

## El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada

Susana Lloret-Segura\*, Adoración Ferreres-Traver, Ana Hernández-Baeza e Inés Tomás-Marco

*Unidad de Aplicaciones y Desarrollos Psicométricos.*

*IDOCAL. Instituto de Investigación en Psicología de los Recursos Humanos, del Desarrollo organizacional y de la Calidad de Vida Laboral  
 Facultad de Psicología. Universitat de València (España)*

**Resumen:** El Análisis Factorial Exploratorio es una de las técnicas más usadas en el desarrollo, validación y adaptación de instrumentos de medida psicológicos. Su uso se extendió durante los años 60 y ha ido creciendo de forma exponencial al ritmo que el avance de la informática ha permitido. Los criterios empleados en su uso, como es natural, también han evolucionado. Pero los investigadores interesados en asuntos sustantivos que utilizan rutinariamente esta técnica permanecen en muchos casos ignorantes de todo ello. En las últimas décadas numerosos trabajos han denunciado esta situación. La necesidad de actualizar los criterios clásicos para incorporar aquellos más adecuados es una necesidad urgente para hacer investigación de calidad. En este trabajo se revisan los criterios clásicos y, según el caso, se sustituyen o se complementan con otros más actuales. El objetivo es ofrecer al investigador aplicado interesado una guía actualizada acerca de cómo realizar un Análisis Factorial Exploratorio consonante con la psicometría post-Little Jiffy. Esta revisión y la guía con las recomendaciones correspondientes se han articulado en cuatro grandes bloques: 1) el tipo de datos y la matriz de asociación, 2) el método de estimación de factores, 3) el número de factores a retener, y 4) el método de rotación y asignación de ítems. Al final del artículo hemos incluido una versión breve de la guía.

**Palabras clave:** AFE; ACP; matriz de asociación; estimación de factores; número de factores; rotación de factores.

**Title:** Exploratory Item Factor Analysis: A practical guide revised and updated.

**Abstract:** Exploratory Factor analysis is one of the techniques used in the development, validation and adaptation of psychological measurement instruments. Its use spread during the 1960s and has been growing exponentially thanks to the advancement of information technology. The criteria used, of course, have also evolved. But the applied researchers, who use this technique as a routine, remain often ignorant of all this. In the last few decades numerous studies have denounced this situation. There is an urgent need to update the classic criteria. The incorporation of the most suitable criteria will improve the quality of our research. In this work we review the classic criteria and, depending on the case, we also propose current criteria to replace or complement the former. Our objective is to offer the interested applied researcher updated guidance on how to perform an Exploratory Item Factor Analysis, according to the “post-Little Jiffy” psychometrics. This review and the guide with the corresponding recommendations have been articulated in four large blocks: 1) the data type and the matrix of association, 2) the method of factor estimation, 3) the number of factors to be retained, and 4) the method of rotation and allocation of items. An abridged version of the complete guide is provided at the end of the article.

**Key words:** AFE; PCA; matrix of association; factors' estimation; number of factors; factors' rotation.

### Introducción

El Análisis Factorial exploratorio (AFE) de ítems es una de las técnicas más frecuentemente aplicadas en estudios relacionados con el desarrollo y validación de tests, porque es la técnica por excelencia que se utiliza para explorar el conjunto de variables latentes o factores comunes que explican las respuestas a los ítems de un test. Si se hace una consulta acerca de qué se publica en relación a este tipo de análisis en las revistas científicas se encontrarán dos tendencias claramente separadas. Una, la más numerosa, en que la técnica se aplica para identificar la estructura subyacente a los ítems de un test. Este es un uso instrumental de la técnica de análisis. La otra, menos numerosa, en que se estudian y comparan los diferentes criterios que habitualmente se aplican en la realización de un AFE, con otros de nueva creación o no tan nuevos pero menos populares. Estos estudios investigan qué decisiones son más adecuadas, según las diferentes condiciones en que apliquemos esta técnica. En este caso, la técnica misma es el objeto de estudio, y es así porque las fases por las que pasa el análisis factorial requieren la aplicación de criterios de decisión que, como casi todo, han ido cambiando a lo largo del tiempo, a mejor por supuesto. Pero aunque

han ido cambiando, en la actualidad coexisten los criterios tradicionales, algunos francamente caducos, con los más modernos..., y de ahí viene la metáfora del laberinto: el investigador, desconcertado en unos casos por el gran número de posibilidades que han aparecido entre las que debe decidir, o ignorante en otros casos porque desconoce esas nuevas posibilidades, va abriéndose paso tomando decisiones en cada una de las encrucijadas que se encuentra por el camino hasta llegar al final confiando en su experiencia, en lo que han hecho otros antes que él, o en una combinación de ambas, pero sin saber a ciencia cierta si la salida que finalmente encuentra es la principal o la de servicio.

La paradoja que se da en este contexto es la siguiente: los metodólogos buscan y encuentran mejores criterios para aplicar el AFE, mientras advierten con unanimidad que los clásicos son en muchos casos inadecuados y peligrosos. Es decir, abren nuevos y mejores caminos para encontrar la salida del laberinto, y avisan de que otros son engañosos. En cambio, los investigadores que hacen un uso instrumental del AFE parecen desconocer en gran medida que los nuevos caminos son mejores y que los antiguos son peligrosos, por lo que continúan recorriendo el laberinto, ajenos a las indicaciones que les avisan de por donde tienen que ir.

Claramente lo que está fallando, a nuestro juicio, es la comunicación, o la señalización, si pensamos en el laberinto. De un lado, es necesario dar la máxima difusión a estos nuevos criterios, porque hoy día ya se sabe cuándo y por qué funcionan mejor que los anteriores. De otro lado, es también necesario exigir que se apliquen en las investigaciones actua-

**\* Dirección para correspondencia [Correspondence address]:**

Susana Lloret Segura. Departamento de Metodología de las Ciencias del Comportamiento. Facultad de Psicología. Universitat de València. Avda. Blasco Ibáñez, 21. 46022 Valencia (España).  
 E-mail: [Susana.Lloret@uv.es](mailto:Susana.Lloret@uv.es)

les, por lo mismo, porque hoy día se sabe que funcionan mejor que los clásicos. Como señalamos anteriormente, cualquier camino no vale para salir del laberinto.

El objetivo de este estudio es contribuir a difundir y a aplicar estos nuevos estándares (y algunos no tan nuevos, que por ser sistemáticamente ignorados, lo parecen). Lo haremos de tres modos: primero, ofreciendo una revisión y puesta al día de los estándares clásicos, y presentando los nuevos estándares que los sustituyen. Segundo, revisando y resumiendo el grado en que una variedad de software disponible en el mercado -SPSS, FACTOR (Lorenzo-Seva y Ferrando, 2006), LISREL (Jöreskog y Sörbom, 2007), y MPlus (Muthén y Muthén, 2007)- permite o limita la aplicación de estos nuevos estándares. Y tercero, utilizando ese software para analizar la estructura factorial de tres escalas -D-48 (Anstey, 1959. Adaptado por el Dpto. de I+D+i de TEA Ediciones, 1996), Autoestima y Autoconcepto, y Fuerza y Flexibilidad (Marsh, Richards, Johnson, Roche, y Tremayne, 1994), utilizando datos reales para cada una, con la particularidad de que los datos de esas escalas se ajustan de manera inadecuada, moderada o satisfactoria, según la escala, a los supuestos del modelo del AFE clásico. Estos tres casos prácticos con datos reales nos permitirán comparar las consecuencias de elegir entre un software u otro, y también las consecuencias de elegir entre unas opciones u otras dentro de un mismo software, cuando los datos son más o menos "problemáticos". Nuestro objetivo es claro: ilustrar cómo la aplicación adecuada o inadecuada del AFE puede llevar a conclusiones muy diferentes.

Este objetivo no es nuevo, en las últimas dos décadas al menos 6 diferentes estudios de revisión del uso del AFE en la investigación empírica en Psicología han puesto de manifiesto las limitaciones que caracterizaban a los estudios empíricos que se estaban publicando. Los resumimos a continuación.

Ford, MacCallum y Tait (1986) revisaron los artículos publicados en tres revistas (*Journal of Applied Psychology*, *Personnel Psychology*, *Organizational Behavior and Human Performance*) en el periodo comprendido entre 1975-1984 (revisaron un total de 152 estudios). Ford et al. (1986) concluyeron la aplicación del AFE en los estudios revisados resultaba deficiente. Casi 10 años después, Hinkin (1995) revisó el uso del AFE en el desarrollo de 277 escalas. Estas escalas aparecieron en 75 artículos publicados en revistas académicas relevantes en el periodo comprendido entre 1989 y 1994, y desafortunadamente, llegó a una conclusión similar. Fabrigar, Wegener, MacCallum y Strahan (1999) revisaron 217 artículos publicados en dos revistas de investigación psicológica (*Journal of Personality and Social Psychology*, *Journal of Applied Psychology*) en el periodo comprendido entre 1991-1995. En este trabajo se llegó, una vez más, a una conclusión similar a la que llegaron Ford et al. (1986), aunque se observó una tendencia a seguir alguna de las nuevas recomendaciones. Más recientemente, Conway y Huffcutt (2003) realizaron una revisión del uso del AFE en las mismas tres revistas de psicología organizacional que revisaron Ford et al. (1986)

(*Journal of Applied Psychology*, *Personnel Psychology*, *Organizational Behavior and Human Performance*) en el periodo comprendido entre 1985-1999 (se revisaron un total de 371 estudios). Conway y Huffcutt (2003), volvieron a confirmar la tendencia a utilizar criterios más modernos, aunque fue una tendencia minoritaria. Finalmente Henson y Roberts (2006) revisaron 60 artículos publicados en cuatro revistas (*Educational and Psychological Measurement*, *Journal of Educational Psychology*, *Personality and Individual Differences* y *Psychological Assessment*) publicados hasta 1999. Su conclusión fue que la aplicación del AFE en los estudios revisados ofrecía un patrón de errores comunes que se debían evitar.

Recientemente, Izquierdo, Olea y Abad (2013) presentaron una comunicación que analizó el uso del AFE en los años 2011 y 2012 en las tres revistas españolas con mayor factor de impacto continuado en los últimos 5 años: *International Journal of Health and Clinical Psychology*, *Psicothema* y *Spanish Journal of Psychology*. Sus resultados fueron -una vez más- similares: altas tasas de decisiones erróneas o injustificadas.

A continuación analizaremos en detalle cuáles son esas decisiones erróneas, generalmente basadas en criterios o estándares clásicos ya desfasados, y qué otras recomendaciones -guías- usar en su lugar.

## Determinación de la adecuación del Análisis Factorial Exploratorio

En Psicología, cuando hablamos de tests, hablamos en muchos casos de variables latentes o rasgos que causan las respuestas a los ítems de ese test. El objetivo del test es evaluar el grado en que una persona queda caracterizada por un determinado rasgo o variable latente, llámese extraversión, factor "g", o estrés, a partir de las respuestas observadas a un conjunto particular y bien escogido de ítems -las variables observadas-. (Si el test es multidimensional, entonces tendremos múltiples rasgos o variables latentes). El AFE, el Análisis de Componentes Principales (ACP) y el Análisis Factorial Confirmatorio (AFC) son técnicas utilizadas con este propósito. Pero no son intercambiables. ¿Cuándo elegir cada una de ellas? Veámoslo.

### ¿Cuándo aplicamos Análisis Factorial y cuándo Componentes Principales?

La clave está en el objetivo mismo del análisis. Si el análisis pretende identificar el número y composición de los factores comunes (variables latentes) necesarios para explicar la *varianza común* del conjunto de ítems analizado, entonces lo apropiado es aplicar un AFE. En este caso la representación algebraica del modelo para  $m \leq p$  factores comunes tiene como ecuación:

$$X_1 = v_{1(1)} F_{(1)} + v_{1(2)} F_{(2)} + \dots + v_{1(m)} F_{(m)} + e_1$$

$$X_2 = v_{2(1)} F_{(1)} + v_{2(2)} F_{(2)} + \dots + v_{2(m)} F_{(m)} + e_2$$

:

:

$$X_p = v_{p(1)} F_{(1)} + v_{p(2)} F_{(2)} + \dots + v_{p(m)} F_{(m)} + e_p$$

donde  $X_j$ ,  $F_i$  y  $e_j$  contienen la puntuación de una persona en el ítem  $X_j$ , el factor común  $F_i$ , y el factor específico  $e_j$ ,  $m$ : número de Factores comunes,  $p$ : número de ítems,  $F$ : factor común,  $v_{j(i)}$  peso del factor común  $i$ ésimo asociado a la  $j$ ésima variable observada o ítem,  $i=1, 2, \dots, m$ ;  $j=1, 2, \dots, p$ ;  $e_j$ : factor único,  $j=1, 2, \dots, p$ .

