MET} (661; 11,2%)	19 (3,1%)	180 (6,9%)	235 (14,4%)	126 (22,2%)	101 (20,8%)

Nota. FPP: Los porcentajes de cada patrón de perfiles fueron calculados por columna.

Tabla 7. Estadísticos descriptivos por ola.

Ola	Ítems	Descriptivos				ANOVA medidas repetidas		
		Media (DE)	Min-Máx	Asimetría	Curtosis	F _(2,7422)	η^2 parcial	Diferencias
Ola 1	Item1	5,996(1,106)	1-7	-1,300	1,945	26,189***	0,007	O ₁ =O ₂ >O ₃
	Item2	6,013(0,915)	1-7	-1,088	2,184	34,588***	0,009	O ₁ =O ₂ >O ₃
	Item3*	5,760(1,478)	1-7	-1,220	0,678	0,699	<0,001	O ₁ =O ₂ =O ₃
	Item4	5,675(1,132)	1-7	-1,029	1,310	29,065***	0,008	O ₁ =O ₂ >O ₃
	Item5*	5,665(1,528)	1-7	-1,172	0,465	4,485*	0,001	O ₁ =O ₂ >O ₃
	Item6	5,473(1,243)	1-7	-0,927	0,678	25,051***	0,007	O ₁ >O ₂ >O ₃
	Item7	5,607(1,124)	1-7	-1,156	1,832	46,753***	0,012	O ₁ =O ₂ >O ₃
	Item8*	4,985(1,844)	1-7	-0,587	-0,903	0,486	<0,001	O ₁ =O ₂ =O ₃
	Item9*	5,599(1,603)	1-7	-1,122	0,321	1,567	<0,001	O ₁ =O ₂ =O ₃
	Item10*	5,926(1,479)	1-7	-1,519	1,566	4,470*	0,001	O ₁ =O ₂ >O ₃
Ola 2	Item1	5,986(1,099)	1-7	-1,330	2,080			
	Item2	5,987(0,910)	1-7	-1,037	2,027			
	Item3*	5,742(1,453)	1-7	-1,199	0,687			
	Item4	5,630(1,118)	1-7	-1,019	1,457			
	Item5*	5,668(1,449)	1-7	-1,141	0,531			
	Item6	5,420(1,250)	1-7	-0,924	0,772			
	Item7	5,575(1,112)	1-7	-1,072	1,457			
	Item8*	4,964(1,795)	1-7	-0,553	-0,859			
	Item9*	5,567(1,553)	1-7	-0,998	0,041			
	Item10*	5,887(1,429)	1-7	-1,370	1,158			
Ola 3	Item1	5,876(1,167)	1-7	-1,301	2,001			
	Item2	5,892(0,941)	1-7	-0,873	1,253			
	Item3*	5,733(1,421)	1-7	-1,171	0,677			
	Item4	5,526(1,141)	1-7	-0,832	0,849			
	Item5*	5,601(1,436)	1-7	-1,052	0,378			
	Item6	5,338(1,227)	1-7	-0,772	0,427			
	Item7	5,449(1,143)	1-7	-0,993	1,222			
	Item8*	4,958(1,741)	1-7	-0,504	-0,872			
	Item9*	5,554(1,523)	1-7	-0,968	0,058			
	Item10*	5,855(1,410)	1-7	-1,281	0,906			

Nota. Los ítems inversos están marcados con asterisco.

* p <0,05; ** p<0,01; *** p<0,001

Tabla 8. Índices de ajuste de los modelos confirmatorios por ola y longitudinal.

Condición	Modelos	Índices de ajuste					
		χ^2	gl	CFI	TLI	RMSEA [CI _{90%}]	SRMR
CFA Ola 1	M1. Factor único	1171,197***	33	0,886	0,845	0,096 [0,092-0,101]	0,064
	M2. Bifactor FI	488,311***	28	0,954	0,926	0,067 [0,061-0,072]	0,036
	M3. Bifactor FID	187,258***	23	0,984	0,968	0,044 [0,038-0,050]	0,019
CFA Ola 2	M1. Factor único	986,536***	33	0,905	0,870	0,088 [0,084-0,093]	0,062
	M2. Bifactor FI	318,664***	28	0,971	0,953	0,053 [0,048-0,058]	0,032
	M3. Bifactor FID	106,829***	23	0,992	0,984	0,031 [0,025-0,037]	0,013
CFA Ola 3	M1. Factor único	981,605***	33	0,913	0,882	0,088 [0,083-0,093]	0,057
	M2. Bifactor FI	446,402***	28	0,962	0,938	0,063 [0,058-0,069]	0,034
	M3. Bifactor FID	160,209***	23	0,987	0,975	0,040 [0,034-0,046]	0,016
SEM Ola 1 - Ola 3	M8. Factor único	4391,608***	366	0,912	0,896	0,054 [0,053-0,055]	0,054
	M9. Bifactor FID	1922,429***	336	0,965	0,955	0,036 [0,034-0,037]	0,039
	M10. Bifactor FIDc	971,678***	330	0,986	0,982	0,023 [0,021-0,025]	0,022

Nota. Índices de ajuste: CFI=Índice de ajuste comparativo; TLI=Índice de Tucker-Lewis; RMSEA=Error cuadrático medio de aproximación; SRMR=Raíz cuadrada media estandarizada residual.

* p <0,05; ** p<0,01; *** p<0,001

Tabla 9. Resumen de los modelos de invarianza estimados.

Modelo	Índices de ajuste absoluto					Índices de ajuste relativo				Diferencias			
	χ^2	gl	$\Delta\chi^2$	pnc	Δ gl	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Bifactor-FID													
Configural	971,678***	330				0,986	0,982	0,023	0,022				
Métrica	1008,301***	364	36,623	40,618	34	0,986	0,983	0,022	0,024	<0,001	0,001	0,001	0,002
Escalar	1053,519***	378	45,218	29,423	14	0,985	0,983	0,022	0,024	0,001	<0,001	<0,001	<0,001
Residual	1120,673***	398	67,154	34,129	20	0,984	0,983	0,022	0,029	0,001	<0,001	<0,001	0,005
Varianza factores	1150,061***	404	29,388	26,081	6	0,984	0,983	0,022	0,033	<0,001	<0,001	<0,001	0,004
Covarianza factores	1156,875***	407	6,814	24,452	3	0,984	0,983	0,022	0,033	<0,001	<0,001	<0,001	<0,001
Bifactor-FI													
Configural	1985,839***	348				0,964	0,955	0,036	0,037				
Métrica	2019,411***	374	33,572	36,385	26	0,964	0,958	0,034	0,038	<0,001	0,003	0,002	0,001
Escalar	2081,498***	390	62,087	31,289	16	0,963	0,959	0,034	0,039	0,001	0,001	<0,001	0,001
Residual	2098,901***	410	17,403	34,797	20	0,963	0,961	0,033	0,042	0,001	0,002	0,001	0,003
Varianza factores	2125,638***	414	26,737	25,440	4	0,963	0,961	0,033	0,045	<0,001	<0,001	<0,001	0,003
Covarianza factores	2131,619***	416	5,981	24,354	2	0,963	0,961	0,033	0,045	<0,001	<0,001	<0,001	<0,001

Nota. **Índices de ajuste absoluto:** χ^2 =chi-cuadrado; gl=grados de libertad; pnc=parámetro de no-centralidad. **Índices de ajuste relativo:** CFI=Índice de ajuste comparativo; TLI=Índice de Tucker-Lewis; RMSEA=Error cuadrático medio de aproximación; SRMR=Raíz cuadrada media estandarizada residual.

* p <0.05; ** p<0.01; *** p<0.001

Tabla 10. Índices de ajuste de las soluciones de perfiles latentes para la escala RSES por ola.

Ola	Perfiles	Índices de ajuste				Entropía	m%
		AIC	BIC	VLMR	LMR		
Ola 1	2-perfiles	113903,891	114096,690	<0,001	<0,001	0,924	28,18%
	3-perfiles	111294,578	111555,790	0,002	0,002	0,852	14,76%
	4-perfiles	109689,684	110019,308	<0,001	<0,001	0,886	7,33%
	5-perfiles	108394,950	108792,986	0,005	0,005	0,869	4,07%
Ola 2	2-perfiles	112194,088	112386,887	<0,001	<0,001	0,917	30,47%
	3-perfiles	109547,164	109808,376	<0,001	<0,001	0,863	15,19%
	4-perfiles	108053,540	108383,164	<0,001	<0,001	0,867	10,13%
	5-perfiles	106980,569	107378,606	0,124	0,127	0,864	3,99%
Ola 3	2-perfiles	111423,192	111615,991	<0,001	<0,001	0,924	30,55%
	3-perfiles	108443,363	108704,574	0,006	0,006	0,853	17,73%
	4-perfiles	107024,026	107353,650	0,027	0,028	0,848	11,34%
	5-perfiles	105621,449	106019,486	0,258	0,262	0,878	3,5%

Nota. Índices de ajuste: AIC=Criterio de Información de Akaike; BIC=Criterio de Información Bayesiano; VLMR=Test de Vuong-Lo-Mendell-Rubin; LMR=Test de Lo-Mendell-Rubin. m%=Proporción de participantes de la clase con menor prevalencia.

