

Estudio de Seguimiento de las Recomendaciones sobre Análisis Factorial Exploratorio en RIDEP

A Follow-up Study of the Recommendations on Exploratory Factor Analysis in RIDEP

David Paniagua¹, Jesús M. Alvarado², María Olivares³, Irene Ruiz⁴, Marcos Romero-Suárez⁵ y Raimundo Aguayo-Estremera⁶

Resumen

En el año 2019, se publicó una revisión de la calidad de las publicaciones de la RIDEP que contenían un análisis factorial exploratorio (AFE) en los seis últimos números hasta el segundo volumen de 2018. En dicho trabajo se exploraba la adherencia de los artículos de la revista a los estándares y recomendaciones. En el presente trabajo se amplía la evaluación, analizando todas las publicaciones con un AFE desde el tercer volumen de 2018 hasta el tercer volumen de 2022. Se evaluaron (a) el diseño, (b) aspectos previos al análisis, (c) la ejecución del AFE y (d) la información escrita de 83 AFE publicados en la RIDEP. En general, la calidad de la información de los artículos con AFE publicados en la RIDEP ha mejorado frente a la anterior revisión, pero con elementos donde sigue siendo necesario un esfuerzo. Se debaten algunas recomendaciones para autores, revisores y la revista.

Palabras clave: análisis factorial exploratorio, psicometría, revisión

Abstract

In 2019, a review of the quality of the RIDEP publications that contained an exploratory factor analysis (EFA) was published. They asses the last six issues until the second volume of 2018. In that work, the adherence of the journal articles was explored. In the present work, the evaluation is extended, analyzing all the publications with an EFA from the third volume of 2018 to the third volume of 2022. We evaluated (a) the design, (b) aspects prior to the analysis, (c) the execution of the EFA and (d) the written information of 83 EFA. In general, the quality of the information of the articles with EFA published in the RIDEP has improved compared to the previous review, but there are some elements where an effort is still necessary. Some recommendations for authors, reviewers and the journal are discussed.

Keywords: exploratory factor analysis, psychometrics, review

¹ MSc. Personal Docente Investigador. Facultad de Psicología, Universidad Complutense de Madrid, España. Campus de Somosaguas, Ctra. de Húmera, s/n, 28223 Pozuelo de Alarcón, Madrid, España. Correo: davidpan@ucm.es

² PhD. Personal Docente Investigador. Facultad de Psicología, Universidad Complutense de Madrid, España. Campus de Somosaguas, Ctra. de Húmera, s/n, 28223 Pozuelo de Alarcón, Madrid, España. Correo: jmalvara@ucm.es

³ MSc. Licenciada en Psicología. Facultad de Psicología, Universidad Complutense de Madrid, España. Campus de Somosaguas, Ctra. de Húmera, s/n, 28223 Pozuelo de Alarcón, Madrid, España. Correo: maroli04@ucm.es.

⁴ MSc. Licenciada en Psicología. Facultad de Psicología, Universidad Complutense de Madrid, España. Campus de Somosaguas, Ctra. de Húmera, s/n, 28223 Pozuelo de Alarcón, Madrid, España. Correo: irenru02@ucm.es

⁵ MsC. Licenciada en Psicología. Instituto de Estudios Feministas, Universidad Complutense de Madrid, España. Campus de Somosaguas, Ctra. de Húmera, s/n, 28223 Pozuelo de Alarcón, Madrid, España. Correo : marcro03@ucm.es

⁶ PhD. Personal Docente Investigador. Facultad de Psicología, Universidad Complutense de Madrid, España. Campus de Somosaguas, Ctra. de Húmera, s/n, 28223 Pozuelo de Alarcón, Madrid, España. Correo: raaguayo@ucm.es (Autor de correspondencia)

Introducción

Charles Spearman, en su trabajo seminal de 1904, propuso el primer modelo de análisis factorial. Más tarde, Thurstone (1935, 1947) planteó otra propuesta a la que denominó modelo del factor común. Este modelo matemático ha dado lugar al desarrollo de los modelos de Análisis Factorial Exploratorio (AFE) y Confirmatorio (AFC) y los procedimientos más recientes se siguen basando en él. Han pasado varias décadas desde estos trabajos pioneros y el AFE se ha convertido en una de las técnicas más usadas para fines psicométricos (Frías-Navarro & Pascual, 2012; Goretzko et al., 2021; Izquierdo et al., 2014; Ledesma et al., 2019; Lloret-Segura et al., 2014; Osborne, 2014).

Sin embargo, su enorme popularidad ha ido de la mano de numerosas controversias. Mientras que en unas ocasiones el foco se ha centrado en la propia técnica (Ferrando & Lorenzo-Seva, 2014), en otras el punto de mira se ha puesto en el uso que hacen los investigadores de ella (Frías-Navarro & Pascual, 2012; Izquierdo et al., 2014; Ledesma et al., 2019; Lloret-Segura et al., 2014). Posiblemente, la crítica más frecuente y más fundamentada haya sido la referida a los malos usos y abusos de la técnica por parte de investigadores aplicados.

Una de las razones que pueden haber motivado estos malos usos radica en el desarrollo de gran variedad de nuevos procedimientos de análisis factorial. Ello puede ocasionar que muchos investigadores aplicados no estén al tanto de las últimas novedades en el campo (Lloret-Segura et al., 2014). Por otra parte, con el desarrollo de los modelos de ecuaciones estructurales y de software especializado, el AFE fue percibido como obsoleto durante algunos años (Ferrando & Lorenzo-Seva, 2014). Esto pudo conllevar la pérdida de interés en aplicar las técnicas correctas de AFE por parte de los investigadores y a la repetición de procedimientos obsoletos.