En cambio, si el objetivo es identificar el número y composición de componentes necesario para resumir las puntuaciones observadas en un conjunto grande de variables observadas, entonces lo apropiado es aplicar un ACP. Este método explica el máximo porcentaje de varianza observada en cada ítem a partir de un número menor de componentes que resume esa información. En este caso, la representación algebraica del modelo para  $m \leq p$  componentes principales tiene como ecuación:

$$CP_1 = w_{(1)1} X_1 + w_{(1)2} X_2 + \dots + w_{(1)p} X_p$$

$$CP_2 = w_{(2)1} X_1 + w_{(2)2} X_2 + \dots + w_{(2)p} X_p$$

:

:

$$CP_m = w_{(m)1} X_1 + w_{(m)2} X_2 + \dots + w_{(m)p} X_p$$

donde  $X_j$  y  $CP_j$  contienen la puntuación de una persona en el ítem  $X_j$ , y el componente  $CP_j$ ,  $m$ : número de componentes principales,  $p$ : número de ítems o variables observadas,  $X$ : ítems o variables observadas,  $PC$ : componentes principales,  $w_{j(i)}$  peso elegido para la  $j$ ésima variable observada para maximizar la razón de la varianza del  $CP(i)$  respecto de la varianza total,  $i=1, 2, \dots, m$ ;  $j=1, 2, \dots, p$ .

Las diferencias entre la primera opción (AFE) y la segunda (ACP), son evidentes. Las variables observadas (ítems) son las variables independientes en el ACP, pero las variables dependientes en el AFE. A nivel conceptual (y formal) la diferencia es enorme. En cambio, diferentes razones han llevado durante décadas a una gran confusión entre ambas, de forma que sistemáticamente se ha aplicado una de esas técnicas, el ACP, para conseguir el objetivo que corresponde a la otra, al AFE. Veamos por qué.

En primer lugar, ambas técnicas suelen aparecer juntas como técnicas de reducción de datos en los paquetes estadísticos más comunes. Y eso contribuye a que parezcan intercambiables entre sí, pero la realidad es bien diferente. El ACP es consistente con la conceptualización de medidas formativas, en las que los indicadores se consideran causa de un posible constructo (Bollen y Lennox, 1991; Borsboom, Mellenbergh y van Heerden, 2003; Joliffe, 2002), aunque el modelo en sí no haga referencia explícita a variables latentes. De hecho, ni los componentes son variables "latentes", ni los ítems son ninguna "medida" indirecta de ellos. Los componentes son "compuestos" de las variables observadas que cumplen la misión de reproducir el máximo de varianza de cada variable observada con el mínimo número de compuestos. El objetivo está, pues, en la diagonal de la matriz, en las varianzas de cada variable observada. Y como parte del procedimiento para alcanzar este objetivo está la condición de que estos componentes presenten correlaciones nulas entre ellos.

El AFE parte del supuesto de que las variables observadas son indicadores de cierto número de factores o variables latentes comunes. Suponiendo que analizamos un conjunto de ítems seleccionado para medir un único factor, cada variable observada o ítem que se analiza está cuidadosamente seleccionado para que refleje alguna característica del factor que se pretende medir con él. Y por supuesto la idea básica es que personas con diferentes niveles en el factor común, darán diferentes respuestas a ese ítem, justamente porque el factor causa las diferentes respuestas a ese ítem (el ítem es una manifestación de ese factor). La variable independiente es el factor, que produce diferentes respuestas en los ítems. Los ítems son las variables dependientes en este diseño. Es decir, bajo el AFE se asume que las medidas son "reflectivas" o manifestaciones de los constructos subyacentes (Harman, 1976; Kim y Mueller, 1978, Bollen y Lennox, 1991; Edwards, 2011) Además, por muy cuidadosa que sea la selección de los ítems, es imposible que sean medidas perfectas del factor común correspondiente. Una parte de la variabilidad del ítem estará directamente producida por el factor que mide, pero otra parte no. Bajo el modelo clásico o teoría clásica de tests, es posible estimar la parte de la varianza de cada ítem explicada por el factor común subyacente a ese conjunto de ítems, precisamente a partir de la *varianza común* entre ese ítem y el resto de ítems que miden ese mismo factor (a esta parte de la varianza del ítem se le denomina *comunalidad*). El resto de varianza del ítem es *varianza no común* (denominada *unicidad*), no contribuye a la medida de los factores comunes, y consecuentemente no se incluye en el proceso de identificación y estimación de los factores comunes. Aquí radica la segunda gran diferencia entre el AFE y el ACP: la inclusión de un término de error ( $e$ ). Este término de error no existe en el ACP. El ACP no distingue entre *varianza común* y *varianza no común* (ver Joliffe, 2002, capítulo 7, para un tratamiento detallado y asequible de las semejanzas y diferencias entre ambos modelos de reducción de datos).

Por último, destacar que los factores comunes no se dirigen a explicar la máxima cantidad de varianza de cada ítem (como sucede con los componentes), solo la *varianza común de cada ítem con el resto*, y esa varianza común ya no está en los elementos de la diagonal de la matriz, sino en los elementos de fuera de la diagonal, estén expresados en términos de *covarianzas* o de *correlaciones*. Esos son ahora los elementos que trata de explicar el investigador.

### ¿Cuál es la recomendación clásica?

Claramente cuando analizamos los ítems de un test en busca de la estructura factorial que le corresponde, estamos ubicados en el segundo de los escenarios arriba expuestos, es decir, estamos bajo el modelo del análisis factorial. Lo que sucede es que durante algunas décadas la dificultad computacional de los procedimientos orientados a la estimación de factores comunes (FC) característicos del AFE excedían con mucho la del procedimiento de identificación de componen-

tes principales (CP), característico del ACP hasta el punto de que en ocasiones era imposible la obtención de soluciones (o las soluciones que aparecían eran inadecuadas, como cuando aparecían casos Heywood, es decir, ítems con comunalidades de 1 o incluso mayores, o matrices no definidas positivamente). El uso de CP como método de para encontrar una solución inicial que identificara el número de dimensiones necesarias para explicar el conjunto de ítems no sufría estos problemas, y esa solución podía ser rotada en busca de la estructura simple (ortogonal u oblicua) más adecuada, de la misma manera en que se hubiera rotado una solución inicial obtenida por un procedimiento de FC. El método de CP se convertía así en el método de estimación/extracción de “factores” (en realidad componentes) más sencillo y eficaz. Aparecieron estudios que sugerían que era razonable aplicar CP en el contexto de la estimación de factores cuando 1) el número de ítems por factor era alto, y 2) los ítems contenían poco error de medida, ya que en estas condiciones particulares la solución obtenida por un procedimiento u otro eran virtualmente equivalentes (Velicer y Jackson, 1990; Thompson, 2004). Y se extendió la idea de la equivalencia entre ambas técnicas. Una vez mostrada esta equivalencia para un conjunto de condiciones muy particulares, comenzó el uso intensivo de CP como método de extracción de “factores” en el contexto del AFE cuando esas condiciones se daban, y cuando no..., también. No había muchas más alternativas, así que el acuerdo tácito de usar una técnica por otra se universalizó. Y se convirtió en la recomendación clásica.

### ¿Cuál es la recomendación actual?

Hoy día la eficiencia computacional del ACP frente al AFE ya no es tal. Disponemos de nuevas opciones de estimación de factores (como el método de factorización ULS del que luego nos ocuparemos) que hacen posible la aplicación del AFE en condiciones en que antes era imposible. Así que ya no tiene sentido el uso de CP como método de estimación de factores en el contexto del AFE. Por el contrario, existen estudios empíricos y de simulación que desaconsejan enérgicamente el uso de uno por otro porque pueden llevar a soluciones muy diferentes (Ferrando y Anguiano-Carrasco, 2010; Gorsuch, 1997a; Vigil-Colet et al., 2009).

Cuando se aplica en situaciones en las que un modelo factorial es más adecuado, el hecho de emplear CP como método de estimación de factores supone ignorar el error de medida, lo que aumenta espuriamente las cargas factoriales, los porcentajes de varianza explicados por los factores, y puede producir una sobreestimación de la dimensionalidad del conjunto de ítems (ver, por ejemplo, Abad, Olea, Ponso da y García, 2011; Ferrando y Anguiano-Carrasco, 2010; Gorsuch, 1997a). Esto sucede al intentar encontrar componentes que expliquen el total de la varianza (la varianza común más la varianza de error conjuntamente consideradas) en vez de dar cuenta únicamente de la comunalidad. De una u otra forma, la interpretación de la solución obtenida mediante ACP podría ser errónea.

### ¿Por qué entonces se sigue usando?

Evidentemente porque el mensaje aún no ha llegado a todos los interesados, y porque la situación se presta a esa ambigüedad: los programas de análisis factorial incluyen, como hemos comentado ya, el método de CP entre sus opciones de métodos de estimación de factores como una alternativa más, incluso como la opción por omisión (y por lo tanto la de referencia). Y esta tendencia, desafortunadamente, no ha ido a la baja.

Ford et al. (1986) abarcaron el periodo comprendido entre 1975-1984, y encontraron que el 42% de los estudios revisados usaron CP, el 34% AFE, y el resto no informó del modelo empleado. Hinkin (1995), en el periodo comprendido entre 1989 y 1994, encontró que el 33% de los estudios revisados usaron CP, (sólo un 20% aproximadamente usó el modelo del AFE en alguna de sus variantes), y Fabrigar et al. (1999) encontraron que este porcentaje aumentó hasta el 48%.

Recientemente, Conway y Huffcutt (2003) realizaron una revisión del uso del AFE en las mismas tres revistas de psicología organizacional que revisaron Ford et al. (1986) e informaron de que la elección de este modelo siguió siendo la más frecuente en un 39.6% de los casos. Finalmente Henson y Roberts (2006) ofrecieron un porcentaje mayor de uso del ACP en su estudio: 56.7%. En nuestro país, Izquierdo y colaboradores (2013) presentaron la misma tendencia. Estos resultados revelan claramente que es necesario insistir una vez más en la recomendación de que, si lo que se pretende es analizar la correspondencia entre una serie de ítems y el conjunto de factores que pretenden medir esos ítems, hay que aplicar AFE.

### ¿Cuándo aplicamos Análisis Factorial Exploratorio y cuándo Confirmatorio?

El AFE no permite al investigador definir qué ítems mide qué factores, ni tampoco las relaciones que se suponen entre los propios factores, -más allá de si están o no relacionados entre sí-. Se denomina exploratorio porque sólo podemos determinar el número de factores que esperamos, pero no su composición ni las relaciones que cada uno de los factores mantiene con el resto. En cambio, el AFC se caracteriza por permitir al investigador definir cuántos factores espera, qué factores están relacionados entre sí, y qué ítems están relacionados con cada factor.

### ¿Cuál es la recomendación clásica?

La recomendación clásica (e.g., Mulaik, 1972), que todavía sigue vigente (e.g., Matsunaga, 2010), diferencia entre el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC) en función de su finalidad. Desde esta perspectiva, ambos métodos se utilizan para evaluar la estructura factorial subyacente a una matriz de correlaciones, pero mientras el AFE se utiliza para “construir” la teoría, el

AFC se utiliza para “confirmar” la teoría. De este modo, el AFE se utiliza cuando el investigador conoce poco sobre la variable o constructo objeto de estudio, y esta aproximación le ayuda a identificar los factores latentes que subyacen a las variables manifiestas, así como los patrones de relaciones entre variables latentes y manifiestas. Por otra parte, cuando ya se tiene una idea clara sobre las variables objeto de estudio, el uso del AFC permite probar la estructura hipotetizada, poniendo a prueba si el modelo hipotetizado se ajusta adecuadamente a los datos.

### *¿Cuál es la recomendación actual?*

En la actualidad existen dos tendencias. La primera, derivada directamente de la aproximación clásica, recomienda hacer un uso secuencial de ambos tipos de análisis, siempre que el tamaño de la muestra lo permita. Se trata de dividir la muestra aleatoriamente en dos submuestras y explorar en la primera muestra (con un análisis factorial exploratorio, claro) la estructura factorial subyacente a los ítems, para luego tratar de confirmar esa estructura en la otra mitad de la muestra, esta vez mediante un análisis factorial confirmatorio (Anderson y Gerbing, 1988; Brown, 2006).