Tabla 11. Probabilidad media de clasificación del perfil latente (PMC) para la pertenencia al perfil latente más probable por perfil identificado para la escala RSES por ola.

Ola	Perfiles	Perfiles latentes				Prevalencia	
		AE _A	AE _M	AE _B	AE _{MET}	n	%
Ola 1	AE _A	0,955	0,045	<0,001	0,001	1981	53,37%
	AE _M	0,071	0,900	0,010	0,019	1074	28,93%
	AE _B	<0,001	0,032	0,937	0,031	385	10,37%
	AE _{MET}	0,005	0,047	0,027	0,921	272	7,33%
Ola 2	AE _A	0,941	0,058	<0,001	0,001	1872	50,43%
	AE _M	0,067	0,895	0,010	0,027	1020	27,48%
	AE _B	<0,001	0,023	0,945	0,032	444	11,96%
	AE _{MET}	0,007	0,060	0,038	0,895	376	10,13%
Ola 3	AE _A	0,936	0,063	<0,001	0,001	1658	44,67%
	AE _M	0,086	0,875	0,004	0,035	1145	30,85%
	AE _B	<0,001	0,013	0,935	0,052	421	11,34%
	AE _{MET}	0,002	0,072	0,045	0,882	488	13,15%

Tabla 12. Índices de ajuste de las soluciones de transiciones latentes estimadas.

Modelo	Loglikelihood			Índices de ajuste		
	LogL	FC	PL	BIC	adjBIC	Entropía
LTA Libre	-159961,136	1,887	177	321377,093	320814,673	0,882
LTA Invariante	-156965,775	2,298	103	314778,141	314450,857	0,885
LTA Invariante y estacional	-160203,243	2,651	79	321055,812	320804,788	0,886

Nota. Loglikelihood: LogL=Loglikelihood; FC=Factor de corrección; PL=Parámetros libres. **Índices de ajuste:** BIC=Criterio de Información Bayesiano; adjBIC=Criterio de Información Bayesiano ajustado al tamaño muestral.

* p <0.05; ** p<0.01; *** p<0.001

Tabla 13. Probabilidad media de clasificación de los patrones de tres olas de transición (PMC), probabilidades de transición entre pares de olas (τ), y frecuencia en cada patrón de perfil identificado.

PMC	Ola 1	Ola 2	Ola 3			
			AE _A	AE _M	AE _B	AE _{MET}
AE _A		AE _A	0,925	0,723	0,871	0,795
		AE _M	0,703	0,677	0,741	0,481
		AE _B	0,800	0,716	0,806	0,972
		AE _{MET}	0,783	0,731	0,756	0,714
AE _M		AE _A	0,717	0,605	0,731	0,744
		AE _M	0,675	0,802	0,725	0,700
		AE _B	0,674	0,692	0,770	0,630
		AE _{MET}	0,736	0,700	0,597	0,654
AE _B		AE _A	0,932	0,603	0,000 ^a	0,000 ^a
		AE _M	0,641	0,725	0,704	0,592
		AE _B	0,828	0,720	0,923	0,622
		AE _{MET}	0,932	0,703	0,670	0,719
AE _{MET}		AE _A	0,818	0,684	0,658	0,715
		AE _M	0,586	0,717	0,724	0,615
		AE _B	0,611	0,593	0,864	0,552
		AE _{MET}	0,740	0,700	0,771	0,760

τ	Pares	Ola t	Ola t+1			
			AE _A	AE _M	AE _B	AE _{MET}
O ₁ -O ₂		AE _A	0,842	0,108	0,011	0,039
		AE _M	0,149	0,655	0,096	0,100
		AE _B	0,029	0,184	0,666	0,121
		AE _{MET}	0,169	0,186	0,112	0,533
O ₂ -O ₃		AE _A	0,836	0,109	0,020	0,035
		AE _M	0,149	0,680	0,092	0,079
		AE _B	0,042	0,181	0,709	0,067
		AE _{MET}	0,145	0,249	0,199	0,407

Frecuencia	Ola 1	Ola 2	Ola 3			
			AE _A (n=1805)	AE _M (n=1059)	AE _B (n=508)	AE _{MET} (n=340)
AE _A (n=1983)		AE _A (n=1884)	1478	144	28	46 ^d
		AE _M (n=987)	42	131	9	8 ^d
		AE _B (n=426)	2	4	14	1 ^d
		AE _{MET} (n=415)	24 ^c	21 ^c	6 ^c	25
AE _M (n=954)		AE _A (n=1884)	79	23	4	9 ^d
		AE _M (n=987)	67	492	59	44 ^d
		AE _B (n=426)	8	37	45	7 ^d
		AE _{MET} (n=415)	9 ^c	24 ^c	14 ^c	33
AE _B (n=390)		AE _A (n=1884)	8	1	0	0 ^d
		AE _M (n=987)	8	43	20	4 ^d
		AE _B (n=426)	6	36	216	10 ^d
		AE _{MET} (n=415)	2 ^c	8 ^c	14 ^c	14
AE _{MET} (n=385)		AE _A (n=1884)	43 ^b	9 ^b	3 ^b	9
		AE _M (n=987)	3 ^b	41 ^b	5 ^b	11
		AE _B (n=426)	1 ^b	4 ^b	27 ^b	8
		AE _{MET} (n=415)	25	41	44	111

^a Patrones de transiciones con una prevalencia igual a 0%.

^b Individuos asignados al perfil AE_{MET} sólo en la Ola 1.

^c Individuos asignados al perfil AE_{MET} sólo en la Ola 2.

^d Individuos asignados al perfil AE_{MET} sólo en la Ola 3.

Tabla 14. Índices de ajuste y varianza explicada por factor de los modelos longitudinales estimados por condición.

Condición	Modelo	Índices de ajuste						Varianza explicada								
		χ^2 (gl)	gl	CFI	TLI	RMSEA	SRMR	Ola 1			Ola 2			Ola 3		
								FR	FD	FI	FR	FD	FI	FR	FD	FI
MC-R ₀₁ n=3327	M8. 1-Factor	4036,485***	366	0,912	0,895	0,055	0,053									
	M9. Bifactor FID	1762,603***	336	0,966	0,956	0,036	0,039	0,777	0,117	0,105	0,795	0,109	0,096	0,811	0,107	0,082
	M10. Bifactor FIDc	945,575***	330	0,985	0,980	0,024	0,022	0,722	0,140	0,138	0,734	0,137	0,129	0,749	0,134	0,116
MC-M ₀₁ n=3327	M8. 1-Factor	3409,543***	366	0,927	0,914	0,050	0,044									
	M9. Bifactor FID	1576,544***	336	0,970	0,962	0,033	0,032	0,849	0,084	0,067	0,816	0,083	0,101	0,821	0,088	0,091
	M10. Bifactor FIDc	949,800***	330	0,985	0,980	0,024	0,021	0,813	0,102	0,085	0,767	0,115	0,119	0,771	0,118	0,111
MC-R ₀₂ n=3297	M8. 1-Factor	3918,694***	366	0,913	0,897	0,054	0,053									
	M9. Bifactor FID	1740,719***	336	0,966	0,955	0,036	0,039	0,780	0,116	0,104	0,796	0,106	0,098	0,808	0,107	0,085
	M10. Bifactor FIDc	934,656***	330	0,985	0,980	0,024	0,023	0,726	0,133	0,141	0,735	0,129	0,136	0,744	0,134	0,122
MC-M ₀₂ n=3297	M8. 1-Factor	3174,095***	366	0,932	0,919	0,048	0,043									
	M9. Bifactor FID	1437,263***	336	0,973	0,965	0,032	0,030	0,801	0,098	0,101	0,860	0,074	0,066	0,834	0,083	0,083
	M10. Bifactor FIDc	855,837***	330	0,987	0,983	0,022	0,019	0,767	0,113	0,120	0,814	0,104	0,082	0,784	0,115	0,101
MC-R ₀₃ n=3372	M8. 1-Factor	3996,918***	366	0,913	0,897	0,054	0,053									
	M9. Bifactor FID	1748,160***	336	0,966	0,956	0,035	0,039	0,777	0,116	0,107	0,791	0,106	0,103	0,802	0,108	0,091
	M10. Bifactor FIDc	881,301***	330	0,987	0,983	0,022	0,022	0,718	0,135	0,147	0,728	0,132	0,140	0,745	0,132	0,123
MC-M ₀₃ n=3372	M8. 1-Factor	3423,971***	366	0,928	0,914	0,050	0,046									
	M9. Bifactor FID	1723,575***	336	0,967	0,958	0,035	0,034	0,796	0,104	0,099	0,818	0,086	0,097	0,854	0,076	0,070
	M10. Bifactor FIDc	937,811***	330	0,986	0,981	0,023	0,021	0,753	0,122	0,125	0,766	0,113	0,121	0,823	0,101	0,076
MC-R _{30A} n=3601	M8. 1-Factor	4259,229***	366	0,913	0,897	0,054	0,053									
	M9. Bifactor FID	1855,843***	336	0,966	0,956	0,035	0,039	0,775	0,121	0,104	0,791	0,109	0,100	0,814	0,105	0,081
	M10. Bifactor FIDc	949,793***	330	0,986	0,982	0,023	0,022	0,719	0,140	0,141	0,729	0,134	0,137	0,750	0,131	0,119
MC-M _{30A} n=3601	M8. 1-Factor	3951,664***	366	0,919	0,904	0,052	0,049									
	M9. Bifactor FID	1644,247***	336	0,970	0,962	0,033	0,035	0,784	0,109	0,107	0,804	0,094	0,102	0,822	0,091	0,087
	M10. Bifactor FIDc	930,928***	330	0,986	0,982	0,022	0,021	0,741	0,133	0,126	0,751	0,128	0,121	0,770	0,124	0,106
MC-R _{30V} n=3004	M8. 1-Factor	3718,061***	366	0,910	0,893	0,055	0,054									
	M9. Bifactor FID	1623,251***	336	0,965	0,955	0,036	0,040	0,775	0,115	0,110	0,786	0,113	0,101	0,798	0,111	0,091
	M10. Bifactor FIDc	849,582***	330	0,986	0,982	0,023	0,022	0,720	0,137	0,143	0,726	0,138	0,136	0,734	0,141	0,125
MC-M _{30V} n=3004	M8. 1-Factor	2573,972***	366	0,944	0,933	0,045	0,037									
	M9. Bifactor FID	1421,049***	336	0,972	0,964	0,033	0,029	0,856	0,082	0,062	0,868	0,068	0,064	0,875	0,072	0,053
	M10. Bifactor FIDc	879,063***	330	0,986	0,982	0,024	0,020	0,822	0,094	0,084	0,823	0,092	0,085	0,829	0,096	0,075

