Liberabit, 2025, 31(1), e963 (enero - junio)
ISSN (Digital): 2223-7666
https://doi.org/10.24265/liberabit.2025.v31n1.963

Una guía amigable de psicometría: análisis factorial exploratorio

A Friendly Guide to Psychometrics: Exploratory Factor Analysis

Sergio Dominguez-Lara*
Instituto de Investigación FCCTP, Universidad de San Martín de Porres, Lima, Perú

Recibido: 13 de setiembre de 2024 Aceptado: 19 de febrero de 2025

Resumen

Antecedentes: uno de los procedimientos más destacados para el análisis de la estructura interna de un instrumento de evaluación psicológica es el análisis factorial exploratorio. No obstante, su uso muchas veces se sustenta en procedimientos que ya fueron superados, tanto a nivel teórico como metodológico, y cuya aplicación en esas condiciones representaría una seria amenaza para la validez científica de los hallazgos presentados. **Estado del arte**: se presenta de forma amigable al lector neófito en temas de psicometría las bases conceptuales y metodológicas del análisis factorial exploratorio, desde la selección de la matriz de correlaciones apropiada hasta la denominación de los factores, a fin de orientarlo en el proceso de toma de decisiones. **Conclusiones**: la toma de decisiones dentro de cada análisis puede conducir a resultados más o menos confiables desde el punto de vista psicométrico, por lo que es fundamental realizar un plan de análisis que contemple todas las características del instrumento que se estudiará.

Palabras clave: psicometría, validez, análisis factorial exploratorio, factor.

Abstract

Backgrounds: One of the most important procedures for the analysis of the internal structure of a psychological assessment instrument is exploratory factor analysis. However, its use is often based on procedures that have already been surpassed, both at the theoretical and methodological level, and whose application in these conditions would represent a serious threat to the scientific validity of the findings presented. **State of art**: The conceptual and methodological bases of exploratory factor analysis are presented in a user-friendly way to the novice reader in psychometric topics, from the selection of the appropriate correlation matrix to the naming of the factors, in order to guide him in the decision-making process. **Conclusion**: Decision-making within each analysis can lead to results that are more or less reliable from a psychometric point of view, so it is essential to carry out an analysis plan that takes into account all the characteristics of the instrument to be studied.

Keywords: psychometrics, validity, exploratory factor analysis, factor.

Para citar este artículo:

Dominguez-Lara, S. (2025). Una guía amigable de psicometría: análisis factorial exploratorio. *Liberabit, 31*(1), e963. https://doi.org/10.24265/liberabit.2025.v31n1.963

* sdominguezmpcs@gmail.com

Introducción

La medición se puede definir como la asignación de números a determinadas características de objetos, personas o eventos según reglas establecidas cuya validez es susceptible de ser evaluada empíricamente (Magnusson, 1972; Stevens, 1951). Por otro lado, una definición más completa indica que cuantificar una propiedad de un sistema es *proyectar* el conjunto de grados de esa propiedad sobre un conjunto de números a fin de que el orden y la espaciación de los números reflejen el orden y la espaciación de los grados de la propiedad (Bunge, 2004). En este sentido, queda claro que si esta *propiedad* (e.g., longitud) pertenece a un *objeto* observable (e.g., mesa) es más fácil de medir. Sin embargo, cuando se trata de la medición de fenómenos psicológicos la tarea no es sencilla.

Para empezar, se sabe que en psicología los atributos estudiados (variables psicológicas) no tienen existencia física ni son directamente observables, por lo que reciben la denominación de *constructos*. En vista que estos constructos no tienen una existencia física y sobrepasan la observación, el procedimiento para medirlos es indirecto. Esta situación se extiende a las características (indicadores) del constructo. Sin embargo, si bien se espera que estos constructos tengan una definición conceptual y operacional, el panorama en la psicología actual es distinto, ya que un solo constructo (e.g., estrés) puede tener diferentes definiciones procedentes de distintas teorías explicativas (que proporcionan alguna definición conceptual) y cada una con su respectivo instrumento de medida o test psicológico (que se derivan de la definición operacional propuesta por el investigador).

En cualquier caso, y extrapolando la definición de medición brindada anteriormente, se espera que el aumento o disminución de las puntuaciones obtenidas en el test psicológico reflejen mayor o menor presencia del constructo en la persona evaluada. Sin embargo, es necesario para la psicometría actual explorar las propiedades métricas (validez y confiabilidad) de esos instrumentos de evaluación a fin de respaldar las interpretaciones que realiza el evaluador a partir de las puntuaciones obtenidas en el instrumento utilizado.

La definición vigente (American Educational Research Association et al., 1999, 2014) concibe a la validez como el grado en que las evidencias teóricas y empíricas brindan soporte a las interpretaciones basadas en las puntuaciones de los tests psicológicos (cuestionarios, escalas o inventarios). Esta definición tiene algunos puntos que es necesario ampliar.

Por ejemplo, la validez se refiere a un *grado* (parte de un *continuum*); no se habla en términos absolutos como, por ejemplo, «el test psicológico tiene validez» o «no tiene validez» ni numéricos como «la validez del test es de .95», «la validez del test es estadísticamente significativa (p < .05)» o «la validez fue de r = .60». Estas afirmaciones son poco precisas porque el enfoque actual de validez no la concibe como la cantidad de *algo* ni una dicotomía (*tiene* o *no tiene*); por el contrario, el investigador analiza en conjunto los resultados de diversas estrategias orientadas a la obtención de evidencias de validez y, en base en ello, podrá determinar si lo encontrado justifica (o no) la interpretación de las puntuaciones en torno al constructo evaluado.

Cuando se mencionan a las *evidencias teóricas y empíricas*, se sugiere que se trata de un proceso sustentado por una metodología, pero a su vez que cuenta con una base teórica que respalda tanto al constructo como a los ítems (Muñiz & Fonseca-Pedrero, 2019). De esto se infiere que, si un test fue creado para evaluar un constructo, pero no cuenta con un respaldo teórico, aunque tenga indicadores cuantitativos o cualitativos que podrían considerarse como *buenos* o *favorables* no sería susceptible de interpretación teórica porque no respondería a dos preguntas sencillas: *¿Qué se evalúa?* y *¿Cómo se interpretan esas puntuaciones?* Por otro lado, si el test es creado con un modelo teórico establecido, es necesario saber si esa teoría es compatible con el grupo en el que se pretende usar el test (Chávez-Ventura et al., 2025), y si el constructo, a la luz de esa teoría, está bien representado (ítems) en el nuevo grupo de estudio. Estos son algunos aspectos clave que debe explorar la persona que desea analizar psicométricamente un test.

Estado del arte

Evidencias de validez basadas en la estructura interna: análisis factorial exploratorio

Como se indicó anteriormente, la concepción actual de validez implica conferir esta cualidad a la medición de un constructo mediante diversas estrategias. No obstante, las clasificaciones antiguas de validez consideraban que el tipo de validez denominado *validez de constructo* estaba íntimamente ligado al análisis factorial, y existía respaldo para ello.

Por ejemplo, si alguien desea evaluar la autoeficacia académica, definida como las creencias de cada individuo sobre las capacidades propias para organizar y ejecutar acciones requeridas en el manejo y afrontamiento de situaciones relacionadas con ámbitos académicos (Dominguez-Lara, 2014), puede utilizar la Escala de Autoeficacia Percibida Específica para Situaciones Académicas (EAPESA; Palenzuela, 1983). La redacción de los ítems de la EAPESA (que pueden ser respondidos en términos de *Nunca* [1], *Algunas veces* [2], *Bastantes veces* [3], o *Siempre* [4]) están claramente orientados a este constructo a nivel de contenido (e.g., *Me considero lo suficientemente capacitado para enfrentarme con éxito a cualquier tarea académica*; entre otros), y la evidencia obtenida así lo confirma (Dominguez-Lara et al., 2012). Entonces, ¿es necesario obtener más evidencia?, ¿cómo se puede saber que las respuestas subyacentes a esos ítems son «influenciadas» (conducidas, motivadas, generadas, etc.) por el constructo que se está evaluado (autoeficacia académica)? Se podría realizar analizando las *respuestas de los examinados*.

En este orden de ideas, un supuesto clave es que, a mayor puntuación en el test, mayor presencia en el constructo; o pensado al revés, una mayor presencia del constructo en la persona *generará* respuestas a los ítems más favorables al constructo (e.g., opciones más elevadas, en el caso de escalamiento Likert). Por ejemplo, si se tienen los ítems *Me considero lo suficientemente capacitado para enfrentarme con éxito a cualquier tarea académica*, *Me siento con confianza para abordar situaciones que ponen a prueba mi capacidad académica* y *Creo que soy una persona capacitada y competente en mi vida académica*, una persona (A) con firmes creencias de autoeficacia académica podría responder con 4, 3 y 4 (atendiendo al formato de respuesta mencionado en el párrafo anterior), mientras que una persona (B) con menos confianza en sus capacidades respondería, probablemente, con 2, 1 y 1. ¿Qué ocurrió? Se proponen dos escenarios.

El primero es que el estímulo (ítem) generó una autoevaluación de las propias capacidades y, con base en esta introspección, la persona A se juzgó con la capacidad de abordar las exigencias de su contexto, mientras que B, no. El segundo escenario es que escogieron una opción que refleje mejor aquella *valoración personal* que realizaron al leer el ítem y la marcaron (respuesta).

Entonces, ¿cómo se puede *ver* al constructo *reflejado* en esas respuestas? Atendiendo a la definición de *autoeficacia académica* es poco probable que la persona A marque 4 al primer ítem y 1 al segundo, así como que la persona B marque 1 en el primer ítem y 4 en el segundo. Adicionalmente, se espera que la puntuación de los ítems varíe de una persona a otra (varianza diferente de cero), así como una *asociación directa* entre las respuestas de los ítems (asumiendo que no hay ítems invertidos), lo que puede expresarse numéricamente en un *coeficiente de correlación* (con el requisito indispensable de contar con mucho más que dos personas, como se habrá supuesto). Por ejemplo, la correlación entre las respuestas a los ítems 1 y 2 podría expresarse como r = .60. Del mismo modo, este *coeficiente de correlación* informaría que, por ejemplo, el ítem 1 se relaciona positivamente con el ítem 2, 3, 4 … y 9; que el ítem 2 se relaciona positivamente con el ítem 3, 4 ... y 9; y así, con todos, es posible obtener una matriz de correlaciones.