Desde la segunda tendencia, se cuestiona la distinción entre AFE y AFC con respecto a su finalidad (exploratoria/confirmatoria), indicando que tal diferenciación no es tan clara, y que plantea una serie de problemas (e.g., Ferrando y Anguiano-Carrasco, 2010). Para empezar, es fácil comprender que la mayor parte de las aplicaciones psicométricas del análisis factorial se encuentran en algún punto intermedio entre la ausencia total de información sobre la variables objeto de estudio, y la definición clara de su estructura factorial. Por ello, Ferrando y Anguiano-Carrasco (2010) han propuesto establecer la diferenciación entre las dos aproximaciones, no en función de su finalidad, sino en función de las restricciones impuestas, de modo que el AFE y AFC, en lugar de ser considerados como dos categorías cualitativamente distintas, deberían ser considerados más bien los dos polos de un continuo. Así, el AFE (no restrictivo) impone restricciones mínimas para obtener una solución factorial inicial, que puede ser transformada aplicando diferentes criterios de rotación; y el AFC (restrictivo) impone restricciones mucho más fuertes que permiten poner a prueba una solución única, cuyo ajuste puede ser evaluado utilizando diferentes índices de bondad de ajuste (Hu y Bentler, 1999; Kline, 2005; Marsh, Hau y Wen, 2004).

Continuando con esta segunda tendencia, diferentes autores han puesto de manifiesto que el AFC falla en la confirmación de estructuras factoriales claramente apoyadas por los correspondientes análisis exploratorios, porque es demasiado restrictivo (e.g. Ferrando y Anguiano-Carrasco, 2010; Ferrando y Lorenzo-Seva (2000); Marsh et al., 2009, 2010, 2011, 2014). Recuérdese que en el modelo del AFC, algunas de las hipótesis plantean que las saturaciones de los ítems en los factores que supuestamente no miden son nulas. Ésta es una restricción que no es realista en muchos casos, espe-

cialmente si los factores están correlacionados. Este problema ha sido discutido bastante a fondo en el trabajo de Ferrando y Lorenzo-Seva (2000), y como indican los autores, se trata principalmente de una acumulación de errores de especificación. La mayor parte de ítems que se analizan en la práctica no actúan como marcadores (es decir, no son factorialmente simples), y por lo tanto, presentan cargas cruzadas menores, pero no nulas, en los otros factores que supuestamente no miden. Cuando estas cargas cruzadas se fuerzan a cero, como sucede en el AFC, el ajuste del modelo se deteriora, ya que los residuales se van acumulando con cada error de especificación. El deterioro del ajuste es mayor cuanto más largo el test (más errores acumulados) y más grande la muestra (mayor potencia).

Para resolver este problema, la recomendación actual presenta diferentes alternativas. Una de ellas es el ESEM (*Exploratory Structural Equation Modeling*), que representa un híbrido entre el AFE y el AFC. Se puede decir que el ESEM es una alternativa semi-confirmatoria, ya que se encuentra a medio camino entre ambas estrategias de análisis en el continuo de restricción que se indicaba más arriba, y que integra las ventajas de ambas aproximaciones.

El ESEM (e.g., Marsh et al., 2011; Morin, Marsh y Nagengast, 2013) es una aproximación similar al AFC, pero donde las saturaciones de los ítems en los factores que “no les corresponden”, -i.e. las saturaciones en los factores que supuestamente no miden (o cross-loadings)-, no se fijan a cero. Por tanto, es menos restrictivo que el AFC, y se ajusta más a la realidad de las medidas en Psicología (donde habitualmente los ítems no son “marcadores” o “indicadores” perfectos del constructo que miden, sino que es habitual que los ítems sean factorialmente complejos y presenten pesos menores, aunque no nulos, en los otros factores). Por ello, el ESEM consigue mejores ajustes que el AFC. Podríamos decir que esta aproximación se aleja de la “estructura simple” propuesta por Thurstone (1947) en los inicios del AFE, que tanto tiempo ha dominado el panorama para acercarse más a la tendencia actual de proponer modelos que se aproximen suficientemente a la realidad, pero sin pretender reproducirla, porque es imposible, o implausible.

Otra alternativa, formulada inicialmente por Mulaik (1972) es especificar una solución directamente interpretable utilizando uno o dos marcadores por factor. Esta alternativa ha sido considerada también más recientemente por McDonald (1999; 2000; 2005) con el nombre de “independent-cluster basis” (I-CB). El concepto de I-CB hace referencia a una solución factorial para un test multidimensional, en la que cada factor estará definido por un pequeño número de ítems factorialmente simples (marcadores). En concreto, el requisito es que haya por lo menos 3 marcadores por factor si los factores no están correlacionados, y como mínimo 2 marcadores por factor si los factores están correlacionados (McDonald, 1999). En una solución semi-restringida, el resto de ítems podrían ser factorialmente complejos. Tal como indican Ferrando y Lorenzo-Seva (2013), el cumplimiento de la condición I-CB es suficiente para identificar una solución

sin indeterminaciones rotacionales, y es muy ventajosa en términos de interpretación.

Finalmente, otra alternativa son las rotaciones semi-especificadas o Procusteanas contra una matriz diana, que se pueden aplicar con el programa FACTOR. El lector interesado puede consultar una aplicación práctica de esta alternativa en el trabajo de Ferrando, Varea, y Lorenzo (1999).

No existe pues una recomendación universal respecto a cuándo aplicar AFE o AFC, pero lo que está claro es que sea uno u otro el modelo o la secuencia de modelos elegida, el investigador debe justificar su decisión adecuadamente.

### Aspectos relacionados con el diseño: selección de ítems, tamaño y composición de la muestra, y número de ítems por factor

Como en cualquier investigación, la utilidad y generalizabilidad de los resultados obtenidos con un AFE dependerá de la adecuación del diseño de la investigación, es decir, de la selección de las variables que se van a medir, del procedimiento de muestreo que se va a emplear, y del tamaño de la muestra, entre otras decisiones. En cambio los estudios empíricos suelen descuidar esta fase de la investigación (Ferrando y Anguiano-Carrasco, 2010). Veamos detalladamente cada uno de estos aspectos.

#### Tamaño y composición de la muestra

*¿Cómo se seleccionan los ítems más adecuados para construir el test?*

Uno de los primeros aspectos que ha de decidirse es el subconjunto de ítems, de entre todos los posibles, que va a configurar la versión inicial del test. Si el subconjunto de ítems omite aspectos relevantes de la variable latente que se desea medir, habrá menos varianza común de la que debiera en el análisis consiguiente, y los factores comunes resultantes serán más débiles porque estarán insuficientemente definidos. Si, por el contrario, se introducen ítems irrelevantes, aparecerán factores comunes adicionales o se dificultará la emergencia de los factores comunes objeto de medida. Así que la selección más o menos adecuada de los ítems que se van a incluir juega un papel determinante en la claridad de la estructura factorial identificada.

#### Recomendación clásica:

La recomendación clásica en la selección de los ítems consiste en definir claramente y de forma exhaustiva el constructo a medir y a partir de esa definición seleccionar los ítems de forma que cubran todo los aspectos relevantes en esa definición. Esto es lo que desde el punto de vista clásico se recoge bajo el epígrafe de validez de contenido. Además, también se recomienda utilizar algunos criterios empíricos obtenidos durante la fase de análisis de los ítems, como son

el índice de homogeneidad corregido (correlación ítem-total sin el ítem analizado) y el coeficiente alfa si se elimina el ítem de la escala (o subescala, dependiendo de si el test es unidimensional o está formado por varias subescalas).

#### Recomendación actual:

Las recomendaciones clásicas siguen vigentes actualmente. Lo novedoso es que además se han incorporado nuevas recomendaciones que vienen a depurar aún más el proceso de selección de ítems. La experiencia acumulada revela la importancia de aspectos de tipo sustantivo como metodológico que resumimos a continuación:

A) El uso/abuso de ítems redundantes deteriora la estructura factorial resultante. Los ítems redundantes son ítems que expresan la misma idea con una redacción mínimamente distinta. Tradicionalmente, se usan para evaluar la consistencia de las personas, pero también para elevar la consistencia interna de las escalas (Ferrando y Anguiano-Carrasco, 2010). El problema surge porque estos ítems redundantes comparten, como es natural, más varianza que la que es directamente explicada por el factor común. También parte de la varianza única en estos pares o tripletes de ítems redundantes es compartida. Y cuando esto sucede, aparecen factores comunes adicionales difíciles de identificar y de explicar, especialmente tras rotar la solución inicial.

B) Las distribuciones de los ítems y el número de opciones de respuesta también juegan un papel determinante. Si se va a analizar la matriz de correlaciones de Pearson o la de covarianzas (porque el programa que vamos a usar no permite otra cosa), los ítems han de ser variables continuas. Y si no lo son, como sucede cuando los ítems a analizar son politómicos (p.e. tipo Likert), la recomendación que permite aproximarnos adecuadamente al supuesto de continuidad en este caso es utilizar ítems con al menos cinco alternativas de respuesta y con distribuciones aproximadamente normales. Ítems con menos categorías de respuesta o con distribuciones no normales deben analizarse como corresponde a su naturaleza ordinal, es decir, usando la matriz de correlaciones policórica (para ítems politómicos, o tetracórica, en caso de ítems dicotómicos) (Bandalos y Finney, 2010). Más adelante volveremos sobre este punto.

#### *¿Cuántos ítems conviene incluir por cada factor?*

Guadagnoli y Velicer (1988), MacCallum, Widaman, Preacher y Hong (2001) y MacCallum, Widaman y Zhang y Hong (1999) mostraron que el número de ítems por factor interactúa con el tamaño de las comunalidades de los ítems y con el tamaño de la muestra.

La práctica habitual es seleccionar tres ítems por factor como mínimo. Sin embargo, esta práctica es contraproducente, al comprometer la estabilidad de los resultados (Velicer y Fava, 1998), especialmente cuando el tamaño de la

muestra está por debajo de 150 personas (Costello y Osborne, 2005).

Como norma general, cuántos más ítems existan y midan con precisión un factor, más determinado estará el citado factor y más estable será la solución factorial. Los estudios revisados apuntan un mínimo de 3 o 4 ítems por factor, solo si se dispone de un mínimo de 200 casos (Fabrigar et al., 1999; Ferrando y Anguiano-Carrasco, 2010).

### Idoneidad de la muestra

*¿Qué tamaño de muestra mínimo necesito? ¿Qué aspectos de su composición hay que considerar?*

¿Que muestra se necesita para que una solución sea estable y generalizable? Este es un problema complejo. La cuestión del tamaño y composición muestral en el AFE ha sido objeto de estudio por los investigadores durante décadas. Es habitual utilizar muestras de conveniencia (generalmente estudiantes universitarios), pero tienen dos problemas: la no-representatividad y la atenuación por restricción de rango (Ferrando y Anguiano-Carrasco, 2010). Resulta obvio que las muestras más grandes son mejores que las pequeñas, pero no siempre el investigador puede acceder a tamaños muestrales grandes, lo que ha propiciado un extenso número de estudios ofreciendo un amplio listado de recomendaciones no siempre acertadas acerca del tamaño muestral mínimo en el AFE (Guadagnoli y Velicer, 1988; Hogarty, Hines, Kromrey, Ferron y Mumford, 2005).

#### Recomendación clásica:

Muchas de estas recomendaciones sobre el tamaño muestral han sido objeto de estudio a través de numerosos estudios empíricos y de simulación (Arrindell y van der Ende, 1985; MacCallum et al., 1999; Velicer y Fava, 1998), dónde se distinguen dos enfoques diferentes: 1) los que sugieren un tamaño mínimo (N), y 2) los que defienden la proporción de personas por ítem (N/p).

Algunos estudios clásicos del primer enfoque (Barrett y Kline, 1981; Guadagnoli y Velicer, 1988) sugieren un tamaño muestral mínimo (N), que oscila entre 50 y 400 sujetos. Comrey y Lee (1992, p. 217) sugirieron que: “la adecuación del tamaño muestral podía ser evaluada con la escala siguiente: 50 -muy deficiente; 100 -deficiente; 200 -aceptable; 300 -bueno; 500 -muy bueno, 1000 o más -excelente”. Así que una de las recomendaciones clásicas por excelencia es que un tamaño de 200 casos o más (excepto en muestras clínicas) es suficiente para la mayor parte de los análisis descriptivos y psicométricos de los ítems, si el test a validar no es muy largo. Aunque se recomienda alcanzar los 500 o más casos, siempre que sea posible (MacCallum et al. 1999).