Nota. Índices de ajuste: CFI=Índice de ajuste comparativo; TLI=Índice de Tucker-Lewis; RMSEA=Error cuadrático medio de aproximación; SRMR=Raíz cuadrada media estandarizada residual. **Varianza explicada:** FR=Factor asociado al rasgo; FD=Factor asociado a los ítems directos; FI=Factor asociado a los ítems inversos.

* p <0.05; ** p<0.01; *** p<0.001

Tabla 15. Correlaciones estimadas entre los factores especificados por modelo (M8-M10) en las condiciones muestrales examinadas

Condición	Modelo	FR			FD			FI		
		O ₁ O ₂	O ₁ O ₃	O ₂ O ₃	O ₁ O ₂	O ₁ O ₃	O ₂ O ₃	O ₁ O ₂	O ₁ O ₃	O ₂ O ₃
MC-R _{O1} n=3327	M8. 1-Factor	0,793***	0,727***	0,768***						
	M9. Bifactor FID	0,859***	0,777***	0,822***						
	M10. Bifactor FIDc	0,833***	0,755***	0,793***	0,599***	0,504***	0,584***	0,436***	0,412***	0,467***
MC-M _{O1} n=3327	M8. 1-Factor	0,797***	0,726***	0,769***						
	M9. Bifactor FID	0,843***	0,765***	0,821***						
	M10. Bifactor FIDc	0,824***	0,750***	0,796***	0,619***	0,524***	0,589***	0,335***	0,324***	0,427***
MC-R _{O2} n=3297	M8. 1-Factor	0,787***	0,726***	0,764***						
	M9. Bifactor FID	0,853***	0,775***	0,817***						
	M10. Bifactor FIDc	0,825***	0,753***	0,787***	0,610***	0,496***	0,591***	0,448***	0,407***	0,481***
MC-M _{O2} n=3297	M8. 1-Factor	0,797***	0,734***	0,769***						
	M9. Bifactor FID	0,846***	0,778***	0,812***						
	M10. Bifactor FIDc	0,827***	0,759***	0,793***	0,577***	0,503***	0,604***	0,362***	0,361***	0,368***
MC-R _{O3} n=3372	M8. 1-Factor	0,790***	0,733***	0,770***						
	M9. Bifactor FID	0,860***	0,783***	0,825***						
	M10. Bifactor FIDc	0,828***	0,757***	0,793***	0,609***	0,524***	0,612***	0,461***	0,430***	0,476***
MC-M _{O3} n=3372	M8. 1-Factor	0,794***	0,722***	0,768***						
	M9. Bifactor FID	0,855***	0,758***	0,806***						
	M10. Bifactor FIDc	0,827***	0,745***	0,788***	0,622***	0,502***	0,607***	0,423***	0,369***	0,396***
MC-R _{30A} n=3601	M8. 1-Factor	0,787***	0,723***	0,763***						
	M9. Bifactor FID	0,855***	0,773***	0,817***						
	M10. Bifactor FIDc	0,825***	0,749***	0,786***	0,606***	0,510***	0,589***	0,453***	0,421***	0,482***
MC-M _{30A} n=3601	M8. 1-Factor	0,784***	0,716***	0,756***						
	M9. Bifactor FID	0,849***	0,765***	0,810***						
	M10. Bifactor FIDc	0,824***	0,747***	0,784***	0,605***	0,499***	0,593***	0,347***	0,303***	0,379***
MC-R _{30V} n=3004	M8. 1-Factor	0,783***	0,719***	0,759***						
	M9. Bifactor FID	0,849***	0,772***	0,815***						
	M10. Bifactor FIDc	0,822***	0,747***	0,785***	0,579***	0,516***	0,580***	0,431***	0,410***	0,487***
MC-M _{30V} n=3004	M8. 1-Factor	0,815***	0,743***	0,786***						
	M9. Bifactor FID	0,849***	0,771***	0,819***						
	M10. Bifactor FIDc	0,832***	0,758***	0,802***	0,621***	0,524***	0,609***	0,476***	0,424***	0,446***

Nota. FR=Factor asociado al rasgo; FD=Factor asociado a los ítems directos; FI=Factor asociado a los ítems inversos.

* p <0.05; ** p<0.01; *** p<0.001

Figura 1. Medias estimadas de los ítems directos e inversos de la RSES y LOT-R por perfiles identificados en ambas muestras.

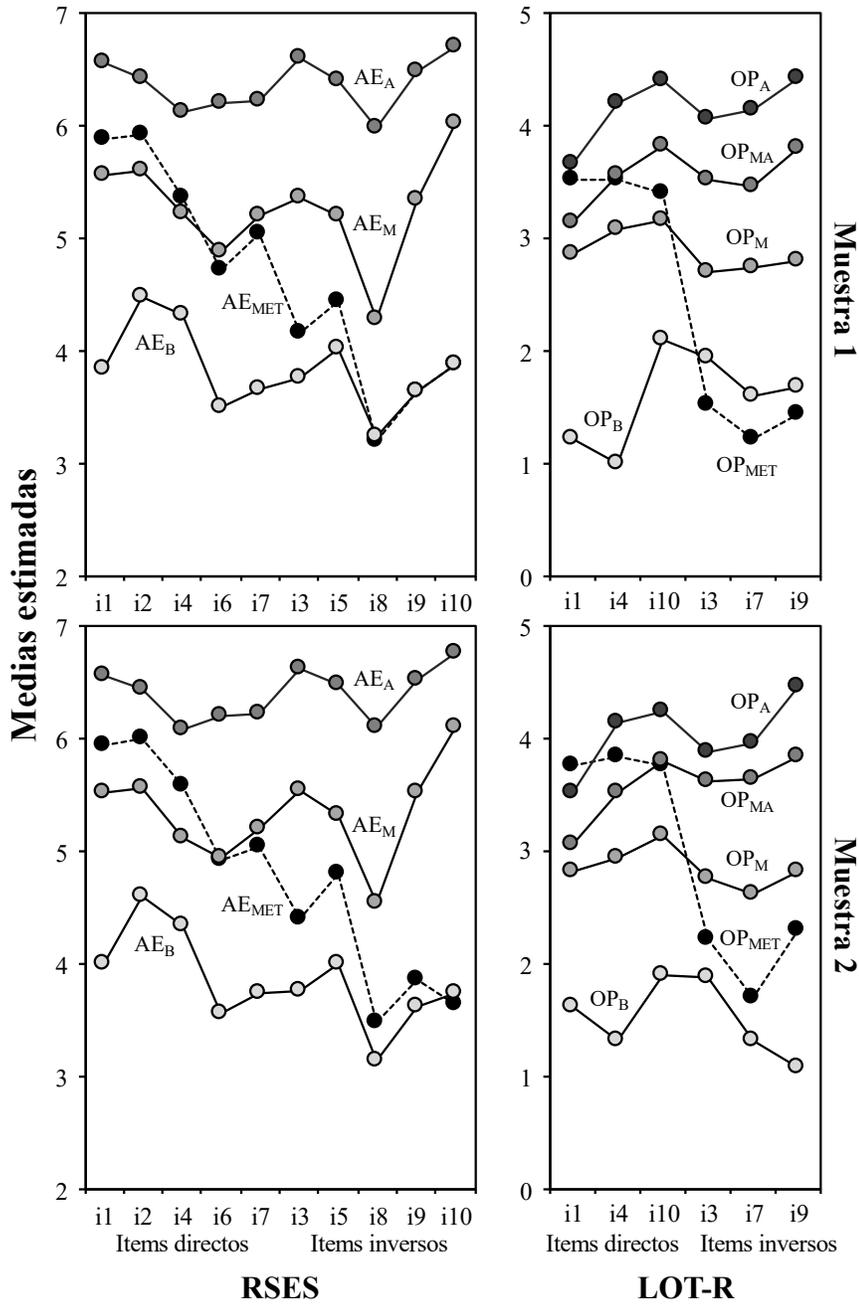


Figura 2. Modelos confirmatorios para las escalas RSES y LOT-R.