De esto podría calcularse el *coeficiente de determinación* asociado a cada correlación (o R^2), el cual indicaría, por ejemplo, que el ítem 1 y el ítem 2 tienen un 36% de varianza compartida, es decir, que tienen *algo en común*; y si se repite el ejercicio (es decir, si se calcula con cada par de ítems), se podría pensar que *aquello* que comparten los ítems es la variable latente o *constructo*, pero para llegar a esa conclusión es necesario recurrir a un método estadístico multivariado: el análisis factorial.

Análisis factorial exploratorio: una definición

De forma genérica, el análisis factorial puede definirse como un conjunto de métodos estadísticos multivariados de interdependencia de variables observables (e.g., ítems) cuyo objetivo es identificar qué variables latentes (e.g., factores) subyacen a un conjunto de datos (Ferrando & Anguiano-Carrasco, 2010). En otras palabras, se desea simplificar la información, es decir, pasar de tener un número grande de variables observadas (e.g., 9 ítems de la EAPESA) a obtener pocas variables latentes (e.g., factor *autoeficacia académica*) para facilitar la interpretación, asumiendo la existencia de variables latentes que son capaces de explicar la magnitud de la asociación entre las variables observadas.

Es conveniente resaltar que se presenta como una *técnica de análisis multivariado de interdependencia* porque analiza de forma conjunta la relación de un gran número de variables observadas con base en la matriz de correlaciones que informa esa relación. Es necesario mencionar que si bien este método fue creado por el psicólogo Spearman (1904), no es exclusivo de la psicometría, ya que se aplica a diversas áreas del conocimiento (e.g., biología, economía, etc.).

Existen dos aproximaciones del análisis factorial mayormente aceptadas: el *análisis factorial exploratorio* (AFE) y *análisis factorial confirmatorio* (AFC). El AFE parte de un supuesto: se conoce que existen variables latentes que explican la varianza de los ítems (Figura 1a), pero no se sabe *cuántas son* ni *qué ítems comparten* la influencia de esas variables latentes, por lo que su objetivo es determinarlo (Figura 1b). El AFC pone a prueba la existencia de una estructura factorial preestablecida, ya sea por determinada teoría que sustentó la construcción o porque la estructura interna fue definida en estudios previos (Nájera et al., 2023).

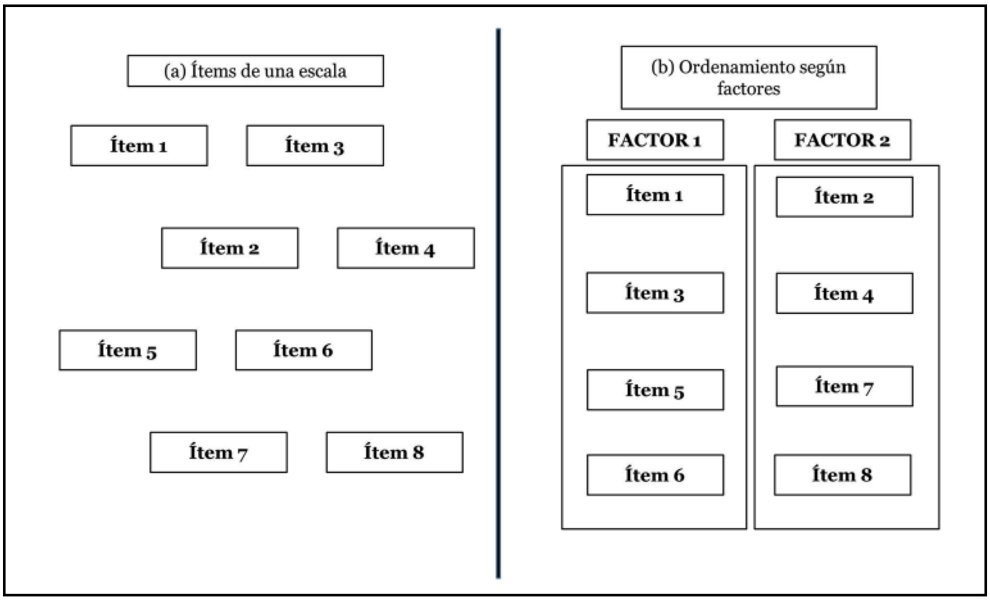
En síntesis, según la Figura 1, se tienen ocho ítems y no se conoce cuántos factores explican su variabilidad. Luego de los análisis, se encuentra que la variabilidad de los ítems 1, 3, 5 y 6 puede ser explicada por un factor (la relación es fuerte entre estos ítems), y la variabilidad de los ítems 2, 4, 7 y 8, por otro factor. Asimismo, para que estén diferenciados se necesitaría, por ejemplo, que la relación entre los ítems 1 y 3, sea más intensa que la relación entre los ítems 1 y 2 o que entre los ítems 2 y 3 (lo que significaría que los ítems 1 y 2 tienen poco en común).

En este orden de ideas, la varianza de los ítems tiene tres componentes: varianza compartida, varianza específica y varianza del error; o en sus nombres más conocidos: comunalidad, especificidad y varianza residual. De estas tres, las dos primeras constituyen la varianza verdadera (es decir, la que se puede interpretar). La *comunalidad* hace referencia a la proporción de varianza del ítem explicada por el factor común (o factores, si se trata de un test multidimensional), la *especificidad* se refiere a la varianza verdadera del ítem no atribuible al factor común, pero con un significado sustantivo, y la *varianza residual* que es atribuida a otras causas no conocidas por el investigador y que no tienen una interpretación. En consecuencia, la varianza no común (*especificidad* + *varianza residual*) recibe el nombre de *unicidad*.

Por ejemplo, cuando la persona responde un *Bastantes veces* [3] al ítem *Me siento con confianza para abordar situaciones que ponen a prueba mi capacidad académica* de la EAPESA, parte de esa respuesta la dio con base en su autoeficacia académica (comunalidad), otra parte en función de la valoración específica de su confianza en abordar los retos académicos (especificidad), y hay un porcentaje de su respuesta que tuvo como origen cualquier otra cognición ajena a la autoeficacia (residual). De hecho, este ítem presentó la más alta comunalidad en el primer estudio psicométrico de la EAPESA en Perú (Dominguez-Lara et al. 2012) con un .603, que se podría entender como «el 60.3% de la respuesta al ítem se explica por el constructo autoeficacia académica».

Figura 1

Estado inicial y final de los ítems en un AFE



Supuestos para realizar el análisis factorial exploratorio: datos

Es necesario indicar que antes de realizar cualquier AFE, existen una serie de condiciones que los datos deben cumplir. Existe abundante literatura al respecto (Pérez & Medrano, 2010; Ferrando & Anguiano-Carrasco, 2010; Lloret-Segura et al., 2014; Mavrou, 2015; Watkins, 2018), por lo que se presentarán de forma resumida en estos puntos para profundizar otros aspectos quizás un poco más relevantes, pero que no suelen tratarse.

En primer lugar, algunos textos clásicos sobre aplicación del análisis factorial en psicometría recomiendan usarlo solo cuando los ítems se consideran variables cuantitativas (> 6 opciones de respuesta; Verdam et al., 2016), aunque la mayor parte de instrumentos posee ítems entre dos (dicotómicos) y cinco (politómicos) opciones de respuesta y podrían ser considerados como variables nominales u ordinales, respectivamente.

Probablemente, la recomendación estuvo influida por dos motivos. El primero, que la matriz de correlaciones usada como base correspondía a coeficientes de correlación de Pearson, los cuales, como se sabe, deben calcularse (de preferencia) con variables cuantitativas. El segundo, que en ese momento de avance computacional solo se disponía de software con métodos diseñados para procesar variables cuantitativas (incluso hasta finales del siglo XX), pese a que ya existían las nociones y desarrollos teóricos sobre las correlaciones tetracóricas y policóricas, que correspondían a variables nominales y ordinales (e.g., Olsson, 1979), respectivamente (Ferrando & Lorenzo-Seva, 2014). Afortunadamente, existen diversos softwares estadísticos que permiten seleccionar métodos orientados a la naturaleza de la variable (LISREL, Mplus, EQS, etc.).

Pese a ser el primer punto, queda claro que la toma de decisiones sobre qué métodos emplear más adelante parte de un hecho aparentemente sin importancia: identificar en qué nivel de medición (nominal, ordinal o cuantitativo) están los ítems.

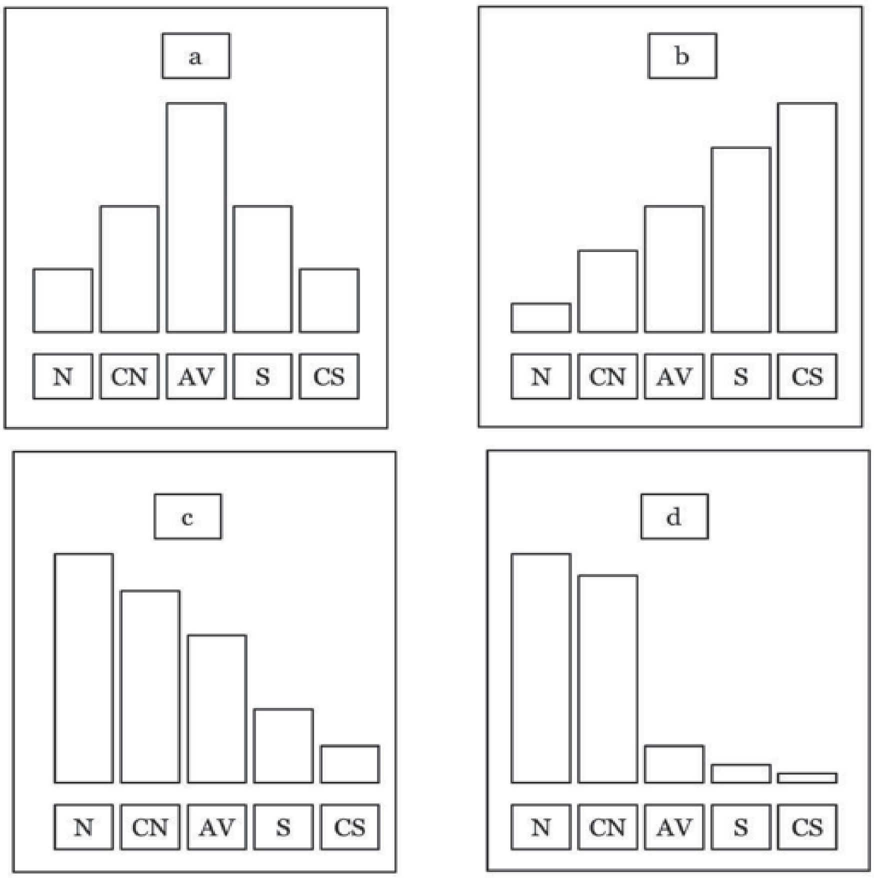
En segundo lugar, y siguiendo la tradición basada en variables cuantitativas, se espera que la distribución de los ítems se aproxime a la normalidad (Figura 2a). Esto es una exigencia poco razonable porque, como se conoce, la distribución esperada de las respuestas a los ítems puede variar dependiendo del tipo de constructo. Por ejemplo, si se trata de ítems asociados a bienestar (e.g., satisfacción con la vida), predominan las respuestas más «altas» y la distribución se torna asimétrica (Figura 2b). Por otro lado, si están vinculados al malestar (e.g., estrés), las respuestas más «bajas» son las que tienen mayor frecuencia, dando como resultado una distribución asimétrica con una orientación opuesta a la anterior (Figura 2c). Por último, si se refiere a constructos con poca manifestación, por ejemplo, en población general (e.g., experiencias perceptuales anómalas), se hace más pronunciada la asimetría (Figura 2d).