Otra de las razones se puede encontrar en los diversos cambios en las recomendaciones acerca de cómo emplear la técnica. Al mismo tiempo que se iban desarrollando nuevos procedimientos, cambiaban los criterios de toma de decisiones para el análisis y la interpretación de los

resultados (Lloret-Segura et al., 2014). Algunos de los cambios se propusieron hace tiempo, por ejemplo: el abandono del método de componentes principales; el de la regla de Kaiser para seleccionar el número de factores; el del método de rotación ortogonal como procedimiento preferente; el de criterios simples, como la proporción de personas por ítem, para determinar el tamaño muestral; y el del uso de la matriz de correlaciones de Pearson independientemente del nivel de medida de los ítems o del cumplimiento del supuesto de normalidad. Otras propuestas y sugerencias, no obstante, son más recientes, por ejemplo: la idea de que AFE y AFC son los dos polos de un mismo continuo; el desarrollo de procedimientos híbridos como los modelos de ecuaciones estructurales exploratorios (ESEM, por sus siglas en inglés) o la rotación procusteana; y la implementación de métodos robustos a la violación del supuesto de normalidad de los ítems, como máxima verosimilitud robusta o mínimos cuadrados ponderados ajustados por la media y la varianza. Se puede encontrar una explicación detallada de estas cuestiones en otros trabajos (v.g., Brown, 2015; Fabrigar & Wegener, 2011; Kline, 2015).

Estos malos usos y abusos de la técnica de AFE podrían estar relacionados con las Prácticas Cuestionables de Investigación (PCI). Bajo este nombre se recogen prácticas o conductas no éticas aplicadas en cualquiera de los procesos de una investigación de forma intencionada con el propósito de conseguir resultados (v.g., publicar un artículo) en detrimento de la calidad de la misma (Fiedler y Schwarz, 2016; John et al., 2012). Son los llamados grados de libertad en la investigación que actúan alineados con los intereses personales del investigador y suelen estar incentivados por factores externos como la presión por publicar (van Dijk et al., 2014), la financiación y políticas éticas débiles o con una implementación laxa en revistas científicas (Saunders y Savulescu, 2007). En este sentido, los investigadores que realizan AFE podrían usar un procedimiento desaconsejado (por ejemplo, análisis de componentes principales, no informar de todos los pesos factoriales o de la fiabilidad) porque consiguen resultados con mayores perspectivas de publicación.

Los dos primeros motivos alegados (profusión de procedimientos novedosos y cambio en las

recomendaciones de uso) ponen de manifiesto la dificultad y la flexibilidad de análisis a la que se enfrentan los investigadores aplicados. Sin embargo, los psicómetras especializados en el campo del análisis factorial no han estado ajenos a estos problemas y han ido desarrollando decálogos de buenas prácticas y guías con recomendaciones. Algunos de ellos están disponibles en inglés (Goretzko et al., 2021; Izquierdo et al., 2014; Osborne, 2014) y otros, más accesibles para el investigador iberoamericano, en español (Ferrando et al., 2022; Ferrando & Anguiano-Carrasco, 2010; Ferrando & Lorenzo-Seva, 2014; Izquierdo et al., 2014; Pérez & Medrano, 2010) y en portugués (Damásio, 2012). De hecho, Ledesma et al. publicaron en 2019 un trabajo en la Revista Iberoamericana de Diagnóstico y Evaluación Psicológica (RIDEP) con un doble objetivo: (a) ofrecer un resumen de algunas de las prácticas recomendadas para hacer AFE, junto con una guía básica para autores y revisores; y (b) estudiar los procedimientos empleados para hacer AFE en artículos publicados en la RIDEP.

Los resultados del estudio de Ledesma et al. (2019) mostraron que, a pesar de haber abundante documentación sobre la aplicación del AFE, por un lado, los investigadores aplicados continuaron cometiendo errores, algunos elementales, en sus aplicaciones; y, por otro lado, los revisores dejaron pasar estos errores y permitieron la publicación de los trabajos. Se encontraron entre otros errores que la mayoría de los estudios usaban el procedimiento conocido como “Little Jiffy” (componentes principales, regla de Kaiser y rotación ortogonal). Además, las omisiones en el informe de investigación fueron frecuentes, como lo ilustra el hecho de que menos de un tercio presentaban la matriz de pesos factoriales completa. Cabe reseñar que estos resultados fueron parecidos a los de otros estudios sobre prácticas de publicación (Frías-Navarro & Pascual, 2012; Izquierdo et al., 2014).

Ledesma et al. (2019) finalizaban su trabajo con una llamada a realizar un esfuerzo conjunto de autores, revisores y editores para mejorar las prácticas en el uso del AFE. A este reto se han sumado los autores de otros trabajos del mismo estilo (Frías-Navarro & Pascual, 2012; Goretzko et al., 2021; Lloret-Segura et al., 2014). En el caso

de la RIDEP, existe una guía aplicada de buenas prácticas desde 2019 (Ledesma et al., 2019). En dicha guía se presentan seis puntos básicos que tratan varias de las prácticas de AFE. Éstos son: (1) datos básicos de aplicación; (2) datos de entrada y análisis preliminares; (3) métodos de extracción; (4) número de factores; (5) métodos de rotación; y (6) tablas con matrices de pesos factoriales. En cada uno de ellos, los autores ofrecen varias recomendaciones para realizar el AFE. A continuación, referimos varias de las recomendaciones a modo de ejemplo (se recomienda fehacientemente leer la guía en su totalidad). En el primer punto, se recomienda informar del tamaño muestral sobre el que se realiza el AFE e incluir los estadísticos descriptivos de los ítems. En el segundo, se sugiere consignar la matriz de correlaciones que se factoriza y recabar al menos 200 participantes si se usan correlaciones policóricas. En el tercero, se desaconseja usar por defecto componentes principales. En el cuarto, se previene sobre el uso automático de la regla de Kaiser. En el quinto, se desaconseja usar la rotación ortogonal por defecto. Finalmente, en el sexto punto se sugiere informar de los pesos factoriales de todos los ítems en cada uno de los factores.