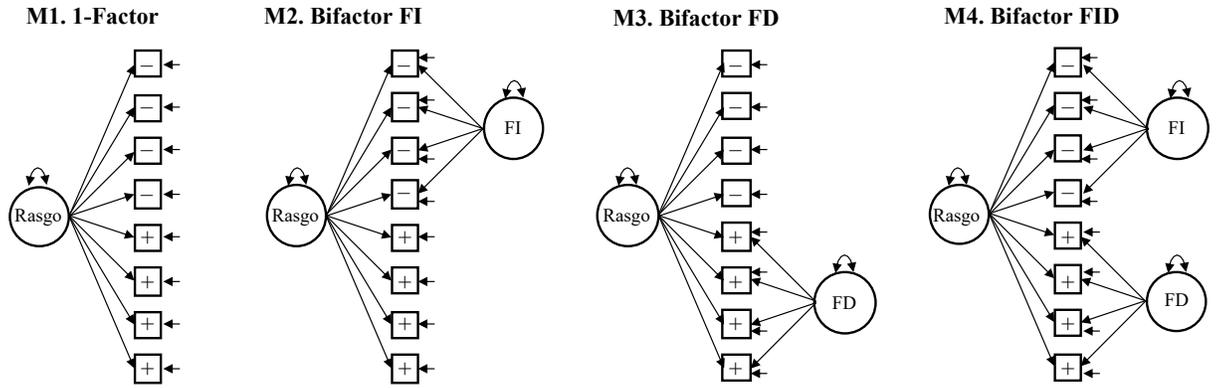


Figura 3. Modelos confirmatorios multirasgo-multimétodo.

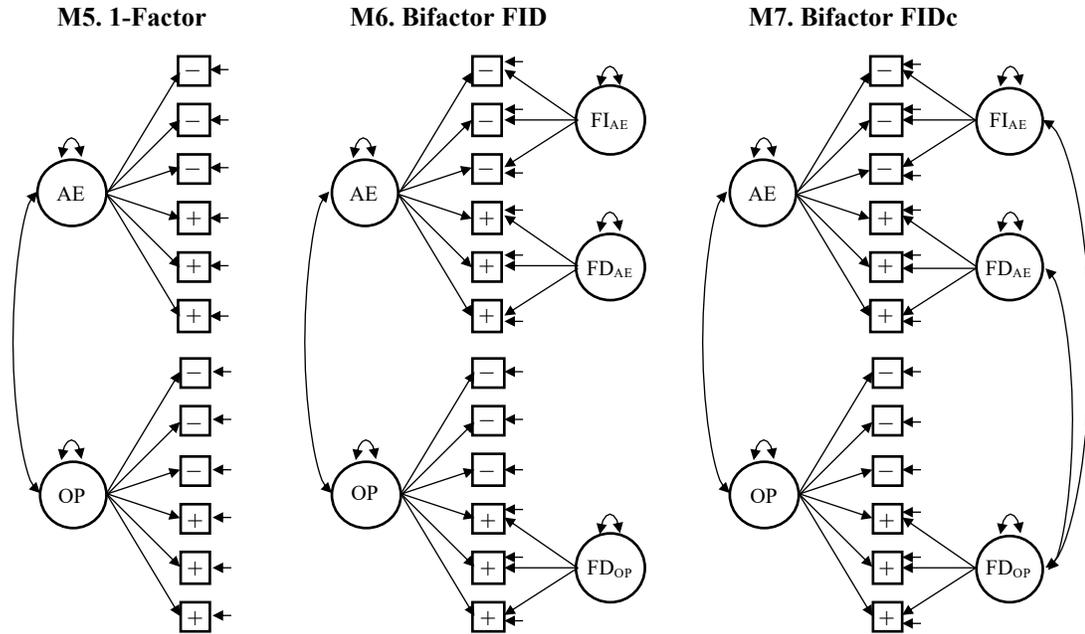


Figura 4. Diagrama de dispersión de las puntuaciones factoriales contenidas en la asociación entre FD_{OP} y FD_{AE} en la condición MC.

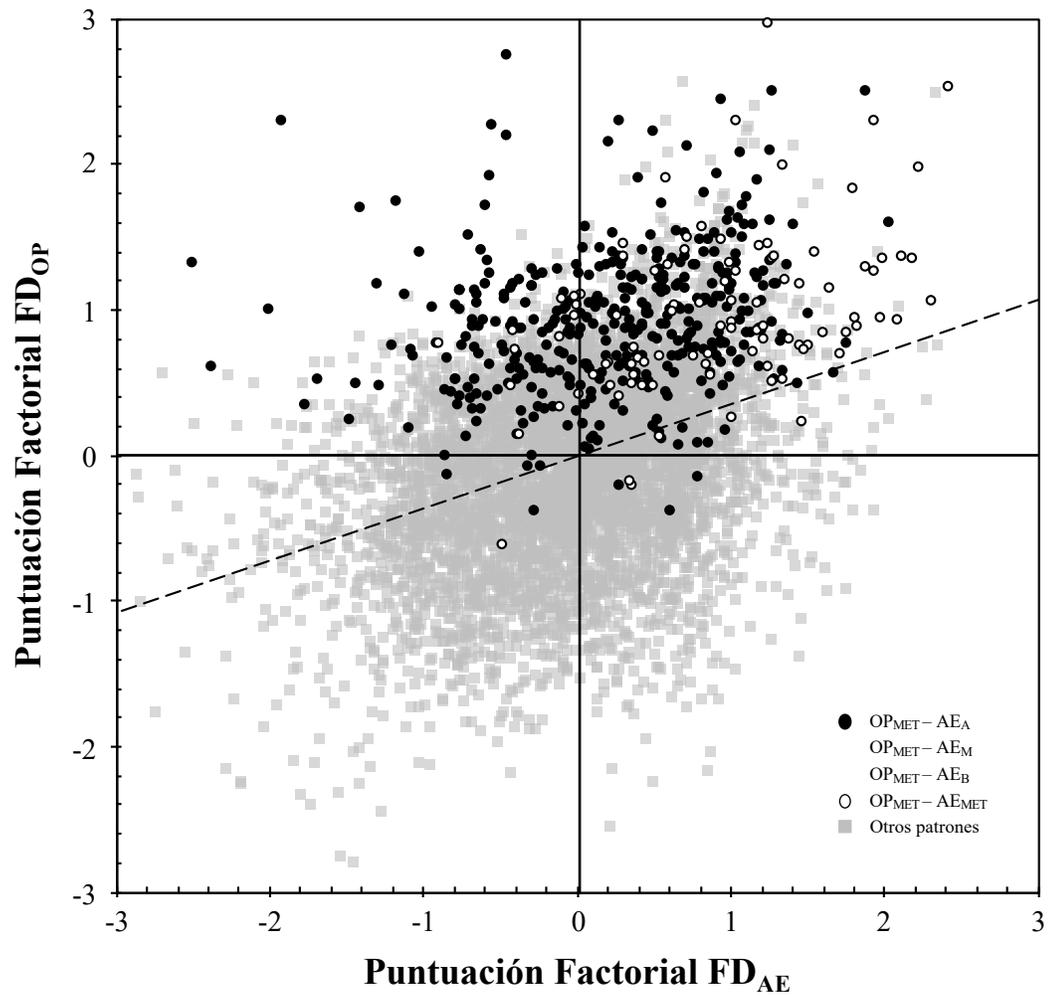


Figura 5. Modelos confirmatorios longitudinales

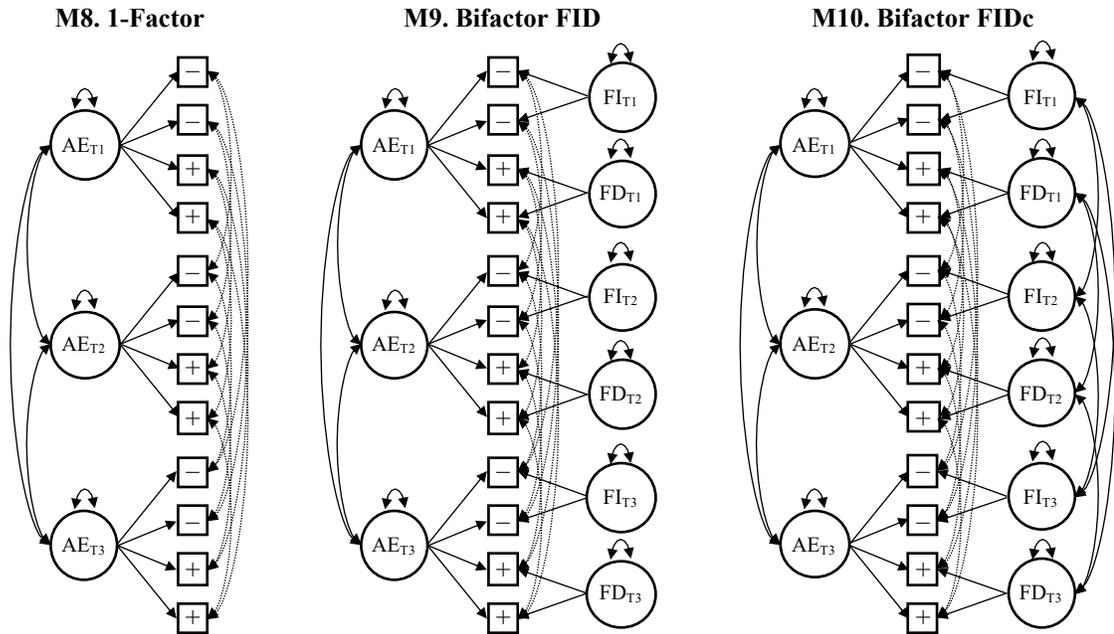


Figura 6. Medias estimadas de los ítems positivos y negativos de la RSES por perfiles identificados en cada ola de medición por tipo de análisis empleado.