Esto último se vincula con una decisión habitual entre algunos investigadores: elegir el tipo de correlación según la cantidad *original* de opciones de respuesta, pero no con base en el *número de opciones respondidas*. Por ejemplo, determinado test puede tener ítems con siete opciones de respuesta, lo que alentaría al investigador a elegir las correlaciones Pearson como base para su análisis; sin embargo, si la cantidad de opciones *respondidas* fue cuatro en todos los ítems, debería usarse otro tipo de matriz de correlaciones (e.g., policórica) (Dominguez-Lara et al., 2020).

Entonces, según lo expuesto, existen algunas propuestas basadas en la magnitud de la asimetría ($g1$) y curtosis ($g2$) univariadas para evaluar la aproximación a la normalidad de los ítems. Por ejemplo, se ha sugerido que ambos estadísticos sean menores que 1.5 (Pérez & Medrano, 2010) o 2 (Gravetter & Wallnau, 2013), o algunos más flexibles, es decir, que la asimetría no sea mayor que 3 y la curtosis no sea mayor que 10 (Kline, 2016), o aceptar asimetrías de hasta 2 y curtosis de 7 (Finney & DiStefano, 2006). En resumen, una distribución asimétrica de los ítems justificaría el uso de matrices policóricas (Ferrando & Lorenzo-Seva, 2019). Por otro lado, también es frecuente el uso del coeficiente de curtosis multivariada de Mardia (1970) como una aproximación práctica al análisis de la normalidad multivariada. Se trata de un estadístico que evalúa el grado en que la distribución contiene casos extremos (*outliers*) teniendo como base la distancia de Mahalanobis (Victor-Edema, 2023), y si bien suele valorarse con base en la prueba estadística de la hipótesis nula (donde se esperaría un *p*-valor mayor que .05), es recomendable tener en cuenta su magnitud. En ese sentido, magnitudes por debajo de 70 indican el cumplimiento de este requisito (Ayán & Ruiz, 2008).

En tercer lugar, se espera que exista una relación significativa entre los ítems, pero que no exista *multicolinealidad*, es decir, correlaciones tan elevadas que sugieren solapamiento, las cuales podrían ser mayores que .70 (Joshanloo et al., 2016) o que .80 (Brown, 2015). En estos casos, una recomendación sería prescindir de aquel ítem que posea la más baja valoración en cuanto a su contenido (e.g., V de Aiken).

Figura 2
Distribución hipotética de un ítem



Nota. N = nunca; CN = casi nunca; AV = a veces; S = siempre; CS = casi siempre.

Software recomendado

Antes de continuar, es necesario mencionar que los softwares deben contemplar algunas características que permitan analizar a los ítems que componen los tests psicológicos. En ese sentido, un estudio muy detallado (Lloret-Segura et al., 2017) recomienda como mejor opción, aunque sin dejar de mencionar algunas posibilidades de mejora, el programa FACTOR (Ferrando & Lorenzo-Seva, 2017; Lorenzo-Seva & Ferrando, 2006, 2013) debido a que es un programa específico, flexible y de libre distribución. Además, de que está constantemente en actualización con los avances más recientes en análisis factorial.

Supuestos para realizar el análisis factorial exploratorio: adecuación de la matriz de correlaciones

En términos individuales, los ítems pueden cumplir con los criterios antes mencionados, pero en vista de que el AFE es una técnica multivariada que toma como insumo la matriz de correlaciones inter-ítem, es necesario conocer si es posible extraer la suficiente información de esta matriz para determinar la existencia de una variable latente. Para ello, son usados dos procedimientos: el test de esfericidad de Bartlett y el estadístico KMO (Kaiser, 1970).

El test de esfericidad de Bartlett es una prueba de hipótesis que evalúa si el conjunto de ítems está correlacionado de forma estadísticamente significativa o no. La hipótesis nula indica que *los ítems no están correlacionados significativamente* o, en otros términos, que la matriz de correlaciones observada (Figura 3a) se aproxima a una matriz de identidad, caracterizada por correlaciones nulas y valores de uno en la diagonal (Figura 3b). Se espera que la hipótesis nula no tenga respaldo ($p < .05$), para poder concluir que los ítems *están correlacionados* significativamente.

Figura 3

Matriz de correlaciones (a) y de identidad (b)

a

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1	.4	.5	.6	.3
Item 2	.4	1	.3	.3	.5
Item 3	.5	.3	1	.6	.5
Item 4	.6	.3	.6	1	.6
Item 5	.3	.5	.5	.6	1

b

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1	0	0	0	0
Item 2	0	1	0	0	0
Item 3	0	0	1	0	0
Item 4	0	0	0	1	0
Item 5	0	0	0	0	1

Sin embargo, esto es el paso inicial, ya que, si se cuenta con un tamaño muestral lo suficientemente alto, incluso con correlaciones de baja magnitud es posible que no se retenga la hipótesis nula. En tal sentido, es necesario contar con un indicador de magnitud de asociación entre ítems: el KMO. Es decir, no basta con saber si los ítems *están asociados o no*; es necesario conocer *qué tan fuerte o débil es la asociación*. La expresión matemática del KMO es la siguiente:

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2} \quad \ldots (2)$$

Donde la expresión que contiene la *r* indica la *suma de correlaciones interítems al cuadrado*, mientras el segundo componente que contiene la *a* representa la *suma de correlaciones parciales interítems al cuadrado*. A modo de recordatorio, y en este contexto, la *correlación parcial* expresa el grado de relación lineal entre dos ítems controlando la influencia de todos los demás. Por ejemplo, para el ítem 1 y 2 (Figura 3), la correlación parcial (a^2_{ij}) sería la asociación entre esos ítems, pero sin considerar sus asociaciones con los ítems 3, 4, 5 y demás (es decir, la potencial varianza compartida o *comunalidad*); y, si se eleva al cuadrado, a^2_{ij} representa la varianza que no comparten los ítems 1 y 2 (*unicidad*).

Visto de este modo, el KMO es la razón entre la *potencial varianza compartida entre ítems* (r^2_{ij}) y la varianza total (*comunalidad* + *unicidad*). En otras palabras, el KMO se podría interpretar como la proporción de la varianza total explicada por el factor común (o factores). Entonces, si el segundo componente del denominador es más elevado que el primero (potencial comunalidad), el KMO será bajo (o al menos, no tan prometedor); mientras que si el segundo componente del denominador es menor que el primero (potencial comunalidad), el KMO será alto (se asume que .80 es el valor mínimo aceptable; Kaiser, 1970).

Ejecución del análisis factorial exploratorio

Hasta este punto, los datos son susceptibles de ser analizados con un AFE. Entonces, el siguiente paso para determinar la estructura factorial es responder a la siguiente pregunta: ¿Cuántos factores deben ser extraídos? ¿Con qué método serán extraídos? ¿Qué tipo de relación se espera entre factores?

En psicometría existe una combinación de métodos para responder las tres preguntas antes mencionadas que en un tiempo fue ampliamente utilizado debido a la aparente facilidad para computarlos con los softwares disponibles. Esta combinación es conocida como *Little Jiffy* y solucionaba las tres etapas clave en cualquier proceso de AFE: para determinar el número de factores se utilizaba la regla de Kaiser (1960); como método de extracción factorial, se empleaba el Análisis de Componentes Principales (ACP); y, para terminar, la rotación varimax. En la actualidad, la aplicación conjunta de estos métodos está ampliamente desestimada (Ferrando & Anguiano-Carrasco, 2010; Hauck-Filho & Valentini, 2020; Lloret-Segura et al., 2014), los motivos se detallan a continuación.

Número de factores. Es un proceso que permite conocer cuántos factores como mínimo (ya que se busca una estructura parsimoniosa) son necesarios para explicar la variabilidad de los ítems. En este punto es conveniente recordar el impacto de la teoría que sustenta el test, ya que se pretende evaluar un constructo multidimensional que presenta cinco dimensiones, se podría esperar que el número de factores sugeridos sea cinco. Sin embargo, no siempre es así, ya que, dependiendo del método, este puede sugerir extraer más o menos factores.

En tal sentido, en la combinación *Little Jiffy* se usaba la *regla de Kaiser*, la cual sugería extraer, como máximo, tantos factores como valores Eigen (también denominado *valor propio* o *Eigen-value* de un factor) mayores que uno existieran. El valor Eigen del primer factor es el resultado de la sumatoria del cuadrado de las cargas factoriales de dicho factor en la matriz no rotada, el valor Eigen del segundo factor se obtiene de la sumatoria del cuadrado de sus cargas factoriales en la matriz no rotada (y así sucesivamente, si hubiera más factores).

Por ejemplo, si había ocho valores Eigen mayores que uno, se podía extraer ocho, siete, seis, etc. factores. Sin embargo, intrínsecamente, presenta algunas falencias. Por ejemplo, no existe un criterio claro en la toma de decisiones. Además, existe evidencia de que sobreestima la cantidad de factores que debe retenerse (Ferrando & Anguiano-Carrasco, 2010; Henson & Roberts, 2006; Zwick & Velicer, 1986), es decir, que sugiere más factores que los que realmente se necesitan. Esto podría afectar la interpretación de la solución factorial y, en consecuencia, la validez de las interpretaciones. Además, existen estudios de simulación que la han descartado como un método analítico objetivo, así como al *Scree Test* (Cattell, 1966), dado que la subjetividad que conlleva el uso de este último le resta credibilidad (Timmerman & Lorenzo-Seva, 2011).