El AFE es una técnica con la que los investigadores tratan de obtener evidencias de validez, en este caso, basadas en la estructura interna del test. La validez se refiere al grado en que la evidencia y la teoría apoyan las interpretaciones de las puntuaciones de los instrumentos para un uso concreto (AERA et al., 2014). Existen varias fuentes de validez (contenido, procesos de respuesta, estructura interna, relación con otras variables y consecuencial). Las evidencias de validez interna evalúan el grado en que las relaciones entre los ítems y los factores se ajustan al constructo en el que se basan las interpretaciones propuestas de la puntuación del test. Así, la aplicación incorrecta del AFE puede conllevar conclusiones erróneas acerca de la estructura factorial de los tests que atenten contra su validez. Es decir, pueden conllevar una falsa evidencia de factores que no son apoyados por la teoría y de usos para los que el test no se propone. Por ejemplo, la eliminación de ítems por bajos pesos factoriales que de haber seguido el procedimiento recomendado habrían

tenido una buena discriminación. Por un lado, eliminar ítems puede ocasionar la subrepresentación del contenido del constructo (Messick, 1995). Por otro lado, mantener ítems que deberían haber sido eliminados puede formar factores irrelevantes con respecto a los contenidos del constructo (Elosua, 2003). Otra de las consecuencias negativas es obtener estructuras factoriales erróneas, que pueden llevar a falta de replicabilidad no solo de las soluciones obtenidas en estudios que aplican AF sino en estudios posteriores que traten de explorar la relación entre el constructo medido y otros (por ejemplo, estudios explicativos que usen modelos de ecuaciones estructurales). Además, obtener una estructura factorial equivocada puede tener consecuencias sociales indeseadas como un diagnóstico erróneo al emplear el test en personas concretas.

De la definición de validez interna (AERA et al., 2014) se desprende que el AFE es una técnica con la que se trata de obtener evidencias acerca de la relación entre ítems y constructos. En este sentido, tanto AFE como AFC forman parte del mismo modelo matemático (modelo del factor común) y ambas técnicas se pueden usar para obtener evidencias de validez interna. No obstante, no ocurre lo mismo con el APC: AF y ACP no son técnicas equivalentes ni intercambiables entre sí (Fabrigar & Wegemer, 2011). Este último no tiene el mismo fundamento matemático y su propósito es radicalmente distinto. Mientras que el AF trata de explicar las correlaciones entre variables observadas a partir de unas pocas variables latentes (i.e., constructos), el ACP sencillamente reduce las puntuaciones de una serie de variables observables en un conjunto más reducido de variables observables (i.e., componentes). Por ello, mientras que el AFE es una técnica adecuada para obtener evidencias de validez interna, el ACP no lo es. Dicho de otra forma, el AF es apropiado cuando se trata con modelos teóricos reflectivos (i.e., un conjunto de indicadores observables cuyas puntuaciones dependen de un factor latente con sentido psicológico) y es inadecuado cuando se usan modelos teóricos formativos (i.e., un conjunto de indicadores observables cuyas puntuaciones pueden agruparse en componentes sin un modelo psicológico subyacente). Por su parte, el ACP es apropiado cuando se usa un modelo formativo,

pero inapropiado cuando el modelo es reflectivo. La inadecuada comprensión de esta diferencia por parte de los investigadores aplicados podría ser una razón que explicara el sobreuso de ACP, sobre todo en los últimos años en los que la complejidad computacional de otros métodos de estimación ya no supone un problema.

Dado que en la RIDEP existe una guía de recomendaciones prácticas para realizar el AFE desde 2019, resultaría útil comprobar si está surtiendo el efecto deseado y estamos (autores, revisores y editores) respondiendo a la llamada que hacían Ledesma et al. (2019) para realizar mejores prácticas. Precisamente el objetivo del presente estudio es analizar las aplicaciones de AFE desde que se publicó dicha guía de buenas prácticas. De esta forma, se podrá valorar si autores y revisores están ajustando su trabajo a las recomendaciones allí propuestas.

El trabajo está organizado de la siguiente forma. En el método se describe el proceso de inclusión de los estudios analizados, así como las prácticas de realización de AFE que tenemos en cuenta para su valoración. En los resultados, se presenta el porcentaje de uso de las prácticas valoradas, comparando los distintos periodos de tiempo estudiados. Siguiendo el trabajo previo de Ledesma et al. (2019), las prácticas aparecen divididas en cuatro grandes categorías: (a) relativas al diseño del estudio y de la escala de medida; (b) sobre los aspectos previos al AFE; (c) sobre las prácticas de realización del AFE y de la fiabilidad; y (d) relativas a la comunicación de resultados en el informe de investigación. Finalmente, se debaten los resultados obtenidos comparando lo encontrado en el presente trabajo con el estudio de Ledesma et al. (2019) y poniendo en relación los problemas de los malos usos de la técnica con la validez de los tests.

Método

Se evaluaron todos los artículos de la RIDEP en los que se había aplicado un AFE desde la publicación del artículo de Ledesma et al. (2019) hasta la redacción de este estudio, es decir, desde el Volumen 3 del 2018 hasta el Volumen 3 del año 2022. Se identificaron un total de 72 artículos que contenían un total de 83 AFE. Esta discrepancia se debe a que, en varios artículos, se

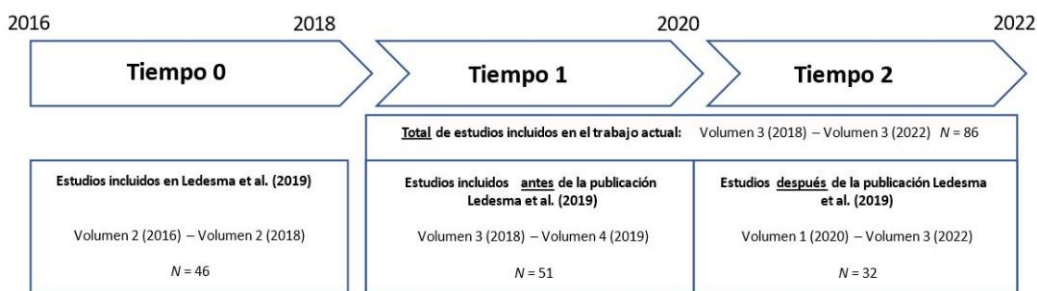


Figura 1. División de los artículos valorados en función del tiempo

realizaron más de un AFE para la misma o distinta escala sobre los cuales los autores aportaron información. Se analizaron todos los AFE por separado (N=83). No se tuvieron en cuenta estudios en los que el modelo teóricamente correcto fuera formativo y se aplicase Análisis de Componentes Principales (ACP).