Por otra parte, desde el segundo enfoque, basado en la ratio personas/ítems (N/p), la recomendación más habitual es la “regla de los 10”: una muestra 10 veces mayor que el número de ítems (Velicer y Fava, 1998), y la proporción de

5:1. En un trabajo clásico Gorsuch (1983) sugirió una ratio de 5 sujetos por variable, y un tamaño no menor de 100 personas.

#### Recomendación actual:

En la actualidad, diferentes estudios de simulación han revelado que el tamaño muestral es un factor que interactúa con otros aspectos del diseño y de la naturaleza de los datos, como es la matriz que sirve de input al AFE, el número de ítems que definen el factor, la homogeneidad de la muestra y, muy especialmente, la comunalidad de los ítems (Fabrigar et al., 1999; Ferrando y Anguiano-Carrasco, 2010; Beavers et al., 2013).

Respecto a la matriz input, se distingue entre la matriz de correlación producto-momento y la matriz de correlación policórica. La primera son medidas calculadas directamente sobre datos empíricos, mientras la de correlaciones policóricas se obtiene a partir de estimadores indirectos de modo iterativo y, en general, son mucho más inestables. Por tanto, un AFE basado en la matriz de correlaciones policóricas requerirá más muestra que un AFE basado en productos-momento para alcanzar el mismo nivel de precisión y estabilidad, si el resto de condicionantes se mantiene constante.

Respecto al resto de aspectos, podemos resumir las recomendaciones como sigue:

- 1) Condición óptima: cuando las saturaciones son superiores a .70, y el número de variables por factor es adecuado –al menos 6 ítems por factor–, un tamaño muestral de 150 o 200 casos parece suficiente para obtener estimaciones precisas de los coeficientes en el AFE (MacCallum et al. 1999; Preacher y MacCallum, 2003). Incluso hay evidencia de que es suficiente con 100 casos (cuando hay tres factores con tres o cuatro ítems cada uno, o cuando hay más ítems y factores pero las comunalidades son superiores a .80 (Bandalos y Finney, 2010; Costello y Osborne, 2005; Guadagnoli y Velicer, 1988)).
- 2) Condición moderada: cuando disponemos de comunalidades entre .40 y .70, y el número de variables por factor es de 3-4 ítems, también se acepta un tamaño de 200 casos, y por último,
- 3) Condición mínima: cuando las comunalidades son bajas, en torno a 0.30, y el número de variables por factor es de 3 ítems, se precisa una muestra mínima de 400 casos (Conway y Huffcutt, 2003; Gorsuch, 2003), incluso de 500 o más para conseguir estimaciones suficientemente precisas (Hogarty et al., 2005).

Actualmente, los criterios clásicos del tipo N/p y las “recetas” tradicionales tipo: 10 veces más sujetos que ítems entre otras, se desaconsejan por completo, ya que no tiene ninguna base sólida (Bandalos y Finney, 2010; Ferrando y Anguiano-Carrasco, 2010). De hecho, no existe una receta clara, ya que el tamaño mínimo recomendado depende de todos los factores enumerados. Lógicamente, cuanto mayor sea el tamaño

muestral del que dispongamos, más confianza tendremos en que la solución obtenida sea estable, especialmente si la comunalidad es baja, o cuando tengamos muchos posibles factores a extraer y/o pocos ítems por factor. Una cosa sí parece clara: si alguien quiere evaluar la calidad de un test se recomienda un tamaño muestral de al menos 200 casos como mínimo, incluso en condiciones óptimas de comunalidades elevadas y factores bien determinados (Ferrando y Anguiano-Carrasco, 2010).

### **Tipo de datos y matriz de asociación. Adecuación de los datos al Análisis Factorial**

El AF clásico se ha desarrollado sobre el supuesto de que los ítems se relacionan linealmente con los factores que miden, y además que las relaciones entre ellos son también lineales (existen variantes de AF no lineal, pero no son el objeto de este trabajo). Si además las variables son continuas o se aproximan suficientemente a esta condición, la matriz de correlaciones producto-momento de Pearson, o la matriz de varianza-covarianza, resumirá de forma adecuada las relaciones entre los ítems (mucho más extendida la primera que la segunda, sobre todo en el AFE, ver Brown (2006), pp 40-42). En cambio, en Psicología es muy frecuente utilizar la matriz de correlaciones de Pearson o la matriz de varianza-covarianza obtenida sobre ítems ordinales sin determinar si sus distribuciones cumplen este requisito. De hecho algunos programas no permiten usar otro tipo de matriz.

#### *Recomendación clásica:*

La recomendación clásica está dirigida a la matriz de correlaciones de Pearson como matriz de input. Este tipo de correlación es adecuada para evaluar la relación lineal entre dos variables continuas y preferentemente de distribución normal. Los ítems no son variables continuas, sino ordinales y discretas. Pero cuando su distribución es aproximadamente normal, el coeficiente de correlación de Pearson sigue siendo una buena manera de estimar la relación dos ítems. ¿Y cuándo es esa distribución aproximadamente normal? En el caso de ítems dicotómicos cuando los índices de dificultad son moderados y homogéneos (entre .40 y .60). O lo que es lo mismo, los ítems presentan dificultades intermedias y consecuentemente distribuciones “simétricas”. Los ítems con distribuciones asimétricas son especialmente problemáticos si además las asimetrías aparecen en ambas direcciones, porque dan lugar a relaciones no lineales. Este problema, típico de tests con ítems muy fáciles y muy difíciles, está muy documentado en la literatura y da lugar a lo que se ha denominado como “factores de dificultad”, factores extra que aparecen justamente por efecto de la asimetría de las distribuciones de los ítems (Embretson y Reise, 2000; Ferrando y Anguiano-Carrasco, 2010; McDonald y Ahlwardt, 1974).

Si los ítems son politómicos (tipo Likert), como ya anticipamos, se considera que la aproximación a la condición ideal de continuidad es razonablemente adecuada cuando el

número de opciones de respuesta es de 5 o más, y de nuevo la distribución de los ítems es aproximadamente normal. En este caso la asimetría también es un problema, porque se atenúan las relaciones lineales entre ítems que miden el mismo rasgo, e incluso pueden dar lugar a relaciones no lineales (consultar West, Finch y Curran, (1995) para una explicación detallada de este aspecto). Otro factor a tener en cuenta es la discriminación de los ítems. Los ítems con discriminaciones intermedias no representan problema, pero cuando las discriminaciones son extremas el coeficiente de correlación de Pearson no es adecuado.

Estas recomendaciones clásicas, en cambio, han sido sistemáticamente ignoradas en trabajos clave en la actualización de los criterios de decisión empleados en el AFE (Conway y Huffcutt, 2003; Fabrigar et al., 1999; Henson y Roberts, 2006; Park, Dailey y Lemus, 2002; Pérez y Medrano, 2010). Estos trabajos no contemplan el tipo de ítems que se analizan ni por lo tanto el tipo de matriz a utilizar. En estos estudios sólo se ha contemplado la posibilidad de analizar la matriz de correlaciones producto-momento de Pearson, que como vemos no es siempre la opción adecuada. Esto, junto con el hecho de que el programa más usado, SPSS, tampoco incluya en su configuración original la posibilidad de analizar otro tipo de matriz que no sea la matriz de correlaciones producto-momento (o la matriz de varianzas-covarianzas), ha contribuido a que la práctica habitual sea calcular la matriz de correlaciones producto-momento de Pearson sobre ítems ordinales, dicotómicos o politómicos, sin estudiar previamente las distribuciones de los ítems.

#### *Recomendación actual:*

En la actualidad se sigue recomendando revisar las distribuciones de los ítems como paso previo, pero ahora además es posible usar otro tipo de matrices de asociación si es conveniente: policórica (para ítems politómicos) o tetracórica, (en caso de ítems dicotómicos), (Brown, 2006; Bandalos y Finney, 2010, Ferrando y Anguiano-Carrasco, 2010). Como en otras ocasiones, la aproximación a la normalidad que se exige a las variables politómicos es mayor o menor dependiendo del autor que consultemos. Algunos autores recomiendan las distribuciones con coeficientes de asimetría y curtosis en el rango (-1,1) (e.g., Ferrando y Anguiano-Carrasco, 2010; Muthén y Kaplan, 1985, 1992). Otros en cambio consideran aceptable valores en el rango (-1.5, 1.5) (Forero et al., 2009), o incluso el rango [-2, 2] (Muthén y Kaplan, 1985, 1992; Bandalos y Finney, 2010). En lo que sí hay acuerdo, como comentamos con anterioridad, es en que el impacto negativo de la asimetría interactúa con otros factores como el tamaño de la muestra o el número de ítems que definen cada factor (West, et al., 1995; Forero et al., 2009) de modo que la exigencia debe crecer a medida que el resto de condiciones se vuelven más desfavorables.

Se recuerda aquí que un AFE basado en la matriz de correlaciones policóricas requerirá más muestra que un AFE basado en la matriz de correlaciones productos-momento de

Pearson para alcanzar el mismo nivel de precisión y estabilidad, por eso, si la muestra es pequeña (200 sujetos) y las distribuciones son adecuadas se recomienda factorizar la matriz de correlaciones de Pearson. En caso de duda puede ser clarificador realizar el análisis sobre las dos matrices.

Algunos ejemplos de software específico que permite analizar la matriz de correlaciones más adecuada es LISREL (Jöreskog y Sörbom 2007) y MPlus (Muthén y Muthén 1998-2012). También existe software específico y de libre distribución, como FACTOR (Lorenzo-Seva y Ferrando, 2006) o el paquete “Psych” en R (Revelle, 2014). Además, es resaltable que los usuarios de SPSS cuentan con TETRA-COM (Lorenzo-Seva y Ferrando, 2012a), un programa para SPSS que estima la matriz tetracórica. Recientemente ha aparecido una rutina desarrollada por Basto y Pereira (2012), que permite también implementar la matriz de correlaciones policórica en el análisis con SPSS.

Por último, pero no menos importante, cualquiera que sea el caso, se recomienda analizar también las distribuciones bivariadas de cada par de ítems, como señalaron recientemente Pérez y Medrano (2010), para identificar patrones de relaciones no lineales entre los ítems. Estos patrones violarían el supuesto de linealidad del AFE y de nuevo añadirían ruido a la matriz a analizar y confusión a la estructura factorial identificada.

### Adecuación de los datos al Análisis Factorial

Cualquiera que sea la matriz que se vaya a factorizar, es necesario previamente comprobar su grado de adecuación al AF. Una de las formas más comunes de evaluar este aspecto es a través del cálculo de la medida KMO de Kaiser (1970). Esta refleja la influencia de todos los factores que hemos comentado: tamaño de las correlaciones entre los ítems, tamaño de la muestra, número de factores y número de ítems. Esta medida de adecuación indica cuán grande es la correlación entre las variables medidas. Si las correlaciones son suficientemente grandes, la matriz se considera adecuada para su factorización porque ofrecerá resultados estables, replicables en otras muestras diferentes, independientemente del tamaño de la muestra, o del número de factores, o del número de ítems. De otro modo, si KMO es suficientemente grande, los resultados no serán casuales. Kaiser consideraba una matriz con valores para KMO por debajo de .50 inadecuada para el AF; mediocre si estos valores oscilaban entre .60 y .69; y satisfactoria solo valores de .80 en adelante. En cambio, en nuestra experiencia son muchos autores que han recogido únicamente el primero de los valores (.50), como punto de corte (véase por ejemplo Ferguson y Cox, 1993; Hair, Anderson, Tatham y Black, 2005; Tabachnick y Fidell, 2001). Otros autores consideran más adecuado incrementar el listón a .70 e incluso .80 (junto por ejemplo Costello y Osborne, 2005; Ferrando y Anguiano-Carrasco, 2010). Los resultados del estudio empírico que presentaremos en la segunda parte de este trabajo apoyan este valor.

### ¿Cuál es el método de estimación de factores más adecuado?

Los métodos de estimación de factores recomendados habitualmente son Máxima Verosimilitud (MV) y Mínimos Cuadrados Ordinarios (MCO), si bien este último, en realidad, aglutina un conjunto de métodos entre los que destacan ejes principales y mínimos cuadrados no ponderados, basados en un mismo principio: minimizar la suma de cuadrados de las diferencias entre las correlaciones observadas y las reproducidas por el modelo; es decir hacer que los residuales sean lo más próximos a 0 posible (Ferrando y Anguiano-Carrasco, 2010).

Máxima verosimilitud: Este método es inferencial (Lawley y Maxwell, 1971). Es un método de estimación factorial que proporciona las estimaciones de los parámetros que con mayor probabilidad han producido la matriz de correlaciones observada, si la muestra procede de una distribución normal multivariada con  $m$  factores latentes. Las correlaciones se ponderan por la inversa de la unicidad de las variables, y se emplea un algoritmo iterativo para la estimación de los parámetros. Este método, frente a otros, tiene la ventaja de que permite contrastar el ajuste del modelo a los datos a través de un índice que sigue una distribución ji-cuadrado, y obtener los errores típicos y pruebas de significación alrededor de los parámetros estimados.