En tal sentido, el análisis paralelo (AP; Horn, 1965) parece ser una opción más potente que la *regla de Kaiser*, ya que proporciona un respaldo empírico para decidir cuántos factores extraer, además de brindar resultados confiables cuando hay desviaciones respecto a la normalidad y cuando los ítems son ordinales. No obstante, una versión más reciente de este método (*optimal implementation*) basado en el análisis factorial de rango mínimo (Minimum Rank Factor Analysis) es más precisa en la estimación del número de factores (Timmerman & Lorenzo-Seva, 2011).

De forma esquemática, el método comienza calculando el *porcentaje real de varianza* (PRV) explicada por cada factor con los datos empíricos. Después de ello, a partir de una matriz de correlaciones con datos aleatorios (generados por el mismo programa estadístico, considerando el mismo tamaño muestral y número de ítems), se calculan los *porcentajes aleatorios de varianza* (PAV), también asociados a cada factor. Luego de ello, se compara el primer PRV (real) con el primer PAV (aleatorio) y si es mayor el real, se continúa comparando. Se deja de comparar cuando el PAV es mayor que el PRV, y si la última comparación fue en el factor k, se tomarán (k-1) factores. El fundamento es que lo aleatorio no podría explicar más que lo real.

A fin de explicar los resultados del análisis paralelo, y de aquí en lo sucesivo, se consideraron los datos del análisis de un test que evalúa *afectividad* con 10 ítems (originalmente los ítems 2, 6, 7, 8, 9 evalúan *afecto negativo*, y el resto el *afecto positivo*). La base de datos usada se brinda como material complementario.

En este caso, la PRV del primer factor fue de 53.028 (Figura 4a), mientras que la PAV también del primer factor fue 21.007% (Figura 4b) y al ser menor que PRV, se continúa con la siguiente fila. La PRV del segundo factor fue 20.424%, mientras que la PAV fue 17.666%; en tal sentido, se continúa. El tercer PRV fue 7.506% y es largamente menor que 15.126%, por lo que ahí se detiene la comparación. En ese sentido, al ser dos PRV que superan a sus homólogos aleatorios, se extraerán dos factores.

Figura 4
Análisis paralelo

```
PARALLEL ANALYSIS (PA) BASED ON MINIMUM RANK FACTOR ANALYSIS
(Timmerman & Lorenzo-Seva, 2011)

Implementation details:

        Correlation matrices analized:              Polychoric correlation matric
        Number of random correlation matrices:      500
        Method to obtain random correlation matrices: Permutation of the raw data (
                    a                b

Variable   Real-data       Mean of random    95 percentile of random
           % of variance   % of variance     % of variance

   1        53.0278*       21.0072           24.5569
   2        20.4236*       17.6661           20.0521
   3         7.5061        15.1256           16.8904
   4         5.9292        12.8726           14.5226
   5         4.3319        10.7966           12.2460
   6         4.2793         8.7663           10.3366
   7         2.4032         6.7110            8.4065
   8         1.3925         4.6006            6.4819
   9         0.7064         2.4539            4.5079

*  Advised number of dimensions:    2
```

Con todo, el uso del AP también podría traer dudas al usuario. Por ejemplo, si teóricamente se espera extraer seis factores pero el AP solo sugirió uno, ¿qué ocurrió? En estos casos se debe partir del supuesto base del análisis factorial: se busca explicar la variabilidad de los ítems con la menor cantidad de factores posible con base en su asociación. De ese modo, si la varianza de los ítems se explicaría mejor por un solo factor es porque la correlación entre todos los ítems es tal que no permite agruparlos en dimensiones independientes. Otra causa podrían ser los factores de dificultad o de endosamiento (Burga, 2006; Schweizer et al., 2021), según el cual los ítems con un similar patrón de respuesta (ítems *muy fáciles* o *muy difíciles*, para test de respuesta máxima; opciones *bajas* o *altas*, para test de respuesta máxima) se agruparán entre sí independientemente de su contenido. Además, si fue utilizado el análisis paralelo, este brinda una cantidad mucho menor que lo obtenido con la regla de Kaiser.

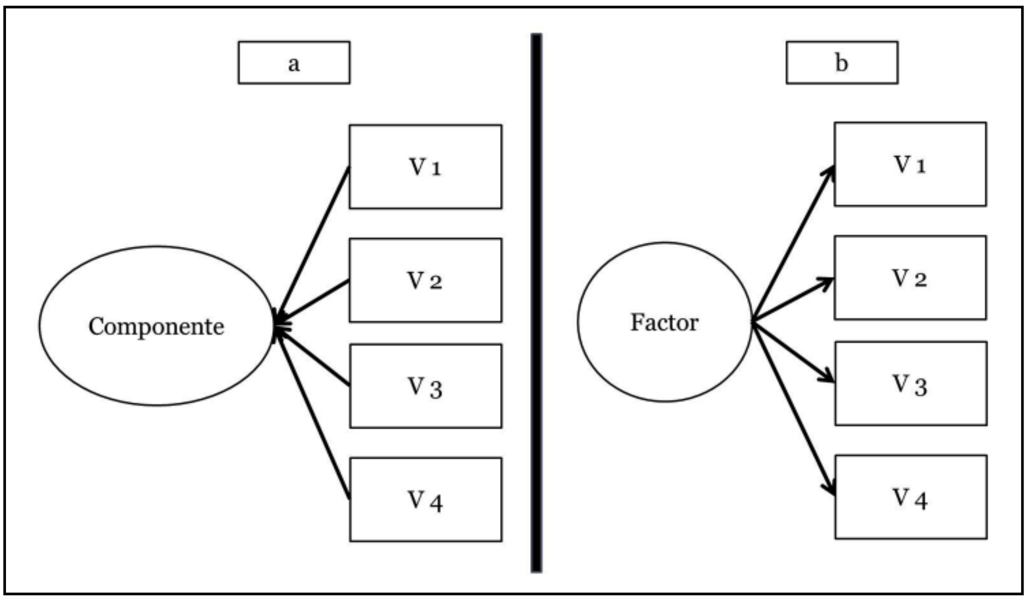
Análisis factorial exploratorio: métodos de extracción factorial

Se refiere a la selección del método más conveniente para reproducir de la forma más precisa la varianza común. Para realizar tal tarea, el ACP es el menos recomendado por la literatura para estudiar la estructura interna de un instrumento debido a su lógica analítica.

El ACP es un método estadístico que consiste en reducir un conjunto de datos a dimensiones más específicas llamadas componentes principales (Greenacre et al., 2022) y su aplicación en psicometría tiene como objetivo configurar *componentes* o dimensiones (que no son *factores* como se entienden en psicometría) como medidas aditivas construidas a partir de las variables observables o ítems (Figura 5a). Por el contrario, otros procedimientos que corresponden al análisis factorial propiamente dicho sí buscan conocer qué factor común explica la varianza de los ítems y son capaces de *aislar* la varianza común (comunalidad) (Figura 5b) (Lloret-Segura et al., 2014; López-Aguado & Gutiérrez-Provecho, 2019).

Figura 5

Diferencia entre análisis de componentes principales (a) y análisis factorial (b)



Adicionalmente, porque analiza la varianza total asumiendo que los ítems son medidas libres de error (Ferrando & Anguiano-Carrasco, 2010; Gruijters, 2019), lo que va en contra de los postulados de la medición psicológica, además de elevar de forma espuria la magnitud de las cargas factoriales (Costello & Osborne, 2005; Pérez & Medrano, 2010; Watkins, 2018) debido a que no separa comunalidad de especificidad, y no permite una interpretación adecuada del constructo evaluado.

En ese sentido, entre los métodos más recomendados se encuentran el de Máxima verosimilitud (MV) y el conjunto de métodos denominados Mínimos Cuadrados Ordinarios (MCO) como Ejes Principales, Mínimos Cuadrados no Ponderados, etc., así como el Análisis Factorial de Rango Mínimo (Ferrando & Anguiano-Carrasco, 2010), o el Mínimos Cuadrados de Diagonal Ponderada (DWLS; Christoffersson, 1977). El método de MV fue creado para variables métricas, continuas y con distribución normal multivariada, lo que en ocasiones no es realista con los ítems, como ya se vio anteriormente. No obstante, los MCO pueden usarse con ítems en nivel de medición nominal u ordinal (Ferrando & Lorenzo-Seva, 2014). Para mayores especificaciones técnicas de los métodos se recomiendan algunas lecturas (Ferrando & Anguiano-Carrasco, 2010; Lloret-Segura et al., 2014; Watkins, 2018).

Rotación factorial. Cuando se realiza la extracción de factores, la primera solución factorial (o distribución de cargas factoriales) normalmente favorece al primer factor, dado que explica la mayor cantidad de varianza y es necesario distribuir esa varianza entre los demás factores para procurar una estructura más simple e interpretable, es decir, cuando cada ítem es influido predominantemente por un factor. En este punto, conviene aclarar un aspecto de suma importancia: ¿Qué son las cargas factoriales?

La carga factorial, representada habitualmente con λ (lambda), se entiende como el grado de influencia que tiene el factor sobre el ítem (Figura 5b), y su valor va de -1 a +1 (análogo al coeficiente β en un análisis de regresión). El cuadrado de la carga factorial representa la comunalidad (λ^2), y mediante ese indicador es posible conocer cuánta variabilidad del ítem es explicada por la variable latente.

Entonces, como todo indicador cuantitativo revisado anteriormente, tiene un valor mínimo (*punto de corte*) considerado como aceptable. Esta magnitud es muy variable, ya que puede considerar puntos fijos (en valor absoluto) como .30 (9% de varianza explicada) (McDonald, 1985), entre .30 (9% de varianza explicada) y .40 (16% de varianza explicada) (Brown, 2006), .50 (25% de varianza explicada) (Costello & Osborne, 2005; Dominguez-Lara, 2018; Hair et al., 2006), > .70 (49% de varianza explicada) (Hair et al., 2006); o rangos que van desde .32 (pobre), .45 (aceptable), .55 (bueno), .63 (muy bueno) o .71 (excelente) (Tabachnick & Fidell, 2007). Sin embargo, como toda asignación arbitraria de puntos de corte, es probable que esto afecte algunas propiedades importantes, como la confiabilidad de constructo, lo que se verá más adelante.