El análisis llevado a cabo parte del realizado en Ledesma et al. (2019) y se completa con una ampliación sobre varios aspectos. En primer lugar, se analizaron cuestiones de diseño, como el tamaño muestral, la inclusión de AFE y AFC dentro del mismo artículo y las características propias de la escala (número de ítems de la escala, la ratio entre el tamaño muestral y el número de ítems, el número de ítems por factor y el número de categorías de respuesta). En segundo lugar, se abordaron las cuestiones ligadas con los aspectos previos al AFE y de adecuación de los datos (descriptivos de los ítems, índice *KMO* y valores perdidos). En tercer lugar, se observaron los aspectos de realización de AFE (método de estimación de parámetros, de retención y rotación de factores), junto con los índices de fiabilidad. Finalmente, se registraron aspectos vinculados al informe de investigación que ayudan a mejorar la reproducibilidad del estudio y replicabilidad de los hallazgos, por ejemplo, el software, la matriz de correlaciones, la correlación entre factores, la matriz de pesos factoriales e información sobre los resultados del modelo (porcentaje de varianza explicada e índices de ajuste).

Teniendo en cuenta que el artículo de Ledesma et al. (2019) se publicó en el Volumen 3 del 2019, se decidió mostrar la evolución de las prácticas de AFE clasificando la información en tres momentos temporales: (a) Tiempo 0, que contiene los datos recogidos por Ledesma et al. y que incluye las publicaciones de la RIDEP desde el Volumen 2 de 2016 hasta el Volumen 2 de

2018 (ambos incluidos); (b) Tiempo 1, que contiene los datos recogidos en esta investigación y que incluye publicaciones entre, incluyendo ambos, el Volumen 3 de 2018 (Volumen siguiente al registrado en Tiempo 0) hasta el Volumen 4 de 2019 (un Volumen después de la publicación de los datos del Tiempo 0); y (c) Tiempo 2, que contiene los datos recogidos en esta investigación con publicaciones desde el Volumen 1 del 2020 hasta el Volumen 3 del 2022 (ambos incluidos). Además, como resumen del estudio actual también se muestran los datos de los 83 AFE tomados en conjunto. Nótese que los autores de los artículos del Tiempo 1 no han podido hacerse eco del estudio de Ledesma et al. (2019).

Resultados

La evolución detallada de las prácticas de realización de AFE a lo largo del tiempo (desde 2016 hasta 2022) se puede ver en las Tablas 1, 2, 3 y 4, que muestran los porcentajes desagregados por las distintas prácticas estudiadas y por los tres momentos temporales. A continuación, destacamos algunos de los resultados más interesantes por sus consecuencias para la interpretación del AFE. La descripción de las prácticas se encuentra dividida en función de las categorías mencionadas anteriormente: elementos del diseño del estudio y de la escala de medida, aspectos previos al AFE, elementos propios del AFE y de la fiabilidad y aspectos de comunicación de resultados en el informe.

En relación a los elementos del diseño (Tabla 1), el 85.5% de todos los estudios registrados superan los 200 sujetos para el AFE (*Mdn*=363.5; rango=[79 - 2055]). El 44.6% realizaron solo AFE y, de entre aquellos que hicieron AFE y AFC, el 69.6% indicaron que los análisis se hicieron con muestras compuestas por distintos sujetos. Sobre

Tabla 1. Evolución de las Prácticas sobre el Diseño del Estudio y de la Escala

	Ledesma et al.		Estudio actual		Total N=83
	Tiempo 0 N=46	Tiempo 1 N=51	Tiempo 2 N=32		
	%	%	%	%	
Tamaño muestral					
< 200	11.1	11.8	12.5	12.1	
200 - 300	26.7	17.7	40.6	26.5	
301 - 600	40.0	37.3	34.3	36.1	
> 601	22.2	29.4	12.4	22.9	
No informa	-	3.9	0.00	2.4	
Uso de AFE Y AFC					
Solo AFE	63.1	35.3	59.4	44.6	
Con distinta muestra	-	41.2	34.4	38.6	
Resto	-	23.5	6.2	16.9	
Ratio n/variables					
<10	50.0	17.65	18.8	18.1	
Ratio ítems por factor					
Menos de 5 en algún factor	-	35.3	43.8	38.6	
Más de 5 ítems en todos	-	62.8	53.1	59.0	
No indica	-	2.0	3.1	2.4	
Categorías de respuesta					
3	-	9.8	0.0	6.0	
4	-	17.7	25.0	20.5	
5	-	37.3	65.6	48.2	
6	-	9.8	0.0	6.0	
7	-	13.7	6.3	10.8	
No indica	-	11.8	3.1	8.4	

Tabla 2. Evolución de las Prácticas sobre los Aspectos Previos al Análisis

	Ledesma et al.		Estudio actual		Total N=83
	Tiempo 0 N=46	Tiempo 1 N=51	Tiempo 2 N=32		
	%	%	%	%	
Análisis de descriptivos					
Analizan e informan	13.0	45.1	68.8	54.2	
Mención sin informar	9.0	7.8	3.1	6.0	
No nombran	78.0	47.1	28.1	39.8	
Información de los ítems					
Índices de discriminación	-	21.6	56.3	34.9	
Alfa si se elimina el elemento	-	11.8	28.1	18.1	
Media	-	43.1	62.5	50.6	
Desviación típica	-	43.1	62.5	50.6	
Asimetría	-	31.4	50.0	38.6	
Curtosis	-	31.4	50.0	38.6	
KMO					
<.80	2.0	11.8	12.5	12.1	
>.80	74.0	72.6	68.8	71.1	
No informan	24.0	15.7	18.8	16.9	
Información sobre					
Valores perdidos	-	13.7	6.3	10.8	

el diseño de las escalas, el 81.9% mostraban un ratio de sujetos por ítems superior a 10, el 38.6% tenía menos de cinco ítems en algún factor y el número de categorías de respuesta más frecuente fue cinco.