Una de las desventajas de MV, sin embargo, es que requiere el cumplimiento del supuesto de normalidad multivariada. Algunos autores como Finney y DiStefano, (2006) recomiendan contrastar este supuesto de normalidad multivariada -como hace el programa FACTOR (Lorenzo-Seva y Ferrando, 2006) y los programas del entorno de los modelos de ecuaciones estructurales MEE-. Yuan y Bentler (1998) son menos estrictos, y sugieren que este supuesto es poco realista en psicología. En cualquier caso, existe abundante literatura basada en estudios de simulación que muestra que el método de MV es robusto al incumplimiento de este supuesto cuando las variables tienen una distribución univariada aproximadamente normal (e.g., Muthén y Kaplan, 1985, 1992; West et al., 1995; Forero, Maydeu-Olivares, y Gallardo-Pujol, 2009). Otra de las desventajas de este método de estimación es que el índice referido a la distribución ji-cuadrado es muy sensible al tamaño muestral, (Tabachnick y Fidell, 2001; Brown, 2006). Además, Ferrando y Anguiano-Carrasco (2010) recuerdan que esta prueba asume que el modelo propuesto con  $m$  factores se ajusta perfectamente a la población, y por tanto, que todo el error es muestral, y no que una parte de ese error sea error por aproximación (el grado en que el modelo es una aproximación razonable a lo que sucede en la población, dados los datos observados en la muestra). Todo ello lleva a rechazar modelos que sí suponen una buena aproximación a la estructura factorial latente, en favor de modelos con más factores de los que tienen significado teórico, es decir, modelos sobrefactorizados. Por eso, en la práctica se suelen considerar otros indicadores de ajuste

te derivados del test ji-cuadrado que evalúan el error de aproximación y el grado de ajuste del modelo. De ellos nos ocuparemos en otra sección.

En consecuencia,

- A) se recomienda usar el método MV, factorizando la matriz de correlaciones producto-momento de Pearson, si los ítems tienen un número suficiente de categorías de respuesta (5 o más), o son continuos (cosa bastante más improbable), y además se cumple razonablemente el supuesto de normalidad, para que pueda observarse la relación lineal asumida en las relaciones bivariadas entre los ítems (Flora, LaBrish y Chalmers, 2012). También se recomienda considerar otros indicadores de ajuste derivados del test ji-cuadrado que evalúen el error de aproximación y el grado de ajuste del modelo.
- B) No se recomienda analizar las matrices de correlaciones policóricas mediante MV, porque aunque las estimaciones de los parámetros son generalmente insesgadas (Bollen, 1989), e incluso reproducen mejor el modelo de medida que el análisis de correlaciones producto-momento cuando algunos de los ítems incumplen el supuesto de normalidad (ver Holgado-Tello et al., 2010) hay que tener en cuenta que las pruebas de bondad de ajuste basadas en ji-cuadrado, así como los errores típicos (y por lo tanto las pruebas de significación de los parámetros) están sesgados (Bollen, 1989; Satorra y Bentler, 1994), por lo que debería evitarse la interpretación de los mismos. Cuando se analiza la matriz de correlaciones policóricas se recomienda usar MCO (e.g. Flora, et al., 2012; Forero et al., 2009; Lee, Zhang y Edwards, 2012)

**Mínimos Cuadrados Ordinarios:** Aquí se agrupan una serie de métodos descriptivos que tienen como denominador común que determinan la solución factorial que hace que los residuales sean tan próximos a cero como sea posible. El uso de estos métodos ha mostrado buenos resultados en la factorización de ítems ordinales cuando se analiza la matriz de correlaciones policóricas (Forero et al., 2009; Lee et al., 2012).

Entre estos métodos el método de Ejes Principales ha sido la opción clásica recomendada cuando no se cumple el supuesto de normalidad, lo cual es más y más probable según se reduce el número de categorías de respuesta (Fabrigar et al. 1999). El método de ejes principales se puede aplicar de forma no iterativa, empleándose como estimaciones más habituales de las comunalidades la correlación múltiple al cuadrado entre cada variable observada y el resto. Estas sustituyen los valores diagonales de la matriz de correlaciones, dando lugar a lo que se denomina matriz de correlaciones reducida, que es el input para el análisis factorial. Sin embargo, en la mayoría de programas, en sus versiones más recientes, el método de ejes principales se aplica por defecto de forma iterativa. Tras estimar las comunalidades iniciales, la matriz reducida se auto-descompone en sus autovalores y autovectores; estos autovectores se reescalan y pasan a for-

mar la matriz factorial en la primera iteración. A partir de esta matriz factorial se reestiman las comunalidades, que sustituyen a las iniciales. El proceso iterativo continúa hasta que las diferencias entre las comunalidades de dos iteraciones sucesivas son tan pequeñas que se alcanza el criterio de convergencia, o hasta que se llega a un número máximo de iteraciones. El objetivo es conseguir la mejor estimación posible de las comunalidades a partir del número de factores retenidos. La adecuación del método de ejes principales, en cualquier caso, depende de la calidad de las estimaciones iniciales de las comunalidades. Cuando se aplica de forma iterativa, se consiguen resultados más similares a los obtenidos mediante otros métodos como el método de mínimos cuadrados no ponderados (ULS -Unweighted Least Squares, Joreskog, 1977). Este método, en vez de usar la matriz reducida como input, con las comunalidades estimadas en la diagonal, minimiza la suma de los cuadrados de las diferencias entre las matrices de correlaciones observadas y reproducidas. Este método es el más recomendado actualmente (ver por ejemplo Flora et al., 2012), porque funciona bien cuando se trabaja con muestras pequeñas incluso cuando el número de variables es elevado, especialmente si el número de factores a retener es pequeño (Jung, 2013). Además evita la aparición de casos Heywood (saturaciones mayores que la unidad y varianzas de error negativas), más frecuentes con otros métodos de estimación. Cabe señalar que el método ULS es virtualmente intercambiable con el método de Residuales Mínimos (MinRes) (Harman y Jones, 1966, Joreskog, 1977). Dependiendo del software que se utilice, dispondremos de un método de estimación u otro.

Ferrando y Anguiano-Carrasco (2010) afirmaron que “en una situación en que (a) las variables tienen distribuciones aceptables, (b) la solución está bien determinada, y (c) el modelo propuesto es razonablemente correcto, las soluciones MCO y MV serán prácticamente idénticas. En este caso, MV tiene la ventaja de que permite obtener indicadores adicionales muy útiles en la evaluación del ajuste.” (p. 28) En caso contrario, aparecerán problemas de convergencia y estimaciones inaceptables e indicadores poco fiables. En estos casos los autores recomiendan utilizar MCO.

Cabe añadir que si bien se han propuesto algunos métodos de estimación robustos a la violación del supuesto de normalidad multivariada (e.g., Mínimos Cuadrados Ponderados (WLS en inglés), Mínimos Cuadrados Ponderados Robustos (o WLSMV en inglés) o Máxima Verosimilitud Robusta), éstos se han aplicado más en modelos confirmatorios que en exploratorios. Algunos programas como Mplus sí tienen implementados algunos de estos métodos tanto para el AFC como para el AFE. Estos métodos robustos (la utilización de uno u otro dependería del tamaño muestral) son los más recomendados cuando se analizan datos ordinales alejados de la normalidad en los modelos de ecuaciones estructurales en general, y en el AFC, en particular (e.g., Curran, West y Finch, 1996; Flora y Curran, 2004). Sin embargo, en el caso del AFE todavía es necesario realizar estudios de simulación que permitan comparar las ventajas de estos

métodos robustos frente a otros como ULS (más allá de la posibilidad de ofrecer índices de bondad de ajuste) o incluso MV, cuando los datos sigan una distribución que se desvíe de la normalidad. Lo mismo puede decirse de las estimaciones bayesianas (Muthen y Asparouhov, 2012).

### ¿Cómo se selecciona el número de factores más adecuado?

Este es quizá de todos, el aspecto más determinante de un AFE. El número de factores comunes que hacen falta para explicar las relaciones entre los ítems, y la composición de esos factores, son las dos cuestiones centrales en la interpretabilidad de la estructura factorial obtenida en el análisis. Si se retienen menos factores de los debidos, los patrones de saturaciones resultantes se vuelven más difíciles de interpretar, y por lo tanto los factores identificados resultan confusos, y si se retienen más, entonces se están “fabricando” variables latentes con poco sentido teórico o sustantivo. La decisión es importante, y es quizá por ello uno de los aspectos en que más opciones han aparecido. Veámoslo.

#### *Recomendación clásica:*

La recomendación clásica en este punto es utilizar la regla de Kaiser: seleccionaremos los factores con valores propios mayores que 1 que extraemos de la matriz de correlaciones original (y no de la matriz reducida, es decir, la matriz con las comunalidades en la diagonal). La regla de Kaiser se ha venido usando de forma rutinaria dentro del pack conocido como “Little Jiffy” que comprende la aplicación del ACP con la regla de Kaiser y el método de rotación Varimax. Esta regla fue propuesta por Kaiser en 1957 como alternativa computacionalmente eficaz a las opciones más adecuadas, pero computacionalmente mucho menos eficientes basadas en AFE y en rotaciones oblicuas. Claro que eso fue en 1957, y mucho han cambiado las cosas como para que sigamos empleando aún los mismos procedimientos. Sin embargo, los estudios que han revisado la forma en que se viene aplicando el AFE en la investigación desde 1975 hasta actualidad muestran cómo este criterio es el más utilizado solo o en combinación con algún otro (Conway y Huffcutt, 2003; Fabrigar et al., 1999; Henson y Roberts, 2006; Park et al., 2002; Pérez y Medrano, 2010) tanto cuando se utiliza ACP como cuando se utiliza AFE. Afortunadamente, los estudios más recientes señalan que la frecuencia con que se usa esta regla como único criterio está disminuyendo en favor del aumento de casos en que se utiliza en combinación con otros criterios (Conway y Huttcutt, 2003; Henson y Roberts, 2006).

El criterio que más se combina con la regla de Kaiser es el gráfico de sedimentación o Scree test (Cattell, 1966). Este criterio también se usa con AFE, aunque en realidad lo que se analiza como en el caso de la regla de Kaiser suele ser la

matriz de correlaciones original, y no la reducida (Ford et al., 1986).

Los otros criterios clásicos que se han usado en menor medida en combinación con la regla de Kaiser son: 1) la solución que ofrece la mejor interpretación posible, y 2) el número de factores basado a priori en la teoría (Conway y Huffcutt, 2003; Fabrigar et al., 1999; Ford et al., 1986).

#### *Recomendación actual:*

Hoy en día, la regla de Kaiser es la más desaconsejada de todas las opciones posibles, aunque a pesar de ello sigue siendo la más utilizada (Costello y Osborne, 2005). Por ejemplo, Lorenzo-Seva, Timmerman y Kiers (2011) directamente excluyeron este criterio del estudio de simulación que llevaron a cabo por ser claramente inadecuado. Además, Ferrando y Anguiano-Carrasco (2010) afirmaron que uno de los inconvenientes más importantes de este procedimiento reside en que el número de factores que identifica está en relación directa con el número de ítems que se analicen; si se analizan  $n$  variables, el número de factores que obtendrá esta regla oscilará entre  $n/5$  y  $n/3$ , sea la escala unidimensional o no. Igualmente, Lorenzo-Seva et al. (2011) excluyeron el Scree test, en este caso porque la subjetividad que entraña resulta difícilmente programable, por lo que no se puede introducir en un estudio de simulación como el que ellos llevaron a cabo. Fabrigar et al. (1999) apoyaron su buen funcionamiento cuando los factores comunes están claramente definidos.

En la actualidad se recomienda con insistencia que la decisión acerca del número de factores se tome tras considerar: 1) varios criterios objetivos, y siempre teniendo en cuenta 2) la interpretabilidad de la solución encontrada y 3) la teoría de partida (Lorenzo-Seva et al., 2011).

Los criterios objetivos disponibles tras el análisis varían mucho dependiendo en gran medida del programa que se use, y en menor medida del método de estimación de factores.