En este sentido, si bien es deseable que la carga factorial tenga una magnitud moderada o alta para concluir que el constructo está representado de forma aceptable por el ítem, es posible que la carga factorial sea mayor que la unidad, lo que se le conoce como *caso Heywood*. Las causas de esta situación son diversas como *outliers* (Bollen, 1987), fallas en la convergencia del modelo (Boomsma & Hoogland, 2001), malas especificaciones (e.g., Kolenikov & Bollen, 2012) o fluctuaciones muestrales (Anderson & Gerbing, 1984), y si bien existen procedimiento para lidiar con ellos en el marco del AFC (e.g., especificar a la unidad la carga del ítem), su tratamiento a nivel de AFE no es claro aún. Por ello, podrían ser eliminados debido a que formarían parte de una solución impropia.

Retomando la idea de la rotación, estos procedimientos pueden agruparse en ortogonales (e.g., varimax) y oblicuas (e.g., promin). Las primeras asumen completa independencia entre factores (correlación cero), mientras que las otras permiten que la correlación interfactorial pueda estimarse (Goretzko et al., 2021). En ese sentido, y dado que en psicología los fenómenos estudiados frecuentemente se hallan relacionados, al menos hoy en día no se justificaría el uso de una estrategia ortogonal (como la rotación varimax). Por ejemplo, algunos modelos teóricos como el de los cinco grandes factores (Goldberg, 1990) postulan inicialmente que se trataba de dimensiones independientes, pero la evidencia indica que existe asociación entre ellos (ver Dominguez-Lara, Merino-Soto, et al., 2018).

Para ejemplificar el proceso, se consideraron los resultados del análisis estructural donde se implementó una rotación oblicua Promin (Lorenzo-Seva, 1999) con el programa FACTOR (Ferrando & Lorenzo-Seva, 2017) con los 10 ítems del test usado de ejemplo para el AP. Como se puede apreciar en la matriz no rotada (Figura 6a), las cargas factoriales en el primer factor son mucho mayores que en el segundo, lo que es coherente con lo mencionado anteriormente. Luego de la rotación (Figura 6b), los ítems se encuentran más orientados hacia uno u otro factor, dependiendo de la afinidad entre ítems. Por ejemplo, el ítem 3 presentaba cargas en el factor 1 y 2 de .520 y -.649, respectivamente; antes de la rotación y después de la misma, la carga factorial más elevada corresponde al segundo factor (.885) y la más baja, al primero (.145).

Figura 6
Ejemplo de rotación factorial

a

UNROTATED LOADING MATRIX

Variable	F 1	F 2
V 1	0.668	-0.270
V 2	-0.711	-0.386
V 3	0.520	-0.649
V 4	0.541	-0.554
V 5	0.537	-0.282
V 6	-0.780	-0.276
V 7	-0.806	-0.328
V 8	-0.759	-0.399
V 9	-0.849	-0.284
V 10	0.675	-0.486

b

ROTATED LOADING MATRIX

Variable	F 1	F 2
V 1	-0.277	0.556
V 2	0.848	0.101
V 3	0.145	0.885
V 4	0.051	0.795
V 5	-0.169	0.514
V 6	0.809	-0.041
V 7	0.871	0.002
V 8	0.894	0.095
V 9	0.867	-0.062
V 10	-0.105	0.781

c

ROTATED LOADING MATRIX
(loadings lower than absolute 0.300 omitted)

Variable	F 1	F 2
V 1		0.556
V 2	0.848	
V 3		0.885
V 4		0.795
V 5		0.514
V 6	0.809	
V 7	0.871	
V 8	0.894	
V 9	0.867	
V 10		0.781

Correlación entre factores. Cuando se determina, empírica o teóricamente, que los factores se podrían asociar, se plantea una rotación oblicua. Como resultado, además de las cargas factoriales, también es calculada la magnitud de la correlación entre factores (ϕ). Este indicador muchas veces no es informado ni interpretado, pero da cuenta de algo fundamental para entender la configuración del test: la diferenciación empírica de los factores o hasta qué punto los factores pueden diferenciarse o no son redundantes entre sí.

En primer lugar, de la misma forma que el coeficiente de correlación de Pearson de apartados anteriores, la correlación interfactorial refleja el grado de asociación entre los constructos (variables latentes) y en este caso es de -.434. El signo tiene sentido, ya que son características que aparecen de forma inversamente proporcional: se espera que, a mayor experiencia de afecto positivo, disminuya la de afecto negativo; y la magnitud también, ya que expresa una asociación diferente de cero. En segundo lugar, al elevar al cuadrado esa correlación

(ϕ^2) da información sobre la varianza compartida entre factores. En este caso, da .188.

Hecho esto, se puede concluir que dos factores están bien diferenciados cuando la varianza explicada por cada factor de forma independiente (AVE; varianza promedio extraída por factor) es mayor que la varianza compartida entre factores (ϕ^2) (Fornell & Larcker, 1981; Shrestha, 2021). El cálculo de la AVE es simple: es el promedio del cuadrado de las cargas factoriales (comunalidad). La AVE para el factor afecto negativo es de .737, mientras que para el afecto positivo fue de .520. Ambas AVE fueron mayores que .188 (ϕ^2), por lo que los factores están diferenciados.

De forma general, cuando la correlación entre factores es mayor que .70 (Joshanloo et al., 2016), podría ser un indicio de solapamiento entre factores. Por ejemplo, la CD-RISC (Connor & Davidson, 2003) fue creada originalmente como una medida de la resiliencia con cinco dimensiones. Sin embargo, el método para determinar el número de factores fue la regla de Kaiser (lo que ya es un indicio de sobreestimación), y en un estudio reciente se pudo demostrar (aunque con un AFC) que los factores originales se correlacionaban de forma muy elevada, y que un solo factor es capaz de explicar mejor la variabilidad de los ítems de la CD-RISC (para mayor información, revisar Dominguez-Lara et al., 2019). Con todo, en los otros estudios instrumentales de la CD-RISC este aspecto no se exploró ni se cuestionó.

Porcentaje de varianza explicada. La varianza explicada fue considerada un indicador relevante de dimensionalidad que durante mucho tiempo se tomó como referencia en los reportes de AFE. Por ejemplo, la literatura muestra una serie de recomendaciones que indicaban porcentajes mínimos de varianza explicada del primer factor para considerar a la medida como esencialmente unidimensional: entre 17% y 40% usando matrices de correlaciones *phi* (Zwick, 1985), 20% (Reckase, 1979), entre 30% y 40% usando matrices de correlaciones tetracóricas (Zwick, 1985) o 40% de la varianza (Carmines & Zeller, 1979); y esa varianza explicada por el primer factor se obtenía con base en el valor Eigen, concepto ya visto anteriormente, y que es recomendable abordar dado que el lector ya está más familiarizado con conceptos como matriz de cargas factoriales, rotación, comunalidad, etc.

Como se mencionó antes, el valor Eigen del primer factor es el resultado de la sumatoria del cuadrado de las cargas factoriales de dicho factor en la matriz no rotada, el valor Eigen del segundo factor se obtiene de la sumatoria del cuadrado de sus cargas factoriales en la matriz no rotada (y así sucesivamente, si hubiera más factores). Entonces, considerando los datos de la Figura 6a, el primer valor Eigen sería de 4.814 y el segundo de 1.690, y cuando cada valor Eigen se divide entre el número de ítems del test, se obtiene que el porcentaje de varianza explicada por los dos factores es de 48.14% y 16.90%, respectivamente.

Sin embargo, el reporte de la varianza explicada en la actualidad no es viable cuando se usan procedimientos vinculados al análisis factorial propiamente dicho (e.g., MCO) porque la probabilidad de que aparezcan valores Eigen negativos es muy elevada (Lloret-Segura et al., 2014; Lorenzo-Seva, 2013). De esto se colige por qué no fue considerada como un método para determinar el número de factores, pese a la recomendación muy extendida (e.g, Campo-Arias et al., 2009) de que deben retenerse factores cuyos valores Eigen estén asociados al menos a la explicación del 50% de la varianza (Streiner, 1994).

Antes de finalizar este apartado, surge la pregunta: ¿Por qué aún se usa el *Little Jiffy* en psicometría si existe evidencia desfavorable desde hace casi 40 años? Si bien es cierto existen estudios que resaltan algunas debilidades de los métodos, así como la aparente contradicción entre objetivos (hallar la *variable latente*) y medios (el ACP trabaja con *variables aditivas*), es posible ensayar algunas respuestas.

- *Navaja de Occam*. Uno de los motivos de su difusión probablemente sea porque muchos programas estadísticos comerciales (e.g., SPSS) incluyen, en el mismo *bloque analítico*, las técnicas de AFE y el ACP (Lloret-Segura et al., 2014); lo que llevaría al usuario a pensar que persiguen los objetivos similares. Además, al proveer parámetros estadísticos aparentemente más favorables en el ACP (e.g., «cargas factoriales» de mayor magnitud), opta por él. Es decir, se aplica la *navaja de Occam* a la psicometría: si tienes dos métodos, usa el más *sencillo*.

- *Argumentum ad populum*. Lloret-Segura et al. (2014) reportaron la frecuencia del uso de determinadas técnicas asociadas al análisis factorial en estudios instrumentales publicados en España. Alrededor del 50% utiliza el ACP, mientras que un hallazgo similar se obtuvo en una revista especializada en evaluación psicológica (Ledesma et al., 2019). De ello, que se refuerce el argumento *si la mayoría lo hace, entonces está bien*. Es más, algunos estudios usan esta premisa para justificar su uso: «*Existen varios métodos para extraer los factores iniciales desde la matriz de correlación, siendo sin duda el método de Análisis de Componentes Principales el más utilizado*.» (Garmedia, 2007, p. 60).

- *Manuales disponibles*. Además, algunos textos clásicos y de uso extendido señalan que el ACP presenta, *en la práctica*, resultados *similares* al AFE (Nunnally & Bernstein, 1994). Asimismo, al ser recomendado en manuales de uso extendido, el lector novel que se encuentre con el siguiente argumento: «*En la confección de instrumentos el análisis más recomendado suele ser el de Componentes Principales (Armor, 1974; Nunnally, 1978; Carmines y Zeller, 1979; Spector, 1992), aunque otros autores suelen preferir el análisis de Factores Comunes pues da unas estimaciones de los pesos más conservadoras al tener en cuenta el error o varianza no compartida de las variables (Gorsuch, 1986a)*» (Morales, 2011, p. 11), podrá ver reforzado los argumentos antes mencionados. Pese a que ya se ha revisado con anterioridad que la lógica analítica entre las técnicas de AFE y ACP es distinta (Figura 5). Afortunadamente, existen cada vez más espacios donde se brindan recomendaciones más precisas sobre su uso (Ledesma et al., 2019; López-Aguado & Gutiérrez-Provecho, 2019).