En relación a los aspectos de análisis previos al AFE (Tabla 2), el 54.2% de los estudios informaron de estadísticos descriptivos de los ítems. Lo más frecuente fue encontrar información relativa a la diada media y desviación típica (50.6%) y a la otra diada asimetría y curtosis

(38.6%). El 83.1% de los estudios informó del índice *KMO*, que tuvo un valor de la mediana igual a .87 y un rango de entre .51 y .97. Solo el 10.8% de los estudios indicaron alguna evaluación para detectar la presencia de valores perdidos o tomar decisiones al respecto.

Sobre los aspectos de realización del AFE (Tabla 3), el método de estimación más utilizado fue componentes principales (32.5%), seguido de ULS (19.3%). El 8.4% de los estudios registrados no informaron del método de estimación utilizado,

Tabla 3. Evolución de las Prácticas sobre la Ejecución de los Análisis

	Ledesma et al.	Estudio actual		
	Tiempo 0 N=46	Tiempo 1 N=51	Tiempo 2 N=32	Total N=83
	%	%	%	%
Matriz				
Pearson	4.0	31.4	34.4	32.5
Policóricas	22.0	37.3	50.0	42.3
No informa	74.0	31.4	15.6	25.3
Estimación de parámetros				
Componentes principales	58.0	31.4	34.4	32.5
Ejes principales	-	9.8	18.8	13.3
Máxima Verosimilitud	-	19.6	12.5	16.9
ULS	-	19.6	18.8	19.3
WLSMV	-	2.0	0.0	1.2
DWLS	-	0.0	6.3	2.4
Minimum Rank	-	3.9	9.4	6.0
No indican	-	13.7	0.0	8.4
Decisión sobre la retención				
Scree test	36.0	9.8	25.0	15.7
Autovalor > 1	20.0	21.6	21.9	21.7
Análisis Paralelo	17.0	31.4	31.3	31.3
No indica	20.0	45.1	31.3	39.8
Se basa en KMO	-	2.0	0.0	1.2
Rotación				
Oblicuo	43.0	43.1	65.6	51.8
Varimax	48.0	29.4	21.9	26.5
No informan	4.0	13.7	3.1	9.6
Unidimensionales	-	13.7	12.5	13.3
Fiabilidad				
Solo Alfa	-	92.2	75.0	85.5
Solo Omega	-	0.0	6.3	2.4
Ambos	-	5.9	18.8	10.8
Ninguno	-	2.0	0.00	1.2

Tabla 4. Evolución de las Prácticas sobre la Información Publicada

	Ledesma et al.	Estudio actual		
	Tiempo 0 N=46	Tiempo 1 N=51	Tiempo 2 N=32	Total N=83
	%	%	%	%
Software usado en el AFE				
SPSS	58.7	56.9	62.5	59.0
Factor	17.0	23.5	21.9	22.9
R	-	2.0	12.5	6.0
Mplus	-	5.9	3.1	4.8
Stata	-	2.0	0.0	1.2
JASP	-	0.0	9.4	3.6
No nombran	24.0	23.5	3.1	15.7
Correlación entre factor				
Muestran	-	31.4	34.4	32.5
No informan	-	37.3	50.0	42.2
Unidimensionales/Varimax	-	31.4	15.6	25.3
Matriz Pesos				
Completa	29.0	47.1	65.6	54.2
Incompleta	44.0	35.3	31.3	33.7
No informan	27.0	17.7	3.1	12.1
Información sobre				
% Varianza explicada		68.6	71.9	69.9
Ajuste del modelo				
Solución sin modelo nulo (GFI)	-	7.8	12.5	9.6
Comparación con nulo (TLI o CFI)	-	7.8	15.6	10.8
Complejidad RMSEA o AIC o BIC	-	9.8	18.8	13.3
No informan	-	90.2	78.1	85.5

aunque el 100% los estudios del Tiempo 2 sí lo hicieron. En esta línea, el 69.9% no indicaron el tipo de matriz de correlaciones usada, siendo la de

correlaciones policóricas aquella de la que más se informó (20.3%). El método de retención de factores más usado fue el análisis paralelo

(31.3%), aunque se observó que el 39,8% de los estudios no lo indicaron. El tipo de rotación que más se utilizó fue el oblicuo (51.8%), siendo diversas rotaciones las usadas al efecto. Esta variedad de procedimientos de rotación fue exclusiva de la rotación oblicua, dado que cuando se optaba por una ortogonal el procedimiento usado siempre fue Varimax. En cuanto a la fiabilidad, el índice más usado fue el alfa de Cronbach (85.5%), combinándolo con alguna variante del coeficiente omega en el 10.8% de los estudios. Solo el 1.2% de los estudios no informaba de ningún tipo de índice de fiabilidad.

En relación a las características vinculadas al informe de investigación (Tabla 4), el software más utilizado fue SPSS (59%), aunque un 15.7% no llegó a informar al respecto. La mayoría de los estudios no informaron de las correlaciones entre factores en los modelos oblicuos (42.2%). El 54.2% de los estudios informaron de la matriz de pesos completa. Finalmente, el 14.5% de los estudios informó sobre índices de ajuste. Teniendo en cuenta los índices recomendados por Ferrando et al. (2022), el 9.6% informaron de la solución sin comparación con el modelo nulo (*GFI*), el 10.8% valoraron la comparación con el modelo nulo (*TLI* o *CFI*) y el 13.3% evaluaron la complejidad con *RMSEA* (13.3%).

Discusión

En junio de 2019 se publicó en el tercer Volumen de la RIDEP un estudio que valoraba las prácticas en las aplicaciones de AFE. En dicho trabajo, Ledesma et al. (2019) concluyen sugiriendo la necesidad de promover mejores prácticas. Según los autores, algunas de las decisiones de los investigadores que aplican AFE son cuestionables y llevan a procedimientos desaconsejados por los expertos. Además, registraron deficiencias relevantes en la comunicación de la información. Para superar estos problemas, propusieron una guía aplicada de buenas prácticas para autores y revisores. Conscientes de la complejidad de la situación, Ledesma et al. finalizaban con una invitación a trabajar para promover mejores prácticas entre toda la comunidad científica: autores, revisores y editores. Fruto de esa invitación, y de la voluntad de la política editorial de la RIDEP, nace este

trabajo que tiene por objetivo analizar la evolución de las prácticas para realizar AFE desde que se publicó la guía de buenas prácticas hasta el presente.