El Análisis Paralelo (AP) selecciona los componentes o factores comunes que presentan valores propios mayores que los que se obtendrían por azar (Horn, 1965). La idea es: si la matriz analizada procediera de una población en que los ítems no estuvieran relacionados, ¿qué valores propios podrían alcanzar los factores comunes extraídos de esa matriz? Y ¿cuántos factores comunes de los que hemos obtenido sobre la matriz analizada superan esos valores propios “espurios”? La respuesta a esta pregunta ofrece el número correcto de factores comunes. Esta técnica fue desarrollada para ser utilizada sobre la matriz de correlaciones original (como la regla de Kaiser o el Scree test) y no sobre la matriz de correlaciones reducida. No obstante, se ha recomendado su uso también para identificar el número de factores comunes (Hayton et al., 2004; Lorenzo-Seva et al. 2013; Peres-Neto, Jackson y Sommers, 2005; Velicer et al., 2000).

En cambio, a pesar de ser recomendado en numerosos estudios, este criterio se utiliza escasamente. Una vez más la

explicación hay que buscarla en la accesibilidad del procedimiento, ya que SPSS, el programa más utilizado por los investigadores en ciencias sociales, no lo implementa en su repertorio. Afortunadamente, ya hay una rutina en R que permite estimar AP y otros criterios para la identificación del número de factores en el contexto del programa SPSS (Basto y Pereira, 2012), programas para SPSS y SAS que determinan el número de componentes mediante AP y MAP de Velicer (O'Connor, 2000), y un programa específicamente destinado a AP en Internet: ViSta-PARAN (Ledesma y Valero-Mora, 2007). El programa de libre distribución FACTOR (Lorenzo-Seva y Ferrando, 2006) también implementa los últimos adelantos en este y otros aspectos del análisis factorial y además, es también de libre distribución.

Tres criterios actualmente recomendados parten de una lógica diferente al AP. Todos ellos se basan en la evaluación de las correlaciones residuales. La idea es que si se ha extraído el número de factores comunes adecuado, no debe quedar idealmente varianza común y por lo tanto las correlaciones residuales tenderán a cero. Estos criterios son: 1) la inspección directa de la distribución de los residuales tipificados (mediana y valores extremos, que indicarán entre qué variables está el problema); 2) la raíz media cuadrática residual (RMCR) y 3) el índice gamma o GFI (Tanaka y Huba, 1989). Los dos últimos surgieron en el contexto de un marco más general: los MEE.

El RMCR es un estadístico descriptivo que condensa la información contenida en la matriz de correlaciones residuales. Es una medida de la magnitud media de las correlaciones residuales. Habitualmente, se ha utilizado la indicación de Harman (1980) que propone como valor de referencia .05 para considerar un ajuste aceptable. Lorenzo-Seva y Ferrando (2006) recomendaron, en cambio, utilizar el criterio propuesto inicialmente por Kelley (1935); Kelley utilizó como valor de referencia el error típico de un coeficiente de correlación de cero en la población de que proceden los datos, y que es aproximadamente  $1/\sqrt{N}$  (ver Ferrando y Anguiano-Carrasco, 2010). El RMCR puede utilizarse con cualquier método de estimación de factores, aunque no está implementado en los programas generalistas como SPSS.

El GFI es una medida de bondad de ajuste normada que oscila entre 0 y 1 que también puede usarse con la mayor parte de los métodos de extracción de factores, aunque tampoco es habitual en programas generalistas como SPSS. Indica la proporción de covariación entre los ítems explicada por el modelo propuesto, es por tanto una especie de coeficiente de determinación multivariado. Valores superiores a 0.95 son indicadores de buen ajuste del modelo (Ruiz, Pardo y San Martín, 2010).

El RMSEA (Steiger y Lind, 1980) es un índice basado en el estadístico ji-cuadrado, por lo que solo puede obtenerse con aquellos procedimientos de estimación de factores que lo ofrecen, lo que depende del software que utilizemos (por ejemplo FACTOR lo incluye para ULS y MV, pero no está disponible en los programas generalistas como SPSS, aunque sí puede ser calculado por el usuario). Encontramos reco-

mendaciones de su uso en los trabajos de Browne y Cudeck (1993), Fabrigar et al. (1999), Ferrando y Anguiano-Carrasco (2010) y Lorenzo-Seva et al. (2011). Este índice estima el error de aproximación del modelo propuesto. Se trata de un índice relativo a los grados de libertad del modelo, por lo que puede favorecer la selección de modelos más complejos. Valores por debajo de .05 se consideran excelentes, en tanto que valores mayores que .08 indicarían ajuste insuficiente.

Por último, otro criterio aplicado en el contexto del ACP que se ha generalizado también al AFE, es el Minimum Average Partial Test (MAP) (Velicer, 1976). Este procedimiento se centra en identificar el número de componentes que ofrezca la correlación parcial mínima entre los residuales resultantes. Ziwck y Velicer (1986) indicaron que este procedimiento fue adecuado para identificar el número de componentes. También se ha aplicado a la identificación del número de factores comunes. Este criterio se encuentra en la rutina en R antes mencionada que permite mejorar las prestaciones del programa SPSS en el AFE (Basto y Pereira, 2012), en el programa desarrollado por O'Connor (2000), y también en el programa FACTOR (Lorenzo-Seva y Ferrando, 2006).

En cuanto a la interpretabilidad de la solución, diremos que es fundamental en la valoración de lo que sugieran los indicadores objetivos que hayamos considerado. De poco sirve que un modelo con 2 factores ajuste mejor que otro con 3, si el tercer factor está pobremente definido (con menos de 3 ítems con saturaciones superiores a .30, por ejemplo), o no puede ser interpretado por carecer de contenido. En este sentido hay que destacar la tendencia en las últimas décadas (Fabrigar et al. 1999, Lorenzo-Seva et al. 2013) hacia distinguir entre factores comunes mayores y menores. Los factores comunes mayores son los que interesa retener, porque son los que explican una parte realmente sustantiva de los ítems que forman la escala. Los factores comunes menores también explican una parte de la varianza común, pero es una parte pequeña, que no llega a ser interpretable en términos sustantivos en el contexto de lo que se desea medir con ese conjunto de ítems. El ejemplo de Lorenzo-Seva et al. (2013) es clarificador: si dos ítems que miden diferentes rasgos de personalidad tienen como denominador común la situación en que se sitúa a la persona, es posible que ese denominador común produzca cierta covarianza entre las respuestas de las personas, y aparezca un factor común menor que explique esa covarianza. Aunque claro está, ese factor común no tendrá significado sustantivo, sino circunstancial. En consecuencia, la recomendación actual ya no es explicar la mayor parte de varianza común posible, sino la mayor parte de la varianza común posible de explicar con el número adecuado de factores comunes, que serán aquellos factores que tengan sentido. Se trata, como el lector habrá ya adivinado, del eterno dilema entre parsimonia y plausibilidad.

Un criterio que no se recomienda es la interpretación de la varianza explicada, porque no es un indicador satisfactorio de la adecuación del número de factores comunes identificados, como pudiera parecer a simple vista. La varianza común

a los ítems no se puede repartir entre los diferentes factores comunes (excepto bajo un método de estimación: el Análisis Factorial de Rango Mínimo). Si lo hiciéramos algunos de los valores propios serían negativos, porque de otro modo es muy probable que la suma de la varianza común explicada por cada factor sobrepasase la varianza común total, precisamente por eso, porque es común, y aparece “contada” en la parte común de la varianza en diferentes ítems. Puede consultarse una explicación más detallada en Lorenzo-Seva, U. (2013). También puede consultarse Bandalos y Finney (2010).

En resumen, la recomendación es utilizar una combinación de criterios objetivos y de criterios basados en la teoría y la interpretabilidad de la solución, y usar esta combinación para comparar las diferentes soluciones alternativas. La solución óptima será la que presente la combinación más plausible y a la vez más parsimoniosa. Hay que tener en cuenta que el programa que se utilice para el análisis limitará en gran medida, junto con el método de estimación de los factores comunes, el tipo concreto de criterios objetivos que podamos usar. La recomendación es abandonar aquellos programas que no permitan usar los mejores métodos, a menos que sepamos que hay una macro que lo remedie. Se desaconseja el criterio más utilizado en todos los estudios revisados: la regla de Kaiser, y recomendamos incorporar los procedimientos basados en: 1) el AP, 2) el método MAP o 3) el indicador de ajuste RMSEA, si es posible. No se pueden hacer recomendaciones de carácter universal más allá. El lector interesado puede encontrarse información más detallada sobre estos y otros métodos y sobre su funcionamiento en un estudio de simulación en Lorenzo-Seva et al. (2013).

## ¿Qué tipo de rotación de factores y qué criterio de asignación de ítems a los factores debemos emplear?

Después de la fase de estimación de factores, la solución es rotada para conseguir la mayor simplicidad e interpretabilidad.

### *Recomendación clásica:*

Thurstone (1947) sugirió que los factores fueran rotados en un espacio multidimensional para conseguir la solución con la mejor estructura simple. La rotación factorial puede ser ortogonal u oblicua. El método de rotación ortogonal asume la independencia de los factores; mientras que el método de rotación oblicua permite la correlación entre factores.

El uso de un tipo u otro de rotación tiene implicaciones prácticas a la hora de ofrecer los resultados de un AFE. La contribución de un ítem particular a un factor determinado se indica tanto con los coeficientes patrón (*factor pattern coefficients*) como con los coeficientes estructura (*factor structure coefficients*).

Los primeros se encuentran en la matriz de configuración, y los segundos en la matriz de estructura. En el AFE, la matriz de estructura ofrece las correlaciones entre las variables observadas (ítems) y las variables latentes (factores). Cuando se utiliza rotación ortogonal, como se asume que los factores no están correlacionados, la matriz de estructura y la matriz de configuración son exactamente iguales, por lo que será suficiente con aportar la matriz de coeficientes patrón. Sin embargo, cuando se utiliza rotación oblicua, la matriz de correlaciones entre factores no es una matriz identidad, de tal suerte que la matriz de estructura y la matriz de configuración no ofrecerán los mismos coeficientes. En rotación oblicua, se deberá aportar ambas matrices, y la interpretación de los resultados se deberá hacer en primer lugar teniendo en cuenta los coeficientes de estructura, y posteriormente se examinarán los coeficientes patrón (Courville y Thompson, 2001; Gorsuch, 1983; Henson y Roberts, 2006; Thompson y Borrello, 1985).

Los criterios de rotación más ampliamente conocidos, y que se encuentran disponibles en la mayor parte de los programas comerciales de análisis estadístico, son: Varimax (Kaiser, 1958) en rotación ortogonal, y Oblimin directo (Clarkson y Jennrich, 1988) y Promax (Hendrickson y White, 1964) en rotación oblicua. Varimax ha sido propuesto como un criterio de rotación ortogonal cuando no hay un factor dominante, mientras que Quartimax ha sido propuesto como un criterio ortogonal alternativo cuando se espera un solo factor general (Carroll, 1953). Por su parte, Equamax combina ambos criterios (un solo factor/varios factores) y lleva a soluciones intermedias. También se han formulado versiones ponderadas del criterio Varimax (e.g., Cureton y Mulaik, 1975).

En rotación oblicua, el criterio más utilizado ha sido Oblimin directo. El criterio Quartimin (Jennrich y Sampson, 1966) es equivalente al criterio Quartimax, pero en rotación oblicua. Otros criterios han sido propuestos recientemente, muchos de ellos prácticamente desconocidos para los usuarios habituales de AFE, como, por ejemplo, Geomin (Yates, 1987), Promin (Lorenzo-Seva, 1999) y Oblimin ponderado (Lorenzo-Seva, 2000). Para una presentación detallada de diferentes criterios de rotación recomendamos los trabajos de Browne (2001), y Sass y Schmitt (2010).

Aunque tradicionalmente se han formulado los diferentes criterios de rotación bajo un método concreto (ortogonal u oblicuo), desarrollos de software más recientes ponen de manifiesto que un mismo criterio puede estar disponible utilizando tanto el método ortogonal como el método oblicuo (e.g. Mplus; Muthén y Muthén, 1998-2012).

Históricamente, los investigadores han seleccionado el criterio de rotación a utilizar en el AFE en función de la popularidad de un criterio determinado en ese momento particular (e.g., Varimax, parte integrante del “pack Little Jiffy” mencionado con anterioridad), o en función de la recomendación de utilizar rotación ortogonal cuando los factores son independientes, y rotación oblicua cuando los factores están relacionados.