- *Negación del usuario*. Hace un tiempo se envió una carta al editor a una revista (Dominguez-Lara, Fernández-Arata et al., 2018) exponiendo esta problemática de índole metodológico en un instrumento ampliamente utilizado en Latinoamérica para realizar un diagnóstico organizacional: Inventario de Violencia y Acoso Psicológico en el Trabajo (IVAPT). Sin embargo, los argumentos de réplica (Pando & Calderón, 2018) se orientaron a reforzar la idea de que el AFE y ACP son similares («diferentes métodos, entre ellos el análisis factorial y de componentes, que son considerados métodos muy similares»), y el *argumentum ad populum* («el método conocido como Little Jiffy es útil no solo por estar extendido en su uso», «los datos ofrecidos en la validación se sustentan en un instrumento con evidencias de validez obtenidas a través del método más comúnmente usado por las ciencias del comportamiento»), e incluso tergiversando argumentos de la carta original cuando mencionan «Little Jiffy fue desarrollado no a partir de los resultados de un software, sino como producto del análisis del problema matemático de los factores», sin haber referido eso el envío inicial; además de otros argumentos que invitamos al lector a revisar.

Análisis factorial exploratorio: denominación de los factores

Luego del procesamiento inicial y de la obtención de la estructura factorial preliminar es necesario determinar, con base en los indicadores cuantitativos de las cargas factoriales, qué ítems son explicados predominantemente por determinado factor, acercándose a una estructura simple, con cargas factoriales de valencia coherete (positivas o negativas), y facilidad de interpretación (Anastasi, 1990). Por ello, probablemente es la etapa más importante porque es la que brindará al potencial usuario un instrumento con sentido empírico y teórico. Teóricamente, una estructura simple (Thurstone, 1947) se refiere al escenario en el que, en un test multidimensional, cada ítem es influido exclusivamente por un factor, es decir, que la carga factorial en dicho factor es diferente de cero, y las cargas secundarias (es decir, la influencia de los *otros factores*) son cero (Figura 7a) (Ferrando & Anguiano-Carrasco, 2010).

Normalmente, inicia con la selección de la más alta carga factorial asociada a un factor que supere el criterio mínimo establecido con anterioridad (e.g., .32). Posteriormente, se agrupan los ítems que tengan la más alta carga en el primer factor, luego, en el segundo, etc. Después, con base en *algo común*, se otorga el nombre a un factor. Sin embargo, en este punto es poco probable que se deje algo al azar. Para llegar a este procedimiento, el test atravesó un proceso de análisis asociado al contenido de los ítems, por lo que se espera que los ítems se agrupen en torno a su factor teórico, aunque en tests con bases teóricas deficientes, ítems mal redactados o descuidos en el proceso de recolección de datos, las soluciones factoriales suelen estar pobremente definidas (ver Muñiz & Fonseca-Pedrero, 2019).

Como se apreció en el apartado anterior (Figura 6b), en ocasiones, identificar *qué ítems son explicados predominantemente por determinado factor* no es una tarea complicada. Atendiendo la tercera parte de la imagen (Figura 6c), se remarca que el punto de corte usado por el programa estadístico para determinar una carga factorial como *mínimamente aceptable* es .30 (*Loading lower than absolute .300 omitted* [*Las cargas con valor absoluto menor que .300 fueron omitidas*]). En otras palabras, según el software usado se considera que el ítem es un buen representante del constructo si este último explica por lo menos el 9% de su variabilidad, es decir, su comunalidad, considerando que, como se mencionó en apartados anteriores, se calcula elevando al cuadrado la carga factorial (comunalidad: $.30^2 = .09 = 9\%$). Así, los cinco ítems (2, 6, 7, 8 y 9) son influidos en mayor medida por el primer factor, y otros cinco ítems (1, 3, 4, 5 y 10) por el segundo factor. Como se puede intuir, el primer factor será denominado *afecto negativo*, mientras que el segundo, *afecto positivo*.

Antes de terminar el análisis factorial exploratorio…

Existen cuatro puntos adicionales a mencionar: 1) La posibilidad de que las cargas factoriales sean menores que el punto de corte establecido; 2) La estructura resultante podría no ser una *estructura simple*, en el sentido estricto del término; 3) La magnitud de las cargas factoriales puede afectar la confiabilidad del constructo; 4) El ajuste estadístico del modelo debe ser considerado.

- *Cargas factoriales con magnitudes por debajo de lo esperado.* Esta situación es frecuente, ya que, independientemente de la valoración de los jueces expertos, el análisis conjunto puede determinar que algunos ítems no son influidos satisfactoriamente por el factor o, en otras palabras, no son buenos representantes del constructo. En ese caso, lo que se propone habitualmente es la *eliminación por pasos*, es decir, comenzar con aquel que tiene la más baja carga factorial (no de forma simultánea) y volver a realizar el análisis.
- *Simplicidad factorial*. Se indicó que la estructura simple implica que cada ítem solo es influido por

un factor (carga factorial diferente de cero), y con valores cero en las cargas secundarias (o la influencia de los *otros factores*) (Figura 7a). Sin embargo, conforme a lo expuesto hasta ahora, es un escenario poco probable, ya que en un modelo multidimensional los factores deben estar asociados para poder considerarlos como parte integrante de un constructo de mayor envergadura, por lo que es más realista pensar que las cargas secundarias serán distintas de cero (Figura 7b) (Kaiser, 1974). Por ello, un ítem puede ser influido por dos o más factores porque, incluso cuando el número de factores es consistente (o aproximado) con la teoría, un ítem podría representar de forma significativa a más de un factor.

Figura 7
Simplicidad factorial

a

	F1	F2	F3
V1	X	0	0
V2	X	0	0
V3	X	0	0
V4	0	X	0
V5	0	X	0
V6	0	X	0
V7	0	0	X
V8	0	0	X
V9	0	0	X

b

	F1	F2	F3
V1	X	n	n
V2	X	n	n
V3	X	n	n
V4	n	X	n
V5	n	X	n
V6	n	X	n
V7	n	n	X
V8	n	n	X
V9	n	n	X

Retomando la información mostrada en la Figura 6b, sería conveniente centrar la atención en dos ítems: el uno y el cuatro. El ítem uno presenta una carga factorial de -.277 (muy cercana al |.30|) en el primer factor; mientras que, en el segundo, el valor asciende a .556. La conclusión lógica es que el ítem es representado *en mayor medida* por el segundo factor. En cuanto al ítem cuatro, la carga factorial en el primer factor es de .051 (que se puede considerar insignificante), mientras que en el segundo factor fue de .795, y también se puede llegar a una conclusión similar, es decir, que el ítem 4 representa más al segundo factor. Sin embargo, existe una diferencia: la distancia entre las cargas factoriales (en valor absoluto) no es muy grande en el ítem uno, lo que podría aproximarse a lo que se conoce como *complejidad factorial* (Fleming & Merino, 2005) o el hecho de que el ítem esté influido de forma significativa por más de un factor a la vez.

La evaluación se realiza mediante el *Índice de Simplicidad Factorial* (ISF; Kaiser, 1974), que indica hasta qué punto la varianza de un ítem es explicada en mayor medida por un factor. El ISF va del cero a la unidad y son aceptables valores por encima de .70 (Fleming & Merino, 2005), es decir, mientras más cercano a uno, el ítem es factorialmente más simple, y si alcanza la unidad significa que el ítem es influido exclusivamente por un factor. El cálculo puede realizarse con el programa SIMLOAD (Fleming, 2003) o con un módulo en MS Excel creado para tal fin (Dominguez-Lara, 2016). Para terminar con el ejemplo, el ISF del primer ítem fue de .602, mientras que el del cuarto ítem fue de .992, lo que coincide con los comentarios del párrafo anterior.

Sin embargo, se puede dar el caso en que la simplicidad del ítem no sea suficiente (ISF < .70). Esto lleva a varias preguntas, como ¿qué hacer ante la existencia de ítems complejos? En vista que el objetivo final es lograr una estructura simple, una opción sugerida sería eliminarlos y dar preferencia a ítems que representen en mayor grado a un solo factor (Fleming & Merino, 2005). Además, debe tenerse en cuenta que no es deseable que un ítem pertenecezca a dos o más dimensiones en un test (Dominguez-Lara & Merino-Soto, 2018), porque cuando se realice la sumatoria de las puntuaciones de dos dimensiones (y estas compartan ítems), se puede elevar la correlación entre estas de forma artificial.

Si no se tiene en cuenta esta situación, podrían interpretarse incorrectamente algunos resultados. Por ejemplo, un estudio donde se analizó la estructura de un cuestionario para evaluar la importancia de la familia en los cuidados de enfermería concluyó a favor de una estructura de cuatro factores (Pascual-Fernández et al., 2015). No obstante, en un reanálisis se pudo encontrar un número importantes de ítems complejos: de los 26 ítems, 15 tenían un ISF menor que .80 (57.69%) y de estos, 11 presentaron un ISF menor que .70 (ver Dominguez-Lara, 2016).

- *Relación entre las cargas factoriales y la confiabilidad del constructo*. En términos generales, un coeficiente de confiabilidad representa la proporción de varianza verdadera de una medición (Lord & Novick, 1968), y aunque habitualmente se le asocie exclusivamente con el uso del coeficiente α orientado a las puntuaciones observadas (Cronbach, 1951), el uso del término confiabilidad también es factible a nivel de variables latentes.

  Existen algunos requisitos para estimar apropiadamente el a, y uno de los más relevantes es que los ítems sean tan equivalentes, es decir, que las cargas factoriales en una dimensión sean estadísticamente similares (Dunn et al., 2013), y es probable una infraestimación del coeficiente cuando no se cumple este supuesto. En contraste, coeficientes como el w (McDonald, 1999) funcionan apropiadamente para medidas congenéricas (es decir, que los ítems evalúen una dimensión en común), incluso, asimétricas (Trizano-Hermosilla & Alvarado, 2016), por lo que no dependen de que exista la tau equivalencia (ver Viladrich et al., 2017). Su expresión matemática es:

$$\omega = \frac{\left(\sum \lambda_j\right)^2}{\left(\sum \lambda_j\right)^2 + \sum \sigma^2_{\varepsilon j}} \quad \ldots (3)$$

Donde λ_j hace referencia a las cargas factoriales y σ^2_ε a la varianza residual. Su valor está entre 0 y 1, y podría interpretarse como el grado en que el constructo es reflejado por sus indicadores observables (Hancock & Mueller, 2001) o qué proporción de la varianza de los ítems es explicada por el constructo. El reporte de esta información es importante, y es un indicador de mayor confianza respecto a la representatividad de los ítems respecto al constructo.