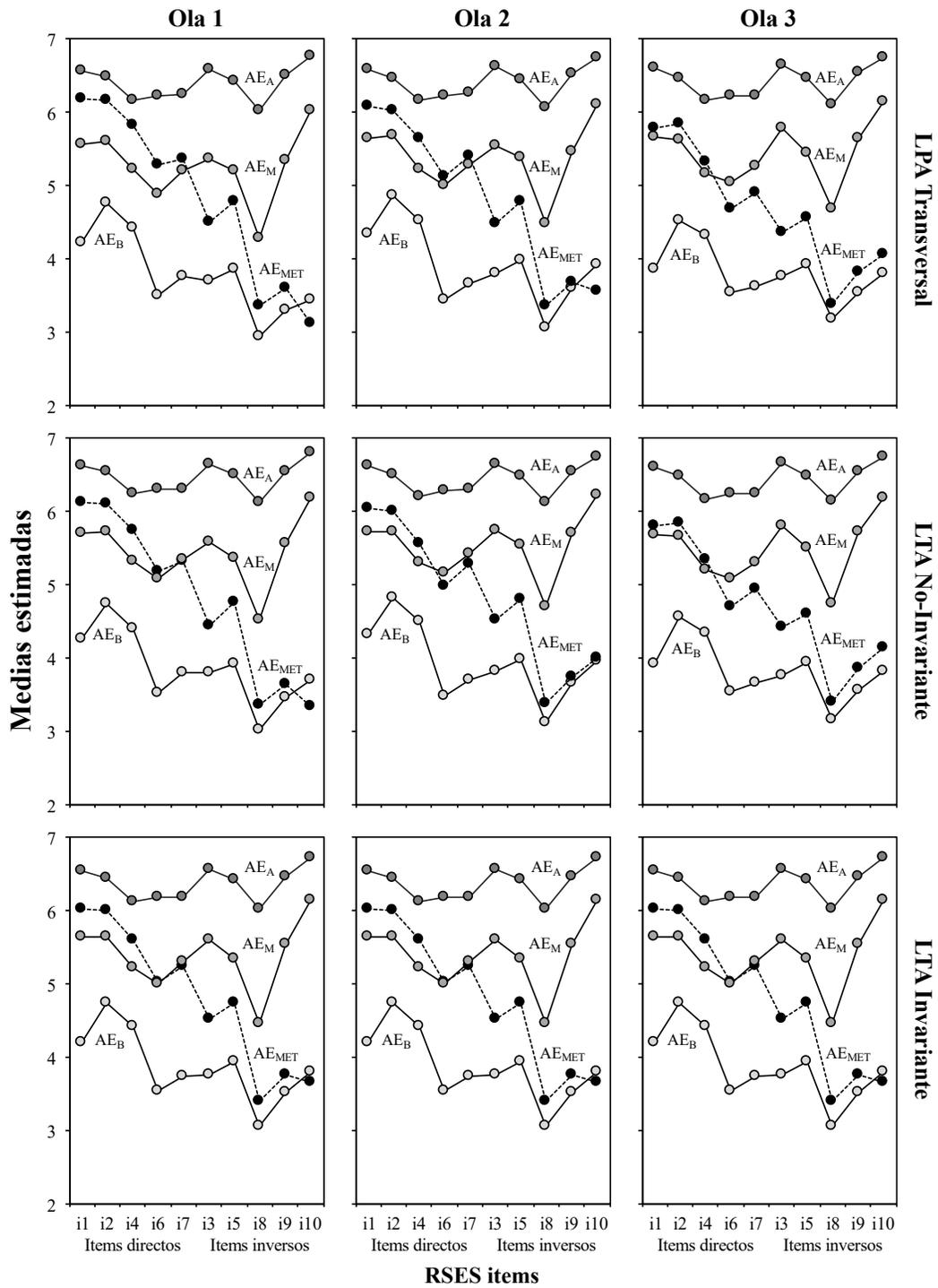


Figura 7. Diagrama de dispersión de las puntuaciones factoriales contenidas en la asociación entre FI_{T1} y FI_{T2} en la condición MC.

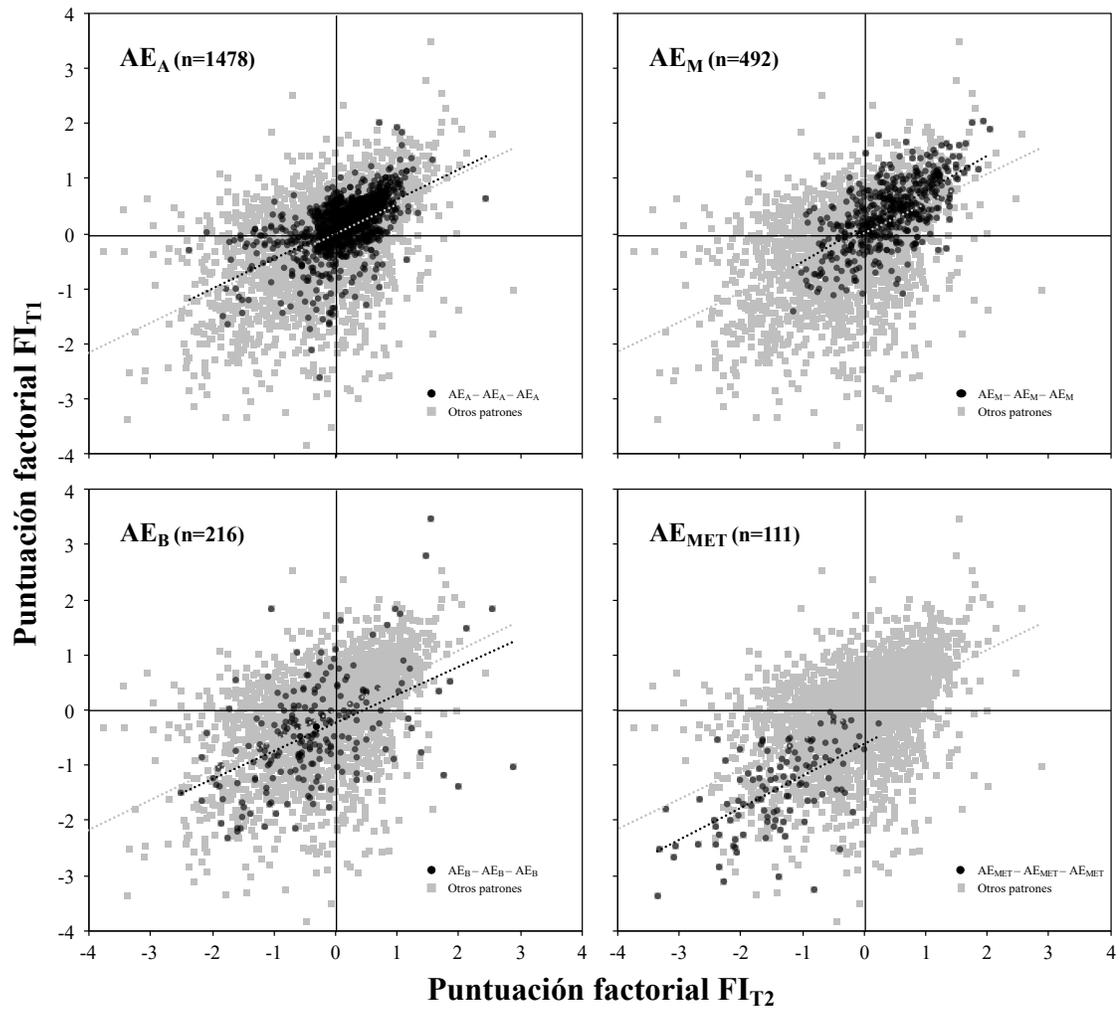


Figura 8. Distribución de porcentajes acumulados de las puntuaciones factoriales de los subgrupos por ola y condición muestral. Los valores entre paréntesis representan la puntuación factorial promedio y la proporción de participantes con puntuaciones entre -1 y +1.