### Recomendación actual:

En las últimas dos décadas, los estudios de revisión del uso del AFE han puesto de manifiesto una evolución donde se ha pasado de un uso mayoritario de la rotación ortogonal (concretamente el criterio Varimax), a utilizar cada vez más la rotación oblicua. Ford et al. (1986) concluyeron que, en los estudios revisados, se aplicó mayoritariamente la rotación ortogonal. Concretamente, Ford et al. (1986) encontraron que alrededor del 80% de los AFEs revisados utilizaron rotación ortogonal, mientras que aproximadamente el 12% o bien utilizaron rotación oblicua o no rotaron la solución factorial (el 8% restante no ofrecían información al respecto). Ya que en este estudio se incluye en la misma categoría los que usan rotación oblicua y los que no rotan, se puede concluir que el porcentaje de análisis en los que se usó rotación oblicua fue inferior al 12%.

Fabrigar et al. (1999) llegaron a una conclusión similar a la de Ford, et al. (1986), pero se observó una tendencia a seguir la recomendación de utilizar rotación oblicua. Los resultados indicaron que el 48.3% de los AFEs evaluados utilizaron rotación ortogonal (concretamente Varimax), y el 20.6% utilizaron rotación oblicua.

Recientemente, Conway y Huffcutt (2003), volvieron a confirmar la tendencia decreciente de la rotación ortogonal. El porcentaje de AFEs analizados que utilizaron rotación ortogonal había bajado al 41.2%. Pero no pudieron confirmar el incremento en el porcentaje de utilización de la rotación oblicua. En su estudio encontraron que se había utilizado solo en el 18% de los casos. Contrariamente a lo esperado este resultado indicaba un ligero descenso de la rotación oblicua. Conway y Huffcutt (2003) explicaban este resultado aludiendo al porcentaje de trabajos que no indicaban la rotación empleada. En efecto, en su estudio este porcentaje fue del 18%, mientras que en el estudio de Fabrigar et al. (1999) ese porcentaje fue solo del 8%.

Esta evolución en el uso de criterios de rotación se debe a que tradicionalmente se pensaba que la rotación ortogonal producía estructuras más simples y más fácilmente interpretables (e.g., Nunnally, 1978). Sin embargo, diferentes estudios (Fabrigar et al., 1999; Finch, 2006; Henson y Roberts, 2006; Matsunaga, 2010; Park et al., 2002; Preacher y MacCallum, 2003) pusieron de manifiesto que esta afirmación no era cierta, sino que la rotación oblicua es capaz de presentar estructuras más claras, simples e interpretables.

Estos estudios aconsejan utilizar rotación oblicua independientemente del modelo teórico del que se parta (factores independientes o relacionados). Los argumentos para ello son varios: 1) Casi todos los fenómenos que se estudian en las ciencias sociales y de la salud están más o menos interrelacionados entre sí, por lo que encontrar relaciones de ortogonalidad perfecta es difícil. De ello se deduce, que imponer una solución factorial ortogonal puede ser muy probablemente alejarse de la realidad; 2) Si el constructo bajo estudio realmente presenta una estructura de factores independientes, esta ortogonalidad se verá reflejada en los resultados (al

permitir correlaciones entre factores utilizando una aproximación oblicua, las correlaciones obtenidas serán bajas); 3) Finalmente, si las correlaciones entre factores fueran consistentemente bajas (por debajo de .30 ó .20), se propone repetir el análisis utilizando una solución ortogonal. Y en el caso de que las dos soluciones fueran similares, atendiendo al criterio de parsimonia, se aconseja aceptar provisionalmente la solución ortogonal (Ferrando y Anguiano-Carrasco, 2010).

Hasta aquí queda claro que la recomendación es comenzar aplicando criterios de rotación oblicua. Pero de todos los que se han formulados, ¿cuál es el mejor? La realidad es que esta pregunta parece no tener respuesta. Recientemente, se ha indicado que la casuística de estructuras factoriales hipotetizadas es muy amplia, y que por lo tanto los criterios de rotación a aplicar en cada caso particular también serán diferentes (e.g., Browne, 2001). En esta línea, el trabajo de Sass y Schmitt (2010) pretendieron comparar diferentes criterios de rotación en situaciones diferentes (e.g., estructura simple perfecta, estructura simple aproximada, estructura compleja, estructura general) y proporcionar unas directrices para realizar un AFE. Para ello, estos autores realizaron un estudio de simulación con Mplus donde se evaluaron diferentes criterios de rotación oblicua (Quartimin, CF-Equamax, CF-Facparsim y Geomin) y su capacidad para reproducir diferentes matrices de “patrón factorial” (*factor pattern matrices*). Sass y Schmitt (2010) concluyeron que no se puede dar una respuesta definitiva a la cuestión de qué criterio de rotación produce la “mejor” solución. Es decir, a priori, no hay criterios de rotaciones correctas o incorrectas, ni criterios de rotación que produzcan soluciones mejores o peores. Por el contrario, se debe ser consciente de que la elección de un criterio de rotación u otro, puede tener efectos importantes en los patrones factoriales estimados (*estimated factor pattern loadings*) y en las correlaciones entre factores (*interfactor correlations*). Los diferentes criterios de rotación, en determinadas circunstancias, pueden llevar a matrices de “patrón factorial” muy similares; pero bajo otras circunstancias pueden producir matrices de “patrón factorial” diferentes, e incluso contradictorias. Es en estas situaciones cuando el investigador debe tomar la difícil decisión de elegir el criterio de rotación más adecuado, es decir, el que proporciona la solución más simple y más informativa (Asparouhov y Muthén, 2009). En última instancia, “la selección del “mejor” criterio de rotación debe hacerla el investigador” (Sass y Schmitt, 2010, p. 99). Aunque esta reflexión ya fue corroborada por Browne (2001, p. 145): “*the choice of the best solution therefore cannot be made automatically and without human judgment*”. Es decir, el investigador puede poner a prueba distintas soluciones factoriales utilizando diferentes criterios de rotación, y en base a los resultados obtenidos, seleccionar como “mejor” criterio de rotación, aquel que ofrece la solución factorial más simple y más informativa. Incluso se aconseja, que además de proporcionar una justificación adecuada del criterio de rotación elegido, se ofrezca las diferentes matrices “factor patrón” resultado de aplicar diferentes métodos de rotación, de tal modo que “el lector pueda sacar sus propias conclusiones en

base a las estructuras factoriales alternativas” (Sass y Schmitt, 2010, p. 97).

En cuanto al criterio de asignación de los ítems a los factores -otro aspecto que puede hacer variar enormemente la interpretación de la solución obtenida- la práctica más común es retener saturaciones que estén por encima de .30 ó .40 (Bandalos y Finney, 2010; Guadagnoli y Velicer, 1988; McDonald, 1985). De hecho, Tabachnick y Fidell (2001) afirmaron que .32 podría ser una buena regla general en la saturación mínima a considerar, que equivale aproximadamente al 10% de la varianza explicada. Otros autores son un poco más estrictos y sitúan el punto de corte en .40 como mínimo (MacCallum et al., 1999; Velicer y Fava, 1998; Williams, Brown y Onsmán, 2010) y además recomiendan elevar este punto de corte a medida que la muestra sea inferior a 300 casos. También se recomienda que la discrepancia entre las saturaciones en los dos primeros factores sea .50/.20 ó .60/.20 (es decir, una diferencia de .30-.40). Algunos estudios, especialmente los empíricos, utilizan la discrepancia .60/.40 (Henson y Roberts, 2006; Park et al., 2002).

Los ítems que no superen el criterio o conjunto de criterios establecidos, deben ser revisados en sus dos aspectos: sustantivo y metodológico, para identificar en lo posible el origen de su mal funcionamiento. Con ello se puede valorar si es necesario o aconsejable que se eliminen del test, que se modifiquen de algún modo para incluirlos en una nueva versión del test, o si lo que sucede es que hay que añadir nuevos ítems de contenido semejante para muestrear adecuadamente el contenido del factor que pretendemos medir con ese ítem (Costello y Osborne, 2005), lo que nos lleva a reexaminar la validez de contenido que mencionamos con anterioridad (Bandalos y Finney, 2010). Independientemente de ello, se realizará un nuevo análisis factorial con la escala reducida tras eliminar esos ítems. Idealmente el análisis se repetirá tras eliminar uno de los ítems inadecuados cada vez (Bandalos y Finney, 2010). En ocasiones variaciones pequeñas como eliminar un par de ítems pueden modificar sustancialmente el resultado final del análisis.

Como el lector ya habrá notado, los aspectos tratados hasta aquí son muchos y variados, y aun así no agotan todos los que debieran tratarse en el contexto del AFE. Solo hemos tratado de presentar los más relevantes que el investigador aplicado debe resolver cuando se pone manos a la obra. Hemos dejado fuera, entre otros aspectos, el AF no lineal, sobre el que existe extensa literatura, el asunto de las puntuaciones factoriales, que puede consultarse en el capítulo sobre Análisis factorial Exploratorio de Abad, et al., (2011), tampoco hemos ofrecido recomendaciones acerca de la utilización de paquetes o miniescalas, que pueden consultarse en West et al., (1995), Gorsuch (1997) y en un reciente y extenso trabajo de Little, Rhemtulla, Gibson y Schoemann, (2013), y finalmente, tampoco hemos dado recomendaciones acerca de cuándo y cómo usar el denominado originalmente “Extension Analysis” (Gorsuch, 1997 a, b), para el que existe desde hace ya una década el programa EXTENSION, compatible con SPSS, SAS y MATLAB (O’Connor, 2001).

Remitimos a los lectores interesados a las fuentes citadas para consultar información sobre estos aspectos. A continuación resumiremos todo lo anterior de forma que sea útil al lector.

## Análisis Factorial Exploratorio de Ítems: Breve Guía de uso

Hemos presentado mucha información hasta aquí, y puede que en este punto el objetivo de ofrecer unas recomendaciones claras y actuales se haya desdibujado. Así que como conclusión se resumirán e integrarán en una guía breve las recomendaciones actuales más importantes en torno a cuatro aspectos centrales: 1) el tipo de datos y la matriz de asociación; 2) el método de estimación de factores; 3) el número de factores y finalmente, 4) el método de rotación y asignación de ítems.

### El tipo de datos y la matriz de asociación

#### Recomendación

El investigador aplicado que pretenda realizar un AFE para evaluar la escala o escalas de su interés debe utilizar tamaños de muestra apropiados, ser consecuente con la naturaleza ordinal, politómica (tipo Likert) o dicotómica, de los ítems que quiera analizar usando software apropiado en cada caso, y comprobar la adecuación de sus datos al AF al menos mediante la prueba KMO.

Como norma general, la matriz de correlaciones de Pearson queda limitada al caso en que los ítems sean continuos, o si no lo son, tengan cinco o más alternativas de respuesta y distribuciones aproximadamente normales (un criterio exigente emplea valores para la curtosis y la asimetría en el rango (-1,1)). En la mayor parte de los casos restantes lo aconsejable es usar la matriz de correlaciones policórica o tetracórica, según el caso.

La excepción ocurre en el caso de que la muestra sea pequeña (200 sujetos o menos). En este caso, las correlaciones policóricas pueden resultar poco estables y por lo tanto ser menos recomendables que las propias correlaciones de Pearson, por lo que en ese caso recomendamos relajar esta exigencia y usar valores para asimetría y curtosis mayores (-2,2). El caso de duda, el investigador puede comparar las soluciones obtenidas con ambas matrices y decidir de forma informada cual es la mejor solución.

El tamaño de la muestra no es fácil de anticipar, porque dependerá de las características psicométricas de los ítems, del tipo de matriz de asociación, y además estos dos aspectos están relacionados entre sí. Pero si se desea tener cierta garantía de poder usar la matriz de correlaciones policórica/tetracórica se recomienda no emplear muestras por debajo de los 300 sujetos. En el caso de la matriz de correlaciones de Pearson se recomienda una muestra mínima de 200 casos. Estas recomendaciones tienen solo validez como pun-

to de partida. Situaciones en que las saturaciones sean bajas (menor de .40) y el número de ítems por factor también sea bajo (3 ítems) requieren tamaños muestrales mayores para tener alguna garantía de generalización de los resultados. Como es habitual utilizar muestras de conveniencia hay que tener en cuenta dos problemas: la no-representatividad y la atenuación por restricción de rango.

La adecuación de los datos al AF se considera “suficiente” cuando el resultado de la medida KMO tome valores entre .70-.79, y “satisfactoria” cuando tome valores mayores a .80.

### El método de estimación de factores

#### Recomendación

La recomendación para el investigador informado es de nuevo que adapte el método al tipo de datos que desea analizar: si los ítems son ordinales pero tienen distribuciones aproximadamente normales, -en este caso el criterio es aceptar valores absolutos menores que 2 para la asimetría y la curtosis-, entonces el método adecuado es MV aplicado a la matriz de correlaciones de Pearson, porque es el que ofrece más información (medidas de ajuste, pruebas de significación y errores típicos alrededor de los parámetros estimados).