En ese sentido, aplicando la ecuación 3 a las cargas del primer factor de la Figura 6c (afecto negativo), el coeficiente ω calculado fue de .933, lo que podría interpretarse como *el 93.3% de la varianza de los ítems puede atribuirse al constructo evaluado*. En este punto, la magnitud resulta aceptable. Sin embargo, ¿qué pasaría si con la misma cantidad de ítems ahora las cargas factoriales son de .30 o .40? En el primer caso (las cinco cargas igual a .30), el nuevo ω es de .331; y con cargas igual a .40, de .488. Si bien queda a libertad del investigador elegir el valor mínimo para considerar como aceptable una carga factorial, es necesario que pueda apreciar hasta qué punto las cargas factoriales afectan la confiabilidad del constructo.

En este punto, queda clara la sugerencia de analizar primero la estructura interna y luego la confiabilidad, ya que si se analiza la confiabilidad antes de determinar cuál es la estructura factorial, se está asumiendo que las dimensiones originales existen y, entonces, ¿qué pasaría si los hallazgos del AFE concluyen un número diferente de dimensiones? Esta es una práctica bastante extendida, pero que es necesario revisar en el futuro. Por tal motivo, en estos casos es recomendable analizar la confiabilidad (sea de puntuaciones o de constructo) luego de determinar la estructura factorial con métodos apropiados para no asumir sin evidencia suficiente la estructura establecida *a priori* por los autores del instrumento.

Conclusiones y recomendaciones

Como el lector pudo apreciar, la complejidad en el procesamiento del AFE atañe a cada etapa, y si bien algunos podrían preguntarse *¿es realmente necesario hacer todo eso?*, la respuesta evidentemente es afirmativa por los motivos expuestos al inicio del manuscrito: es un procedimiento muy útil para el investigador porque se puede aproximar a la evaluación empírica de un constructo. De este modo, se concluye que la toma de decisiones dentro de cada análisis puede conducir a resultados más o menos confiables desde el punto de vista psicométrico, por lo que es fundamental realizar un plan de análisis que contemple todas las características del instrumento que se estudiará. No obstante, pese a los constantes avances a nivel metodológico, aún quedan aspectos pendientes por abordar como la optimización de estrategias para remover ítems (Güvendir & Özkan, 2022), el manejo de los casos Heywood mencionados en párrafos anteriores (Cooperman & Waller, 2022) e incluso el manejo de los residuales correlacionados (Ferrando et al., 2022).

Conflicto de intereses

El autor declara no tener conflicto de intereses.

Responsabilidad ética

El presente estudio es teórico. Los datos de pacientes no han sido utilizados.

Contribución de autoría

SDL: redacción y revisión del manuscrito.

Referencias

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Anastasi, A. (1990). *Psychological Testing*. McMillan.

Anderson, J. C., & Gerbing, D. (1984). The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness-of-Fit Indices for Maximum Likelihood Confirmatory Factor Analysis. *Psychometrika*, *49*, 155-173. https://doi.org/10.1007/BF02294170

Ayán, M., & Ruiz, M. (2008). Atenuación de la asimetría y de la curtosis de las puntuaciones observadas mediante transformaciones de variables: Incidencia sobre la estructura factorial. *Psicológica*, *29*, 205-227. https://psycnet.apa.org/record/2008-15048-007

Bollen, K. A. (1987). Outliers and Improper Solutions: A Confirmatory Factor Analysis Example. *Sociological Methods and Research*, *15*(4), 375-384. https://doi.org/10.1177/0049124187015004002

Boomsma, A., & Hoogland, J. J. (2001). The Robustness of LISREL Modeling Revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural Equation Models: Present and Future. A Festschrift in Honor of Karl Jöreskog* (pp. 139-168). Scientific Software International.

Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research*. The Guilford Press.

Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research* (2.ª ed.). The Guilford Press

Bunge, M. (2004). *La investigación científica: su estrategia y su filosofía*. Siglo XXI Editores.

Burga, A. (2006). La unidimensionalidad de un instrumento de medición: perspectiva factorial. *Revista de Psicología*, *24*(1), 53-80. https://doi.org/10.18800/psico.200601.003

Campo-Arias, A., Bustos-Leiton, G., & Romero-Chaparro, A. (2009). Consistencia interna y dimensionalidad de la Escala de Estrés Percibido (EEP-10 y EEP-14) en una muestra de universitarias de Bogotá, Colombia. *Aquichán*, *9*(3), 271-280. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-59972009000300007

Carmines, E., & Zeller, R. (1979). *Reliability and Validity Assessment*. Sage.

Cattell, R. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research*, *1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Chávez-Ventura, G., Polo-López, T., Zegarra-Pereda, L., Balarezo-Aliaga, O., Calderón-Valderrama, C., & Dominguez-Lara, S. (2025). Self-Efficacy Scale for University Teaching in Virtual Environments, SSUTVE. *Heliyon*, *11*(1), e41134. https://doi.org/10.1016/j.heliyon.2024.e41134

Christoffersson, A. (1977). Two-Step Weighted Least Squares Factor Analysis of Dichotomized Variables. *Psychometrika*, *42*(3), 433-438. https://doi.org/10.1007/BF02293660

Connor, K. M., & Davidson, J. R. (2003). Development of a New Resilience Scale: The Connor Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*, *18*(2), 76-82. https://doi.org/10.1002/da.10113

Cooperman, A. W., & Waller, N. G. (2022). Heywood You Go Away! Examining Causes, Effects, and Treatments for Heywood Cases in Exploratory Factor Analysis. *Psychological Methods*, *27*(2), 156-176. https://doi.org/10.1037/met0000384

Costello, A., & Osborne, J. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from your Analysis. *Practical Assessment Research & Evaluation*, *10*(1), 1-9. https://doi.org/10.7275/jyj1-4868

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of a Test. *Psychometrika*, *16*, 297-334. https://doi.org/10.1007/BF02310555

Dominguez-Lara, S. (2014). Autoeficacia para situaciones académicas en estudiantes universitarios peruanos: un enfoque de ecuaciones estructurales. *Revista de Psicología - UCSP*, *4*, 43-54. https://dialnet.unirioja.es/servlet/articulo?codigo=8541596

Dominguez-Lara, S. (2016). Análisis factorial exploratorio y complejidad factorial: más allá de las rotaciones. *Enfermería Clínica*, *26*(6), 401. https://doi.org/10.1016/j.enfcli.2016.06.001

Dominguez-Lara, S. (2018). Propuesta de puntos de corte para cargas factoriales: una perspectiva de fiabilidad de constructo. *Enfermería Clínica*, *28*(6), 401-402. https://doi.org/10.1016/j.enfcli.2018.06.002

Dominguez-Lara, S., & Merino-Soto, C. (2018). Dos versiones breves del Big Five Inventory en universitarios peruanos: BFI-15p y BFI-10p. *Liberabit*, *24*(1), 81-96. http://ojs3.revistaliberabit.com/index.php/Liberabit/article/view/138/98

Dominguez-Lara, S., Villegas, G., Yauri, C., Mattos, E. & Ramírez, F. (2012). Propiedades psicométricas de una escala de autoeficacia para situaciones académicas en estudiantes universitarios peruanos. *Revista de Psicología-UCSP*, *2*(1), 27-39.

Dominguez-Lara, S., Fernández-Arata, M., Merino-Soto, C., Navarro-Loli, J. S., & Calderón, G. (2018). Inventario de Violencia y Acoso Psicológico en el Trabajo (IVAPT) en Colombia: el peligroso Little Jiffy. *Revista Salud Uninorte*, *34*(2), 536-537. http://dx.doi.org/10.14482/sun.34.2.658.47

Dominguez-Lara, S., Merino-Soto, C., Zamudio, B., & Guevara-Cordero, C. (2018). Big Five Inventory en universitarios peruanos: resultados preliminares de su validación. *Psykhe*, *27*(2), 1-12. https://doi.org/10.7764/psykhe.27.2.1052

Dominguez-Lara, S., Gravini-Donado, M., & Torres-Villalobos, G. (2019). Análisis psicométrico de dos versiones de la Connor-Davidson Resilience Scale en estudiantes universitarios peruanos: propuesta del CD-RISC-7. *Revista Argentina de Ciencias del Comportamiento*, *11*(2), 36-51. https://doi.org/10.32348/1852.4206.v11.n2.23774

Dominguez-Lara, S., Sánchez-Villena, A. R., & Fernández-Arata, M. (2020). Psychometric Properties of the UWES-9S in Peruvian College Students. *Acta Colombiana de Psicología*, *23*(2), 7-23. https://doi.org/10.14718/ACP.2020.23.2.2

Dunn, T. J., Baguley, T., & Brunsden, V. (2013). From Alpha to Omega: A Practical Solution to the Pervasive Problema of Internal Consistency Estimation. *British Journal of Psychology*, *105*(3), 399-412. https://doi.org/10.1111/bjop.12046

Ferrando, P. J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, *31*(1), 18-33. https://www.papelesdelpsicologo.es/pdf/1793.pdf

Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales. *Anales de Psicología*, *30*(3), 1170-1175. https://dx.doi.org/10.6018/analesps.30.3.199991

Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, Development and Future Directions. *Psicothema*, *29*(2), 236-240. https://doi.org/10.7334/psicothema2016.304

Ferrando, P. J., & Lorenzo-Seva, U. (2019). Robust Promin: A Method for Diagonally Weighted Rotation. *Liberabit*, *25*(1), 99-106. https://doi.org/10.24265/liberabit.2019.v25n1.08

Ferrando, P. J., Hernandez-Dorado, A., & Lorenzo-Seva, U. (2022). Detecting Correlated Residuals in Exploratory Factor Analysis: New Proposals and a Comparison of Procedures. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(4), 630-638. https://doi.org/10.1080/10705511.2021.2004543

Fleming, J. (2003). Computing Measures of Simplicity of Fit for loadings in Factor-Analytically Derived Scales. *Behavioral Research Methods, instruments, & Computers*, *35*, 520-524. https://doi.org/10.3758/bf03195531

Finney, S. J., & DiStefano, C. (2006). Nonnormal and Categorical Data in Structural Equation Modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *A Second Course in Structural Equation Modeling* (pp. 269-314). Information Age Publishing.