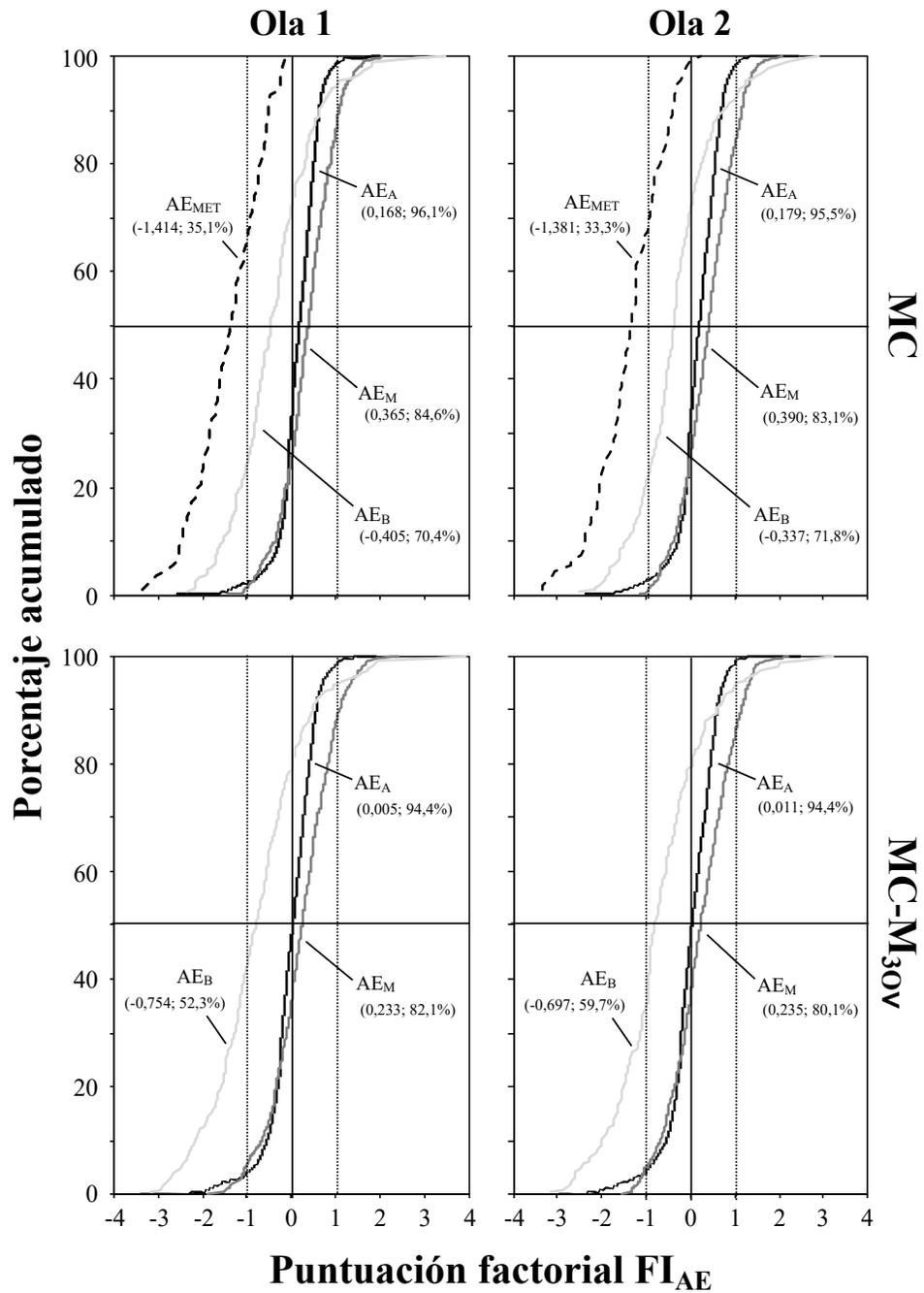
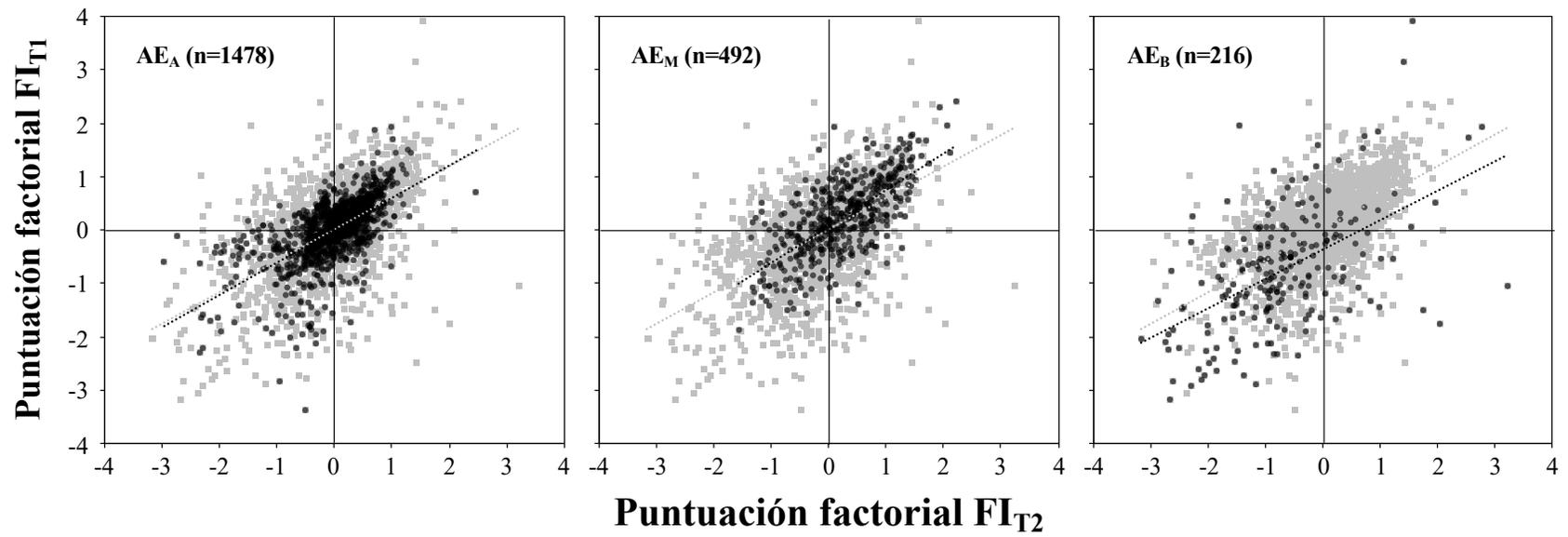


Figura 9. Diagrama de dispersión de las puntuaciones factoriales contenidas en la asociación entre FI_{T1} y FI_{T2} en la condición MC-M.



IV. DISCUSIÓN GENERAL

La presente tesis doctoral tuvo como propósito central abordar la evidencia empírica y las propuestas teóricas planteadas en relación al efecto de redacción de los ítems en el proceso de respuesta y sus consecuencias para la estimación de modelos de medida y la asociación con variables criterio, utilizando la escala de autoestima de Rosenberg (RSES) como modelo de investigación. Para abordar esto, tres objetivos centrales fueron propuestos: (a) evaluar el valor del modelo bifactor para examinar el efecto de redacción, examinando las propiedades del factor general y los factores específicos de redacción, (b) proponer una estrategia analítica alternativa, basada en la identificación de un grupo de individuos, para examinar el efecto de redacción, y (c) emplear esta estrategia para evaluar la consistencia entre atributos y la estabilidad temporal del efecto de redacción.

Los resultados de los estudios llevados a cabo, organizados en tres manuscritos en consonancia con los objetivos propuestos, permitieron aportar evidencia relevante respecto tanto de las estrategias metodológicas empleadas para el estudio y modelamiento de los efectos de redacción como acerca de las conclusiones teóricas que han sido descritas en las últimas dos décadas sobre la naturaleza sustantiva versus efímera de este efecto.

Los resultados de la revisión sistemática llevada a cabo en el primer manuscrito permitieron contribuir a la discusión sobre el valor e interpretabilidad del modelo bifactor y sus factores específicos de redacción. Este trabajo representó, hasta donde sabemos, el primero en evaluar exhaustivamente las propiedades de los factores de redacción y discutir sus implicancias en el uso de esta representación nomológica para dar cuenta del efecto de redacción y su naturaleza subyacente. En primer lugar, consistente con lo reportado en investigaciones previas, los hallazgos revelaron una solución esencialmente unidimensional (Kam, 2021), tanto por la proporción de varianza explicada por el factor general de autoestima como por la proporción de varianza confiable (reflejada en los valores de omega y omega jerárquico) asociada a la variabilidad de las puntuaciones totales que se puede atribuir al factor general por sobre los factores específicos de redacción (Rodríguez, Reise, & Haviland, 2016a,b). Del mismo modo, se estableció un

bajo grado de interpretabilidad y replicabilidad de los factores de redacción de la RSES, así como de las asociaciones estructurales estimables con dichos factores. Este hallazgo, a diferencia del grado de unidimensionalidad de la escala, ha sido escasamente reconocido tanto en aquellos estudios que examinaron la dimensionalidad de la escala de autoestima de Rosenberg como aquellos que emplearon el factor de redacción para examinar su relación con variables criterio.