Los resultados del presente estudio, poniéndolos en relación con las conclusiones y guía aplicada de Ledesma et al. (2019), permiten observar una evolución favorable en algunas prácticas de realización del AFE. Por ejemplo, disminuye el uso del método de componentes principales (punto 3 de la guía de Ledesma et al. 2019) y aumenta el uso de la rotación oblicua (punto 5) y de métodos adecuados para la selección del número de factores (punto 4). Además, la comunicación de resultados ha mejorado en algunos aspectos, lo que favorece la reproducibilidad y replicabilidad. En este sentido, aumenta la información declarada sobre el método de estimación, el software, la matriz de correlaciones (punto 2) y la matriz de pesos (punto 6).

Sin embargo, a pesar de la evolución favorable, consideramos que quedan aspectos que siguen necesitando la atención de autores, revisores y revista. En primer lugar, uno de los principales puntos débiles comprende el conjunto de aspectos que tienen que ver con la toma de decisiones al planificar y realizar el AFE. Según nuestros resultados, el tamaño muestral de los estudios puede parecer adecuado para realizar un AFE (solo un 12,1% de los estudios utilizan muestras con un tamaño muestral menor a 200), pero no se puede llegar a esa conclusión si no se conocen dos aspectos: primero, la matriz de correlaciones utilizada (de Pearson o policóricas); y, segundo, si la distribución de los ítems no permite usar métodos robustos (i.e., se incumple el supuesto de normalidad). En este sentido, menos de un tercio de los estudios informó del tipo de matriz de correlaciones usada, y menos del 40% tuvo en cuenta la asimetría y curtosis de los ítems. Además, sobre el segundo aspecto hay que añadir que el número de categorías de respuesta puede determinar la adecuación del tamaño muestral utilizado (Ondé y Alvarado, 2020).

En línea con los resultados obtenidos por Ledesma et al. (2019), parece haber una serie de decisiones inadecuadas que se producen al mismo tiempo cuando se realiza un AFE. De este modo, un alto número de estudios utilizan, sin

justificación, el análisis de componentes principales (ACP) junto con la rotación Varimax al mismo tiempo se valen del criterio de Kaiser (autovalores mayores a uno) como método de retención de factores (puntos 3, 4 y 5 de la guía aplicada de Ledesma et al., 2019). Es frecuente observar este conjunto de procedimientos, denominado como “Little Jiffy”, en otros estudios de revisión (Frías-Navarro & Pascual, 2012; Goretzko et al. 2021; Izquierdo et al., 2014). Esto sugiere que sigue existiendo una laguna en la comprensión de los modelos que deben usarse para obtener evidencias de validez de un instrumento de medida psicológico. El ACP, al no fundamentarse en el modelo del factor común, no persigue explicar las relaciones entre ítems por medio de factores latentes (i.e., constructos). Por ello, el ACP no debe usarse con modelos reflectivos que son los usados en psicología para medir constructos. Además, al usar la rotación ortogonal se olvida que en psicología es poco realista esperar que los constructos no correlacionen moderadamente entre ellos (Fabrigar & Wegener, 2011). En todo caso, con un método oblicuo siempre se podrán detectar factores independientes observando el valor de la correlación entre los factores (Ferrando et al., 2022).

Aunque se siguen cometiendo estos errores, existe una disminución de malas prácticas en la selección del método de estimación (decremento de ACP) y en la selección de métodos de rotación (aumento de la rotación oblicua). Esta mejora en las aplicaciones del AFE puede estar relacionada con el auge del uso de softwares como R y Factor, cuyas opciones por defecto difieren de SPSS. Por ejemplo, en una aplicación concreta se pudo ver que se utilizó SPSS y no se informó de los motivos para seleccionar el número de factores, se utilizaba la rotación Varimax existiendo una correlación entre factores de .87 y numerosas cargas cruzadas por encima de .64. En este punto creemos importante recalcar que no sugerimos que haya paquetes estadísticos buenos o malos, sino que las opciones de análisis que aparecen inicialmente tienden a ser vistas como adecuadas y a ser utilizadas sobre todo cuando el conocimiento técnico de los investigadores no es suficiente. Así, se han encontrado aplicaciones de AFE con factores representados por dos e incluso

solo un ítem, seguramente debido a que se han extraído tantos factores como autovalores hubiera por encima de uno.

En lo que respecta al análisis de fiabilidad, es reseñable que en la mayoría de estudios se informó de algún tipo de índice, reduciéndose así la práctica inadecuada de inducir la fiabilidad de la escala, es decir, comunicar la fiabilidad obtenida en aplicaciones anteriores (Aguayo et al., 2011; Sánchez-Meca et al., 2021). Esto puede sugerir una mejor comprensión de esta propiedad psicométrica, que se refiere a las puntuaciones obtenidas en una aplicación concreta del test y no al test en sí mismo. Sin embargo, el índice de fiabilidad reportado mayoritariamente fue el alfa de Cronbach, que requiere para su correcto funcionamiento el cumplimiento de supuestos como la unidimensionalidad y la tau-equivalencia de escala junto con la normalidad de los ítems. Estos supuestos raramente se comprueban (en ninguno de los estudios analizados se informaba) y cuando se hace, en pocas ocasiones se cumplen (Flora, 2020; Gignac, 2014, Triziano-Hermosilla y Alvarado, 2016; Toro et al. 2022). Como hemos comentado anteriormente acerca del AFE, las opciones por defecto en los softwares pueden estar relacionadas en este caso con el sobreuso del coeficiente alfa (v.g., en SPSS no se puede calcular el coeficiente omega sin usar macros). También se debería indicar si el coeficiente fue calculado antes o después del resultado del análisis factorial. Con relación al coeficiente omega, tampoco se debería considerar como la panacea de la fiabilidad. Existen distintos coeficientes (total, jerárquico, subescala), que deben ser usados en función del tipo de modelo propuesto (unidimensional, multifactorial). En nuestro caso, ningún artículo mencionó el tipo de coeficiente ni su motivación.