Fleming, J., & Merino, C. (2005). Medidas de simplicidad y de ajuste factorial: un enfoque para la evaluación de escalas construidas factorialmente. *Revista de Psicología*, *23*(2), 249-266.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Garmedia, M. L. (2007). Análisis factorial: una aplicación en el cuestionario de salud general de Goldberg, versión de 12 preguntas. *Revista Chilena de Salud Pública*, *11*(2), 57-65. https://revistasaludpublica.uchile.cl/index.php/RCSP/article/view/3095

Goldberg, L. R. (1990). An Alternative «Description of Personality»: The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, *59*(6), 1216-1229. https://doi.org/10.1037//0022-3514.59.6.1216

Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory Factor Analysis: Current Use, Methodological Developments and Recommendations for Good Practice. *Current Psychology*, *40*, 3510-3521. https://doi.org/10.1007/s12144-019-00300-2

Gravetter, F. J., & Wallnau, L. B. (2013). Introduction to statistics. In J. Hague, T. Matray, T. Williams, & L. Sarkisian (Eds.). *Statistics for the Behavioral Sciences* (pp. 3-36). Cengage Learning.

Greenacre, M., Groenen, P. J., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal Component Analysis. *Nature Reviews Methods Primers*, *2*(1), 100. https://doi.org/10.1038/s43586-022-00184-w

Gruijters, S. L. (2019). Using Principal Component Analysis to Validate Psychological Scales: Bad Statistical Habits We Should Have Broken Yesterday II. *European Health Psychologist*, *20*(5), 544-549.

Güvendir, M. A., & Özkan, Y. Ö. (2022). Item Removal Strategies Conducted in Exploratory Factor Analysis: A Comparative Study. *International Journal of Assessment Tools in Education*, *9*(1), 165-180. https://doi.org/10.21449/ijate.827950

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R., & Tathum, R. (2006). *Multivariate Data Analysis* (6.ª ed.). Prentice Hall.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking Construct Reliability within Latent Variable Systems. En R. Cudeck, S. H. C. Du Toit, & D. Sörbom (eds.), *Structural Equation Modeling: Past and Present. A Festschrift in Honor of Karl G. Jöreskog* (pp. 195-261). Scientific Software International.

Hauck-Filho, N., & Valentini, F. (2020). A análise de componentes principais é útil para selecionar bons itens quando a dimensionalidade dos dados é desconhecida? *Avaliação Psicológica*, *19*(4). https://doi.org/10.15689/ap.2020.1904.ed

Henson, R., & Roberts, J. (2006). Use of Exploratory Factor Analysis in Published Research. Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement*, *66*(3), 393-416. https://doi.org/10.1177/0013164405282485

Horn, J. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, *30*(2), 179-185. https://doi.org/10.1007/BF02289447

Joshanloo, M., Jose, P. E., & Kielpikowski, M. (2016). The Value of Exploratory Structural Equation Modeling in Identifying Factor Overlap in the Mental Health Continuum-Short Form (MHC-SF): A Study with a New Zealand Sample. *Journal of Happiness Studies*, *18*(4), 1061-1074. https://doi.org/10.1007/s10902-016-9767-4

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141-151. https://doi.org/10.1177/001316446002000116

Kaiser, H. F. (1970). A Second Generation Little Jiffy. *Psychometrika*, *35*(4), 401-415. https://doi.org/10.1007/BF02291817

Kaiser, H. F. (1974). An Index of Factorial Simplicity. *Psychometrika*, *39*(1), 31-35. https://doi.org/10.1007/BF02291575

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4.ª ed.). The Guilford Press.

Kolenikov, S., & Bollen, K. A. (2012). Testing Negative Error Variances: Is a Heywood Case a Symptom of Misspecification? *Sociological Methods & Research*, *41*(1), 124-167. https://doi.org/10.1177/0049124112442138

Ledesma, R. D., Ferrando, P. J., & Tosi, J. D. (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://dx.doi.org/10.6018/analesps.30.3.199361

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2017). The Exploratory Factor Analysis of Items: Guided Analysis Based on Empirical Data and Software. *Anales de Psicología*, *33*(2), 417-432. https://doi.org/10.6018/analesps.33.2.270211

Lord, F. M., & Novick, R. (1968). *Statistical Theories of Mental Tests Scores*. Addison-Wesley Educational Publishers.

Lorenzo-Seva, U. (1999). Promin: A Method for Oblique Factor Rotation. *Multivariate Behavioral Research*, *34*(3), 347-365. https://www.tandfonline.com/doi/abs/10.1207/S15327906MBR3403_3

Lorenzo-Seva, U. (2013). *How to Report the Percentage of Explained Common Variance in Exploratory Factor Analysis. Technical Report*. Universitat Rovira i Virgili, Tarragona.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A Computer Program to Fit the Exploratory Factor Analysis Model. *Behavior Research Methods*, *38*, 88-91. https://doi.org/10.3758/BF03192753

Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2: A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement*, *37*, 497-498. https://doi.org/10.1177/0146621613487794

López-Aguado, M., & Gutiérrez-Provecho, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, *12*(2), 1-14. https://dialnet.unirioja.es/servlet/articulo?codigo=7057076

Magnusson, D. (1972). *Teoría de los tests*. Trillas.

Mardia, K. (1970). Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, *57*(3), 519-530. https://doi.org/10.1093/biomet/57.3.519

Mavrou, I. (2015). Análisis factorial exploratorio. Cuestiones conceptuales y metodológicas. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*, *19*, 71-80. https://doi.org/10.26378/rnlael019283

McDonald, R. P. (1985). *Factor Analysis and Related Methods*. Erlbaum Associates.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Erlbaum Associates.

Morales, P. (2011). *El Análisis Factorial en la construcción e interpretación de tests, escalas y cuestionarios*. Universidad Pontificia Comillas.

Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, *31*(1), 7-16. https://doi.org/10.7334/psicothema2018.291

Nájera, P., Abad, F. J., & Sorrel, M. A. (2023). Is Exploratory Factor Analysis always to be Preferred? A Systematic Comparison of Factor Analytic Techniques Throughout the Confirmatory-Exploratory Continuum. *Psychological Methods*, *30*(1), 16-39. https://doi.org/10.1037/met0000579

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3.[a] ed.). McGraw-Hill.

Olsson, U. (1979). Maximum Likelihood Estimation of the Polychoric Correlation Coefficient. *Psychometrika*, *44*(4), 443-460. https://doi.org/10.1007/BF02296207

Palenzuela, D. (1983). Construcción y validación de una escala de autoeficacia percibida específica de situaciones académicas. *Análisis y Modificación de Conducta*, *9*(21), 185-219. https://doi.org/10.33776/amc.v9i21.1649

Pando, M., & Calderón, J. L. (2018). Inventario de Violencia y Acoso Psicológico en el Trabajo (IVAPT) para Colombia. *Revista Salud Uninorte*, *34*(2), 538-540. https://doi.org/10.14482/sun.34.2.658.48

Pascual-Fernández, M. C., Ignacio-Cerro, M. C., Cervantes-Estévez, L., Jiménez-Carrascosa, M. A., Medina-Torres, M., & García, A. M. (2015). Cuestionario para evaluar la importancia de la familia en los cuidados de enfermería. Validación de la versión española (FINC-NA). *Anales del Sistema Sanitario de Navarra*, *38*(1), 31-39. https://recyt.fecyt.es/index.php/ASSN/article/view/29238/71748

Pérez, E., & Medrano, L. (2010). Análisis factorial exploratorio: bases conceptuales y metodológicas. *Revista Argentina de Ciencias del Comportamiento*, *2*(1), 58-66. https://doi.org/10.32348/1852.4206.v2.n1.15924

Reckase, M. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational Statistics*, *4*(3), 207-230. https://doi.org/10.2307/1164671

Schweizer, K., DiStefano, C., & Troche, S. (2021). Conditions Leading to the Observation of a Difficulty Effect and its Consequence for Confirmatory Factor Analysis. *Psychological Test and Assessment Modeling*, *63*(4), 469-483. https://www.psychologie-aktuell.com/fileadmin/Redaktion/Journale/ptam-2021-4/PTAM__4-2021_2.pdf

Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4-11. https://doi.org/10.12691/ajams-9-1-2

Spearman, C. E. (1904). «General Intelligence» Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201-292. https://doi.org/10.2307/1412107

Stevens, S. S. (1951). Mathematics, Measurement, and Psychophysics. In S. S. Stevens (ed.), *Handbook of Experimental Psychology* (pp. 1-49). John Wiley & Sons.

Streiner, D. L. (1994). Figuring Out Factors: The Use and Misuse of Factor Analysis. *The Canadian Journal of Psychiatry*, *39*(3), 135-140. https://doi.org/10.1177/070674379403900303

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5.[a] ed). Pearson Education.

Thurstone, L. L. (1947). *Multiple Factor Analysis*. University of Chicago Press.

Timmerman, M., & Lorenzo-Seva, U. (2011). Dimensionality Assessment of Ordered Polytomous Items with Parallel Analysis. *Psychological Methods*, *16*(2), 209-220. https://doi.org/10.1037/a0023353

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*, 769. https://doi.org/10.3389/fpsyg.2016.00769

Verdam, M. G., Oort, F. J., & Sprangers, M. A. (2016). Using Structural Equation Modeling to Detect Response Shifts and True Change in Discrete Variables: An Application to the Items of the SF-36. *Quality of Life Research*, *25*(6), 1361-1383. https://doi.org/10.1007/s11136-015-1195-0

Victor-Edema, U. A. (2023). Comparative Analysis of Some Approaches to Multivariate Normality Test. *FNAS Journal of Scientific Innovations*, *4*(2), 154-164.

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A Journey Around Alpha and Omega to Estimate Internal Consistency Reliability. *Anales de Psicología*, *33*(3), 755-782. https://doi.org/10.6018/analesps.33.3.268401

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219-246. https://doi.org/10.1177/0095798418771807

Zwick, R. (1985). *Assessment of the Dimensionality of NAEP Year 15 Reading Data*. Educational Testing Service.

Zwick, W., & Velicer, W. (1986). Comparison of Five Rules for Determining the Number of Components to Retain. *Psychological Bulletin*, *99*(3), 432-442. https://doi.org/10.1037/0033-2909.99.3.432

Sergio Dominguez-Lara

Instituto de Investigación de la Facultad de Ciencias de la Comunicación, Turismo y Psicología, Universidad de San Martín de Porres, Lima, Perú.

Docente en la Escuela de Psicología de la Facultad de Ciencias de la Comunicación, Turismo y Psicología. Sus temas de investigación se vinculan con la metodología y la psicometría.

ORCID: https://orcid.org/0000-0002-2083-4278

sdominguezmpcs@gmail.com