Estos resultados ponen serios cuestionamientos a la validez de toda la evidencia reportada sobre la naturaleza aparentemente sustantiva del efecto de redacción basada en las asociaciones del factor de redacción, al menos aquella que ha empleado la RSES como escala balanceada de referencia (*e.g.*, DiStefano & Motl, 2016; 2019a,b).

Tomando en cuenta la cuestionable idoneidad del modelo bifactor para representar el efecto de redacción, en el segundo manuscrito se propuso una estrategia alternativa basada en el modelamiento de mixturas, con el propósito de identificar participantes con respuestas diferenciales y examinar su impacto sobre el modelo de medida. El primer estudio confirmó la presencia, en dos muestras aleatoriamente seleccionadas, de un grupo de participantes que exhibieron respuestas asimétricas, consistentes con la idea de efecto de redacción (Kamoen et al., 2013), y cuya proporción de individuos identificados fue consistente con las descripciones sugeridas por otros autores (Reise et al., 2016; Woods, 2006). Por su parte, los Estudios 2 y 3 examinaron el impacto de este grupo sobre la emergencia de multidimensionalidad y ajuste estadístico de los modelos de medida de la RSES, corroborando las consecuencias del subgrupo identificado con aquellas descritas por el efecto de redacción (Schmitt & Stults, 1985; Woods et al., 2017).

Estos resultados permitieron confirmar el valor del análisis de perfiles latentes como una estrategia válida y útil para modelar explícitamente la heterogeneidad presente en la población de encuestados, produciendo una taxonomía sustantivamente interpretable y comparable de participantes que exhibieron un patrón diferencial de respuestas cuyas consecuencias fueron consistentes con el efecto de redacción. Esta ventaja analítica sitúa al análisis propuesto por sobre otras nociones previamente descritas en la literatura (*e.g.*, IRLS, Reise et al., 2016; FMM, Arias et al., 2020; Steinman et al., 2021).

Más importante aún, el procedimiento analítico propuesto ofrece un repertorio de extensiones que permite examinar la naturaleza subyacente del efecto de redacción, a través de la evaluación de las propiedades más relevantes de la noción del estilo de respuesta, predominante en la literatura durante las últimas dos décadas, situando estos hallazgos como una contribución relevante para el estudio del efecto de redacción frente a las limitaciones y consecuencias indeseadas del modelamiento bifactor descritas en el primer manuscrito.

Finalmente, el tercer manuscrito aprovecha las ventajas y fortalezas descritas del procedimiento de mixturas presentado en el segundo manuscrito para examinar la evidencia de consistencia entre escalas y estabilidad temporal. La estrategia general, aunque extensa, permitió ilustrar exhaustivamente la presencia e impacto de un grupo de encuestados con patrones de respuestas diferenciales en dos escalas, cuya evidencia de presencia de efectos de redacción ha sido previamente descrita en la literatura (Gnambs et al., 2018; Vecchione et al., 2014), confirmando la ubicuidad del efecto y sus consecuencias bajo esta estrategia alternativa de análisis y detección (Biderman et al., 2011).

Posteriormente, una vez corroborada la presencia del efecto en ambas escalas, la estrategia consistió en examinar, una vez replicada la evidencia de asociación entre los factores de redacción de ambas escalas (DiStefano & Motl, 2006; Horan et al., 2003), la presencia de un grupo de participantes con el patrón conjunto de respuestas asimétricas en ambas escalas, evaluar dicha proporción de individuos, y examinar sus consecuencias sobre el ajuste estadístico y la magnitud de la asociación entre los factores de redacción del modelo estructural previamente replicado. Luego, siguiendo la estrategia previamente descrita, se estimó y replicó la evidencia de estabilidad temporal e invarianza longitudinal previamente reportada (Marsh et al., 2010; Gana et al., 2013) y, utilizando el mismo set de datos, se examinó la presencia conjunta de individuos con patrones de respuesta asimétricos, los parámetros de transición obtenidos, y sus consecuencias sobre el modelo longitudinal previamente estimado, en términos de ajuste e impacto sobre la asociación.

En ambos casos, nuestros hallazgos revelaron la presencia de asociaciones espurias entre los factores de redacción de ambas escalas o a nivel longitudinal, cuya presencia y magnitud capitalizaba las puntuaciones conjuntas de individuos que no exhibían ningún grado de sesgo de respuesta o lo exhibían en una de las dos escalas, contradiciendo los presupuestos contenidos en la interpretación sustantiva de las asociaciones estimadas y, con ello, descartando la evidencia de naturaleza sustantiva del efecto de redacción.

Estos hallazgos contradicen la literatura predominante, siendo esta una de las principales contribuciones de la presente tesis en tanto, hasta donde sabemos, ningún estudio previamente reportado ha examinado la evidencia asociada a la consistencia entre escalas o estabilidad temporal del efecto de redacción empleando modelamiento de este tipo ni ha optado, a partir de sus resultados, por sugerir esta interpretación del efecto de redacción y de las asociaciones de los factores de redacción.

BIBLIOGRAFÍA

- Alessandri, G., Vecchione, M., Tisak, J., & Barbaranelli, C. (2011). Investigating the nature of method factors through multiple informants: Evidence for a specific factor? *Multivariate Behavioral Research*, *46*(4), 625-642.
<https://doi.org/10.1080/00273171.2011.589272>
- Arias, V.B., & Arias, B. (2017). The negative wording factor of Core Self-Evaluations Scale (CSES): Methodological artifact, or substantive specific variance? *Personality and Individual Differences*, *109*, 28-34. <https://doi.org/10.1016/j.paid.2016.12.038>
- Arias, V.B., Garrido, L.E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*, 2489-2505.
<https://doi.org/10.3758/s13428-020-01401-8>
- Alessandri, G., Vecchione, M., Donnellan, B.M., & Tisak, J. (2013). An application of the LCLSTM framework to the self-esteem instability case. *Psychometrika*, *78*(4), 769-92. <https://doi.org/10.1007/s11336-013-9326-4>
- Alessandri, G., Vecchione, M., Eisenberg, N., & Laguna, M. (2015). On the factor structure of the Rosenberg (1965) General Self-Esteem Scale. *Psychological Assessment*, *27*(2), 621-635. <https://doi.org/10.1037/pas0000073>
- Baumeister, R.F., Campbell, J.D., Krueger, J.I., Vohs, K.D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, *4*(1), 1-44.
<https://doi.org/10.1111/1529-1006.01431>
- Baumeister, R.F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements. *Perspectives on Psychological Science*, *2*(4), 396-403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Bentler, P.M., Jackson, D.N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, *76*, 186-204. <https://doi.org/10.1037/h0031474>

- Biderman, M.D., Nguyen, N.T., Cunningham, C.J.L., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the Big Five. *Journal of Research in Personality*, 45, 417-429. <https://doi.org/10.1016/j.jrp.2011.05.001>
- Billiet, J., & McClelland, M. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608-628. https://doi.org/10.1207/S15328007SEM0704_5
- Bonifay, W. (2015). An integrative framework of model evaluation (unpublished doctoral dissertation). University of California, Los Angeles. <https://escholarship.org/uc/item/5j04v50q>
- Bonifay, W., & Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behavioral Research*, 52(4), 465-484. <https://doi.org/10.1080/00273171.2017.1309262>
- Bonifay, W., Lane, S.P., & Reise, S.P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184-186. <https://doi.org/10.1177/2167702616657069>
- Cabooter, E., Weijters, B., De Beuckelaer, A., Davidov, E. (2017). Is extreme response style domain specific? Findings from two studies in four countries. *Quality and Quantity*, 51, 2605-2622. <https://doi.org/10.1007/s11135-016-0411-5>
- Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research*, 69, 2574-2584. <https://doi.org/10.1016/j.jbusres.2015.10.138>
- Conroy, D.E. (2001). Progress in the development of a multidimensional measure of fear of failure: The Performance Failure Appraisal Inventory (PFAI). *Anxiety, Stress & Coping: An International Journal*, 14(4), 431-452. <https://doi.org/10.1080/10615800108248365>
- Corwyn, R.F. (2000). The factor structure of global self-esteem among adolescents and adults. *Journal of Research in Personality*, 34(4), 357-379. <https://doi.org/10.1006/jrpe.2000.2291>

- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475-494. <https://doi.org/10.1177/001316444600600405>
- Cronbach, L.J. (1956). Assessment of individual differences. *Annual Review of Psychology*, 7, 173-196. <https://doi.org/10.1146/annurev.ps.07.020156.001133>
- DiStefano, C., & Motl, R.W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13, 440-464. https://doi.org/10.1207/s15328007sem1303_6
- DiStefano, C., & Motl, R.W. (2009a). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling*, 16, 134-146. <https://doi.org/10.1080/10705510802565403>
- DiStefano, C., & Motl, R.W. (2009b). Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences*, 309-313. <https://doi.org/10.1016/j.paid.2008.10.020>
- Donnellan, M.B., Trzesniewski, K.H., & Robins, R.W. (2013). Self-esteem: Enduring issues and controversies. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.). *The Wiley-Blackwell handbook of individual differences*. Hoboken, NJ: Blackwell Publishing Ltd. <https://doi.org/10.1002/9781444343120.ch28>
- Dueber, D.M., Toland, M.D., Lingat, J.E., Love, A.M.A., Qiu, C., Wu, R., & Brown, A.V. (2021). To reverse item orientation or not to reverse item orientation, That is the question. *Assessment*. <https://doi.org/10.1177/10731911211017635>
- Gana, K., Saada, Y., Bailly, N., Joulain, M., Hervé, M., & Alaphilippe, D. (2013). Longitudinal factorial invariance of the Rosenberg self-esteem scale: Determining the nature of method effects due to item wording. *Journal of Research in Personality*, 47(4), 406-416. <https://doi.org/10.1016/j.jrp.2013.03.011>
- Geukes, K., Nestler, S., Hutteman, R., Küfner, A.C.P., & Back, M.D. (2017). Trait personality and state variability: Predicting individual differences in within- and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, 69, 124-138. <https://doi.org/10.1016/j.jrp.2016.06.003>

- Gnambs, T., Scharl, A., & Schroeders, U. (2018). The structure of the Rosenberg self-esteem scale: A cross-cultural meta-analysis. *Zeitschrift für Psychologie*, 226(1), 14-29. <https://doi.org/10.1027/2151-2604/a000317>
- Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S.P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, 35, 1241-1254. [https://doi.org/10.1016/S0191-8869\(02\)00331-8](https://doi.org/10.1016/S0191-8869(02)00331-8)
- Gu, H., Wen, Z., & Fan, X. (2015). The impact of wording effect on reliability and validity of the Core Self-Evaluation Scale (CSES): A bi-factor perspective. *Personality and Individual Differences*, 83, 142-147. <https://doi.org/10.1016/j.paid.2015.04.006>
- Haefel, G.J., & Howard, G.S. (2010). Self-report: Psychology's four-letter word. *The American Journal of Psychology*, 123(2), 181-188. <https://doi.org/10.5406/amerjpsyc.123.2.0181>
- Horan, P.M., DiStefano, C., & Motl, R.W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling*, 10(3), 444-455. https://doi.org/10.1207/S15328007SEM1003_6
- Huang, J.L., Liu, M., & Bowling, N.A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100(3), 828-845. <http://dx.doi.org/10.1037/a0038510>
- Hyland, P., Boduszek, D., Dhingra, K., Shevlin, M., & Egan, A. (2014). A bifactor approach to modelling the Rosenberg Self Esteem scale. *Personality and Individual Differences*, 66, 188-192. <https://doi.org/10.1016/j.paid.2014.03.034>
- Kam, C.C.S. (2018). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociological Methods & Research*, 47(3), 574-597. <https://doi.org/10.1177/0049124115626177>
- Kam, C.C.S. (2021). Bifactor model is not the best-fitting model for self-esteem: Investigation with a novel technique. *Assessment*, 28(7), 1799-1812. <https://doi.org/10.1177/1073191120949916>

- Kamoen, N., Holleman, B., & van den Bergh, H. (2013). Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*, 7(3), 181-189. <https://doi.org/10.18148/srm/2013.v7i3.5034>
- Jackson, D.N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55, 243-252. <https://doi.org/10.1037/h0045996>
- Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409-426. <https://doi.org/10.1007/BF02291366>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 5-55.
- Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, 94, 196-204. <https://doi.org/10.1080/00223891.2011.645936>
- Markus, K.A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.
- Marsh, H.W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, 70, 810-819. <https://doi.org/10.1037/0022-3514.70.4.810>
- Marsh, H.W., Scalas, L.F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment*, 22, 366-381. <https://doi.org/10.1037/a0019225>
- Maydeu-Olivares, A., & Coffman, D.L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344-362. <https://doi.org/10.1037/1082-989X.11.4.344>
- McPherson, J., & Mohr, P. (2005). The role of item extremity in the emergence of keying-related factors: An exploration with the Life Orientation Test. *Psychological Methods*, 10(1), 120-131. <https://doi.org/10.1037/1082-989X.10.1.120>

- Michaelides, M.P., Koutsogiorgi, C., & Panayiotou, G. (2016). Method effects on an adaptation of the Rosenberg Self-Esteem Scale in Greek and the role of personality traits. *Journal of Personality Assessment*, *98*, 178-188.
<https://doi.org/10.1080/00223891.2015.1089248>
- Michaelides, M.P., Zenger, M., Koutsogiorgi, C., Brähler, E., Stöbel-Richter, Y., & Berth, H. (2016). Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Personality and Individual Differences*, *97*, 13-18. <https://doi.org/10.1016/j.paid.2016.03.011>
- Motl, R.W., & Conroy, D.E. (2000). Validity and factorial invariance of the social physique anxiety scale. *Medicine & Science in Sports & Exercise*, *32*(5), 1007-1017. <https://doi.org/10.1097/00005768-200005000-00020>
- Motl, R.W., Conroy, D.E., & Horan, P.M. (2000). The Social Physique Anxiety Scale: An example of the potential consequence of negatively worded items in factorial validity studies. *Journal of Applied Measurement*, *1*(4), 327-345.
- Motl, R.W., Dishman, R.K., Felton, G., & Pate, R.R. (2003). Self-motivation and physical activity among black and white adolescent girls. *Medicine & Science in Sports & Exercise*, *35*(1), 128–136. <https://doi.org/10.1097/00005768-200301000-00020>
- Motl, R.W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*, *9*(4), 562-578. https://doi.org/10.1207/S15328007SEM0904_6
- Muthén, B.O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*(4), 557-585. <https://doi.org/10.1007/BF02296397>
- Nunnally, J.M. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71.
<https://doi.org/10.1136/bmj.n71>

- Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press.
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P.M., MacKenzie, S.B., & Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Quilty, L.C., Oakman, J.M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling, 13*, 99-117. https://doi.org/10.1207/s15328007sem1301_5
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223-237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rodríguez, A., Reise, S.P., & Haviland, M.G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137-150. <https://doi.org/10.1037/met0000045>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Roszkowski, M.J., & Soven, M. (2010). Shifting gears: consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education, 35*(1), 113-130. <https://doi.org/10.1080/02602930802618344>
- Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly, 5*, 91-92. <https://doi.org/10.1086/265467>

- Sackett, P.R., Lievens, F., Van Iddekinge, C.H., & Kuncel, N.R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102*, 254-273. <https://doi.org/10.1037/apl0000151>
- Schmitt, N., & Stults, D.M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9*(4), 367-373. <https://doi.org/10.1177/014662168500900405>
- Schriesheim, C.A., Eisenbach, R.J., & Hill, K.D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement, 51*(1), 67-78. <https://doi.org/10.1177/0013164491511005>
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239. <https://doi.org/10.1111/j.2044-8317.1974.tb00543.x>
- Spector, P.E., Van Katwyk, P.T., Brannick M.T., & Chen P.Y. (1997). When two factors don't reflect two constructs: how item characteristics can produce artifactual factors. *Journal of Management, 23*(5), 659-677. [https://doi.org/10.1016/S0149-2063\(97\)90020-9](https://doi.org/10.1016/S0149-2063(97)90020-9)
- Steinmann, I., Strietholt, R., & Braeken, J. (2021). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological Methods. <https://doi.org/10.1037/met0000392>*
- Tomas, J.M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling, 6*, 84-98. <https://doi.org/10.1080/10705519909540120>
- Tomas, J.M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling, 20*, 299-313. <https://doi.org/10.1080/10705511.2013.769394>
- Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

- Vecchione, M., Alessandri, G., Caprara, G.V., & Tisak, J. (2014). Are method effects permanent or ephemeral in nature? The case of the revised Life Orientation Test. *Structural Equation Modeling, 21*(1), 117-130.
<https://doi.org/10.1080/10705511.2014.859511>
- Watts, A.L., Poore, H.E., & Waldman, I.D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science, 7*(6), 1285-1303. <https://doi.org/10.1177/2167702619855035>
- Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reverse item bias: An integrative model. *Psychological Methods, 18*(3), 320-334.
<https://doi.org/10.1037/a0032121>
- Widaman, K.F. (2012). Exploratory factor analysis and confirmatory factor analysis. In Cooper, H., Camic, P.M., Long, D.L., Panter, A.T., Rindskopf, D., Sher, K.J. (Eds.), *APA handbook of research methods in psychology*. Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/13621-018>.
- Wilson, T.D., & Dunn, E.W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology, 55*, 493-518.
<https://doi.org/10.1146/annurev.psych.55.090902.141954>
- Wood, D., Harms, P.D., Lowman, G.H., & DeSimone, J.A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychology and Personality Science, 8*(4), 454-464.
<https://doi.org/10.1177/1948550617703168>
- Woods, C.M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28*(3), 189-194. <https://doi.org/10.1007/s10862-005-9004-7>