El segundo de los puntos débiles tiene que ver con la forma de comunicar los resultados en el informe de investigación. Ya se ha comentado previamente que la ausencia de información sobre la matriz de correlaciones utilizada condiciona la evaluación de idoneidad del tamaño muestral. Además, se han observado otras omisiones relevantes al indicar el número de categorías de respuesta (el 8.4% de los estudios analizados no lo indicaron), el software utilizado, el método de estimación, el de selección del número de factores

y el tipo de rotación. Incluso a pesar de que hubo un aumento en la comunicación de AP, que es una técnica recomendada para la retención de factores, tampoco es frecuente informar de cuestiones más técnicas como el tipo de AP. Adicionalmente, se observó que esta información, cuando aparecía en el informe, frecuentemente no estaba ubicada correctamente. En este sentido, recomendamos que esta información, relativa al procedimiento de realización del AFE, se comunique en la sección de método y no en la de resultados. De esta forma, el lector puede saber rápidamente el proceso de toma de decisiones que realizaron los autores antes de hacer el AFE y evita que sea necesario explorar el apartado de resultados para encontrar ciertos aspectos como el método de estimación usado.

Es tan importante tomar decisiones correctas al realizar los análisis estadísticos como informar con precisión y exhaustividad suficientes de los procedimientos utilizados (por ejemplo si la administración fue online o física, Freiberg-Hoffmann y Romero-Medina, 2021). y de los resultados obtenidos. La transparencia en la comunicación científica no solo permite al lector formarse un juicio sobre las evidencias encontradas, sino que favorece la reproducibilidad de los estudios y la replicabilidad de los hallazgos (Gagnier et al., 2021; Ioannidis, 2014). Conseguir altas tasas de reproducibilidad es un pilar fundamental para aumentar la credibilidad de los estudios, sin embargo, éstas se sitúan por debajo del 100% en el área de Psicología (Nosek et al., 2022). Aunque no se disponen de estudios en el campo de la psicometría, a la luz de los resultados obtenidos es bastante probable nos encontremos con tasas altas de reproducibilidad en estudios en los que se aplica AFE. De hecho, los resultados del AFE pueden sufrir variaciones relativamente importantes cuando se aplican a distintas muestras, por ejemplo, en los estudios de adaptación a otras lenguas o en los estudios transculturales, especialmente, cuando los estudios a comparar se realizaron en un intervalo temporal de varios años.

Esta falta de transparencia y precisión en el informe de investigación se podría caracterizar como parte de las Prácticas Cuestionables de Investigación (PCI). Como han indicado otros autores (Aguayo, 2018; Picho & Artino, 2016), las

PCI se pueden producir o bien antes de la investigación, con una revisión sesgada de la literatura o una mala elección de instrumentos de medida, o bien después, por la falta de transparencia en los resultados. En el caso del AFE, los autores deberían informar de toda toma de decisión relacionada con su aplicación y presentar los resultados completos, sin omitir información relevante. Así, se podrán diseñar estudios de reproducibilidad que permitan valorar si las variaciones encontradas en los resultados acerca de la estructura factorial eran producto de la heterogeneidad metodológica o del propio instrumento de medida.

De acuerdo con los resultados de este trabajo y otros previos (Frías-Navarro & Pascual, 2012; Goretzko et al., 2021; Lloret-Segura et al., 2014), sigue siendo necesario promover mejores prácticas en las aplicaciones de AFE. Por ello, insistimos en que es de vital importancia que autores y revisores acudan a la abundante literatura previa que contiene guías y decálogos para realizar AFE, disponible tanto en español (Ferrando & Lorenzo-Seva, 2014; Ferrando et al., 2022; Lloret-Segura et al., 2014) como en portugués (Damásio, 2012). En ella encontrarán la información necesaria para tomar las mejores decisiones. La aplicación correcta del AFE no es solo una cuestión meramente técnica sino que de su mala aplicación puede llegarse a conclusiones erróneas que atenten a la validez de los instrumentos de medida. La validez se refiere al grado en que la evidencia y la teoría apoyan las interpretaciones de las puntuaciones de los instrumentos para un uso concreto (AERA et al., 2014). De esta forma, como consecuencia de un mal uso del AFE puede derivarse la eliminación de ítems por bajos pesos factoriales que de haber seguido el procedimiento recomendado (v.g., matrices policóricas o de Pearson) habrían tenido una buena discriminación. Esto tiene consecuencias en la medida del constructo, ya que se puede incurrir en la subrepresentación (Messick, 1995), es decir, eliminar ítems que son fundamentales para una adecuada representación del contenido completo del constructo medido. Igualmente, mantener ítems que deberían haber sido eliminados puede formar factores irrelevantes con respecto a los contenidos del constructo (Elosua, 2003). Otra de las consecuencias de la

aplicación inadecuada del AFE es obtener estructuras factoriales erróneas, en las que por problemas en las distribuciones (falta de normalidad) se produzcan agrupaciones de los ítems por cuestiones ajenas al constructo, por ejemplo, por su nivel de dificultad. Dada una estructura factorial equivocada la aplicación del test puede ocasionar consecuencias indeseadas derivadas del uso de un componente irrelevante para el constructo (v.g., un factor debido al agrupamiento de ítems redactados de forma positiva frente a otros de forma negativa).

Un mal uso de la técnica factorial puede llevar al investigador a soluciones no replicables e incorrectas y a la toma de decisiones equivocadas que atenten a la validez de los instrumentos de medida. En este trabajo hemos observado que los autores siguen usando procedimientos desaconsejados para realizar el AFE y que los informes de investigación no son lo suficientemente precisos y completos. Por ello recomendamos encarecidamente la aplicación del AFE de forma rigurosa, con análisis preliminares del cumplimiento de los supuestos del modelo. Además, creemos necesario reseñar las siguientes sugerencias: (a) la publicación de un decálogo o normas de uso del AFE por parte de la RIDEP y su uso en los procesos de revisión por pares; (b) la inclusión de, al menos, un experto en medición para revisar los artículos con aplicaciones de AFE y AFC; (c) contextualizar la aplicación del AFE en un estudio concreto como uno de los pasos, y tipos de evidencia, en el marco del proceso de validación de un test, incluyendo una mención a futuros estudios que se centren en otras evidencias de validez; y (d) una mayor atención a la transparencia y cuidado con las omisiones en el informe de investigación. Esperamos con ello poder contribuir al esfuerzo conjunto por mejorar las prácticas de realización del AFE.

Referencias

- Aguayo, R. (2018). La investigación en el síndrome de burnout: Reflexión crítica desde una perspectiva metodológica. *Apuntes de Psicología*, 36(1-2), 93-100.
- Aguayo, R., Vargas, C., Emilia, I., & Lozano, L. M. (2011). A meta-analytic reliability generalization study of the Maslach Burnout Inventory. *International Journal of Clinical and Health Psychology*, 11(2), 343-361.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.
- Damáso, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica: Interamerican Journal of Psychological Assessment*, 11(2), 213-228.
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315-321.
- Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.
- Ferrando, P. J., & Anguiano-Carrasco, I. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33.
- Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: Algunas consideraciones adicionales. *Anales de Psicología*, 30, 6.
- Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñiz, J. (2022). Decálogo para el Análisis Factorial de los Ítems de un Test. *Psicothema*, 34.1, 7-17.
<https://doi.org/10.7334/psicothema2021.456>
- Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, 7(1), 45-52.
<https://doi.org/10.1177/1948550615612150>
- Flora, D. B. (2020). Your coefficient Alpha is probably wrong, but Which Coefficient Omega Is Right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484-501.
<https://doi.org/10.1177/2515245920951747>
- Freiberg-Hoffmann, A., & Romero-Medina, A. (2021). Approaches and Study Skills Inventory for Students: Comparación de las propiedades psicométricas entre las versiones

- de lápiz-papel y online en estudiantes universitarios. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 4(64), 17-30.
<https://doi.org/10.21865/RIDEP61.4.11>
- Frías-Navarro, D., & Pascual, M. (2012). Prácticas del análisis factorial exploratorio (afe) en la investigación sobre conducta del consumidor marketing. *Suma Psicológica*, 19(1), 47-58.
- Gagnier, J. J., Lai, J., Mokkink, L. B., & Terwee, C. B. (2021). COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research*, 30(8), 2197-2218.
<https://doi.org/10.1007/s11136-021-02822-4>
- Gignac, G. E. (2014). On the inappropriateness of using items to calculate total scale score reliability via Coefficient Alpha for multidimensional scales. *European Journal of Psychological Assessment*, 30(2), 130-139.
<https://doi.org/10.1027/1015-5759/a000181>
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 40(7), 3510-3521.
<https://doi.org/10.1007/s12144-019-00300-2>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies: A typology of reviews. *Health Information & Libraries Journal*, 26(2), 91-108.
<https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- 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*, 3(2), 166-184.
<https://doi.org/10.1177/2515245919882903>
- Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Medicine*, 11(10), e1001747.
<https://doi.org/10.1371/journal.pmed.1001747>
- Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema*, 26.3, 395-400.
<https://doi.org/10.7334/psicothema2013.349>
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6-23.
<https://doi.org/10.1037/a0014694>
- John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524-532.
<https://doi.org/10.1177/0956797611430953>
- Kline, R. B. (2015). *Principles and practice of structural equation modelling*. Guilford Publications.
- Ledesma, R., Ferrando, P., & Tosi, J. (2019). Uso del Análisis Factorial Exploratorio en RIDEP. Recomendaciones para autores y revisores. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 52(3), 173-180.
<https://doi.org/10.21865/RIDEP52.3.13>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169.
<https://doi.org/10.6018/analesps.30.3.199361>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
<https://doi.org/10.1037/0003-066X.50.9.741>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). *Replicability, robustness, and reproducibility in psychological science*, 33.
<https://doi.org/10.1146/annurev-psych-020821114157>
- Ondé, D., & Alvarado, J. M. (2020). Reconsidering the conditions for conducting confirmatory factor analysis. *The Spanish Journal of Psychology*, 23.
<https://doi.org/10.1017/SJP.2020.56>

- Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. Scotts Valley. <https://doi.org/10.4135/9781412995627.d8>
- Pérez, E. R., & Medrano, L. (2010). Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista Argentina de Ciencias del Comportamiento*, 2(1), 58-66.
- Picho, K., & Artino, A. R. (2016). 7 Deadly sins in educational research. *Journal of Graduate Medical Education*, 8(4), 483-487. <https://doi.org/10.4300/JGME-D-16-00332.1>
- Reio, T. G., & Shuck, B. (2014). Exploratory factor analysis: Implications for theory, research, and practice. *Advances in Developing Human Resources*, 17(1), 12-25.
- Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., López-Pina, J. A., Blázquez-Rincón, D. M., López-Ibáñez, C., & López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods*, 12(4), 516-536. <https://doi.org/10.1002/jrsm.1487>
- Saunders, R., & Savulescu, J. (2008). Research ethics and lessons from Hwanggate: What can we learn from the Korean cloning fraud?. *Journal of Medical Ethics*, 34(3), 214-221. <http://doi.org/10.1136/jme.2007.023721>
- Schreiber, J. B. (2008). Core reporting practices in structural equation modeling. *Research in Social and Administrative Pharmacy*, 4(2), 83-97. <https://doi.org/10.1016/j.sapharm.2007.04.003>
- Thurstone, L. L. (1935). *The vectors of mind*. University of Chicago Press.
- Thurstone, L. L. (1947). *Multiple-factor analysis*. The University Chicago Press.
- Toro, R., Peña-Sarmiento, M., Avendaño-Prieto, B. L., Mejía-Vélez, S., & Bernal-Torres, A. (2022). Análisis empírico del Coeficiente Alfa de Cronbach según opciones de respuesta, muestra y observaciones atípicas. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 2(63), 17-30. <https://doi.org/10.21865/RIDEP63.2.02>
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's Alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00769>
- Van Dijk, D., Manor, O., & Carey, L. (2014). Publication metrics and success on the academic job market. *Current Biology*, 24(11), R516-R517. <https://doi.org/10.1016/j.cub.2014.04.039>