En caso contrario, lo que se recomienda es aplicar un método basado en MCO como ULS. El método ULS permite factorizar matrices en situaciones adversas, incluso con pocos casos y muchos ítems, y sin necesidad de hacer supuestos distribucionales. Este método también se recomienda cuando la solución de MV resulte inadecuada, porque presente problemas de convergencia o casos Heywood, u otras anomalías similares. En estos casos, hay que valorar cuidadosamente si esos problemas en la estimación no están enmascarando unos datos que en realidad no cumplen los supuestos requeridos (como el de normalidad o linealidad), o un modelo mal especificado (Fabrigar et al., 1999).

El método de estimación que se desaconseja enérgica y explícitamente es el CP. CP es un método para explicar la varianza de cada ítem individual. No es un método adecuado al objetivo del AF, que es explicar las correlaciones entre los ítems a partir de la identificación de un conjunto de factores comunes. Ni siquiera usado junto a una rotación oblicua de los componentes identificados.

### Numero de factores a retener

#### Recomendación

Respecto al número de factores también hay una opción que se desaconseja con la misma energía que el método de estimación CP, y es el criterio de Kaiser. Un investigador informado debe emplear una variedad de métodos para decidir: debe retener solo factores comunes “mayores” identificados en función del número de ítems (mínimo 3-4), tama-

ño de la saturación (mínimo .40) y significado de los ítems que los definen, y debe aplicar al menos un criterio objetivo “extra” al margen de los que ofrece un AFE convencional -que son scree test, GFI o RSMR-, o un AFE mediante MEE -hemos mencionado como criterio RMSEA, aunque existen otros criterios que se pueden utilizar-. Cuando la proporción de ítems por factor es pequeña, y el tamaño de la muestra es de 200 casos o más, este criterio objetivo “extra” puede ser un Análisis Paralelo o el método MAP. Teniendo en cuenta toda esta información, la recomendación es:

- 1) obtener el resultado del ajuste del modelo con el número de factores fijado al número esperado,
- 2) obtener el ajuste de otros modelos que puedan parecer plausibles tras el análisis inicial, y
- 3) comparar y decidir.

Se desaconseja explícitamente el uso del criterio de porcentaje de varianza explicada, por ser confuso.

### El método de rotación y asignación de ítems

#### Recomendación

Se recomienda absolutamente la rotación oblicua. Incluso en el caso en que no sea adecuada, porque el resultado obtenido lo pondrá de manifiesto. No ocurre lo mismo con las rotaciones ortogonales. Lo que no está claro es el método de rotación oblicua más adecuado, pues no hay criterios claros para ello. Es el investigador el que deberá probar varios y elegir de nuevo el que presente mejor interpretabilidad.

En cuanto al criterio para interpretar la saturación de un ítem, la recomendación también es clara: nunca por debajo de .40, especialmente si la muestra es menor de 300 casos.

Los ítems que no superen ese valor deberán ser eliminados del análisis y sometidos a un examen desde el punto de vista sustantivo y metodológico para decidir entre tres opciones: eliminar el ítem, revisar el ítem o añadir nuevos ítems que muestreen la faceta relacionada con ese ítem en particular. En cualquier caso habrá que a re-examinar la adecuación del contenido del test al constructo que se pretende medir. Independientemente de ello, se realizará un nuevo análisis factorial con la escala reducida tras eliminar uno de esos ítems cada vez.

Destacaremos una vez más que se desaconseja el uso del pack denominado “Little Jiffy”: ACP más Kaiser más Varimax. No cabe duda que durante cierto tiempo ha sido la forma más efectiva de resumir las relaciones observadas entre los ítems de un test, y que bajo ciertos -improbables- supuestos, podría dar resultados adecuados, pero como ha quedado claro -o eso esperamos- hoy día ya no es así.

### Conclusión

Martínez y Martínez (2009, p. 373) concluyen un muy interesante artículo sobre las nuevas conceptualizaciones de la va-

lidez -y la dificultad con que se abren camino en la comunidad psicométrica-, presentando el siguiente ejemplo: “Jacob Cohen, uno de los más influyentes investigadores en metodología, se ha pasado más de cuarenta años defendiendo el uso de los índices del tamaño del efecto en lugar del clásico “p-valor”, con argumentos que son de una lógica aplastante.” Y como el propio Cohen (1990) explica, a pesar de todo “El problema es que, en la práctica, la investigación actual apenas presta atención al tamaño del efecto” (Cohen, 1990, p. 1310). Eso es también lo que concluyen Sedlmeier y Gigerzner (1989) en su artículo “¿Tienen los estudios sobre la potencia estadística efecto sobre la potencia de los estudios?” que la potencia de los estudios publicados es, como promedio, un insuficiente .44. Puede que el uso de la potencia estadística haya aumentado en estos últimos años, pero no es la práctica habitual. Si algo tan simple como el estudio del tamaño del efecto en la interpretación de los análisis que realizamos habitualmente sigue sin ser aplicado de forma rutinaria, a pesar de la gran cantidad de estudios que muestran

su conveniencia, no nos deberíamos extrañar de lo que pasa en el contexto más complejo del AFE.

Borsboom, en su artículo “El ataque de los psicómetras” (2006), también concluye que los avances psicométricos parecen no haber contribuido apenas al avance en la psicología en los últimos 50 años. Y no es por falta de avances en psicometría, es por falta de aplicación de esos avances en Psicología. Pensamos que la clave está, como afirmó Cohen (1990, p. 1309), “en que el proceso por el que uno concibe, planifica, ejecuta y escribe una investigación debe depender del *juicio informado de uno mismo como científico*” (la cursiva es nuestra).

No se trata de repetir lo que han hecho otros investigadores antes, así no se puede avanzar. Se trata de informarse y tener juicio para tomar las mejores decisiones. Así es como se puede salir del laberinto por la puerta principal, como decíamos al principio de este artículo. Y esa debería ser la máxima de todos los que trabajamos en el mundo de la investigación: investigadores, editores, revisores y profesores. Esta es nuestra pequeña contribución.

## Referencias

- Abad, F. J., Olea, J., Ponsoda, J., y García, C. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Anstey, E. (1959). *Test de Dominós*. Buenos Aires: Paidós.
- Anderson, J. C., y Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411-423.
- Arrindell, W. A. y van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9, 165-178.
- Asparouhov, T. y Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397-438.
- Bandalos, D. L. y Finney, S. J. (2010). Factor Analysis: Exploratory and Confirmatory. En G. R. Hancock y R. O. Mueller (Eds.), *Reviewer's guide to quantitative methods*. Routledge: New York.
- Barrett, P. T. y Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study in Group Behavior*, 1, 23-33.
- Basto, M. y Pereira, J.M. (2012). An SPSS R-Menu for ordinal factor analysis. *Journal of Statistical Software*, 46, 1-29. Recuperado de: <http://www.jstatsoft.org/v46/i04/paper>.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical Considerations for Using Exploratory Factor Analysis in Educational Research. *Practical Assessment, Research & Evaluation*, 18(6). Recuperado de: <http://pareonline.net/getvn.asp?v=18&n=6>.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Borsboom, D., Mellenbergh, G. J. y Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203-219.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71, 425-440.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111-150.
- Browne, M. W. y Cudeck, R. (1993). Alternative ways of assessing model fit. En K. A. Carroll, J. B. (1953). An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 18, 23-28.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Clarkson, D. B. y Jennrich, R. I. (1988) Quartic rotation criteria algorithms. *Psychometrika*, 53, 251-259.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304-1312.
- Comrey, A. L. y Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.
- Conway, J. M. y Huffcutt, A. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147-168.
- Costello, A. B. y Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7), 1-9. Recuperado de: <http://pareonline.net/getvn.asp?v=10&n=7>.
- Courville, T. y Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: Beta is not enough. *Educational and Psychological Measurement*, 61, 229-248.
- Cureton, E. E. y Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika*, 40, 183-195.
- Curran, P. J., West, S. G. y Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.
- Edwards, J. R. (2011). The Fallacy of formative measurement. *Organizational Research Methods*, 14, 370-388.
- Embretson, S. E. y Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., y Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299.
- Ferrando, P. J. y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicólogo*, 31(1), 18-33.
- Ferrando, P. J. y Lorenzo, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions. *Psicologica*, 21, 301-323.
- Ferrando, P. J. y Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory*. Recuperado de <http://psico.fcep.urv.es/utilitats/factor/>.
- Ferrando, P. J., Varea, M. D. y Lorenzo, U. (1999). Evaluación psicométrica del cuestionario de ansiedad y rendimiento (CAR) en una muestra de escolares. *Psicothema*, 11 (1), 225-236.
- Finch, H. (2006). Comparison of the performance of varimax and promax rotations: Factor structure recovery for dichotomous items. *Journal of Educational Measurement*, 43(1), 39-52.
- Finney, S. J. y DiStefano, C. (2006). Nonnormal and categorical data in structural equation models. In G.R. Hancock y R.O. Mueller (Eds.), 1

- second course in structural equation modeling (pp. 269 - 314). Greenwich, CT: Information Age.
- Flora, D. V. y Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 466-491.
- Flora, D. B., LaBrish, C. y Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Quantitative Psychology and Measurement*, 3 (55), 1-21.
- Ford, J. K., MacCallum, R. C. y Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39, 291-314.
- Forero, C. G., Maydeu-Olivares, A. y Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16, 625-641.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd. ed.). Hillsdale, NJ: Erlbaum.
- Gorsuch, R. L. (1997a). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68, 532-560.
- Gorsuch, R. L. (1997b). New procedure for extension analysis in exploratory factor analysis. *Educational and Psychological Measurement*, 57, 725-740.
- Guadagnoli, E. y W. F. Velicer (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265-275.
- Hair, J. F., Anderson, R. E., Tatham, R. L. y Black, W. C. (2005). *Multivariate data analysis*. New York: Prentice All International.
- Hancock, G. R. y Mueller, O. (2010). *The Reviewer's Guide to Quantitative Methods in the Social Sciences*. (3rd ed.). New York, NY: Routledge.
- Harman, H. H. (1980). *Análisis factorial moderno*. Madrid: Saltés.
- Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago, IL: University of Chicago Press.
- Harman, H. H. y Jones, W. H. (1966). Factor analysis by minimizing residuals (Minres). *Psychometrika*, 31, 351-369.
- Hayton, J. C., Allen, D. G. y Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7, 191-205. Recuperado de: <http://orm.sagepub.com/content/7/2/191>.
- Hendrickson, A. E. y White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 17, 65-70.
- Henson, R. K. y Roberts, J. K. (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.
- Hinkin, T. R. (1995). A Review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967-988.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M. y Munford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, 65, 202-226.
- Holgado-Tello, F. C., Chacon-Moscoso, S., Barbero, I. y Vila-Abad, E. (2010). Polycoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, 44 (1), 153-166.
- Horn, J. L. (1965). A rationale and test for the number of factors in a factor analysis. *Psychometrika*, 30, 179-185.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441, 498-520.
- Hu, L. T. y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.
- Izquierdo, I., Olea, J. y Abad, F. J. (2013, Septiembre). *¿Se aplica correctamente el análisis factorial exploratorio en la investigación psicológica?* Comunicación presentada al XIII Congreso de Metodología de las Ciencias Sociales y de la Salud. Tenerife. España. Resumen recuperado de <http://13congresometodologia.es/programa/programa-cientifico/programa-cientifico-general/>
- Jennrich, R. I. y Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, 31, 313-323.
- Jolliffe, I. T. (2002). *Principal component analysis*. Second edition, New York: Springer-Verlag.
- Jöreskog, K. G. y Sörbom, D. (2007). *LISREL 8.80*. [Computer Software]. Lincolnwood, IL: Scientific Software International, Inc.
- Jöreskog, K. G. (1977). Factor analysis by least-squares and maximum-likelihood methods. En K. Enslein, A. Ralston y H.S. Wilf (Eds.), *Statistical methods for digital computers*, Vol. 3. New York: Wiley
- Jung, S. (2013). Exploratory factor analysis with small sample sizes: A comparison of three approaches. *Behavioural Processes*, 97, 90-95
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika*, 35, 401-415.
- Kelley, T. L. (1935). *Essential traits of mental life*. Harvard Studies in Education, vol. 26. Harvard University Press, Cambridge.
- Kim, J. O. y Mueller, C. W. (1978). *Factor analysis*. Beverly Hills, CA: Sage
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd Edition ed.). New York: The Guilford Press.
- Lawley, D. N. y Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworths.
- Ledesma, R. D., y Valero-Mora, P. (2007). Determining the Number of Factors to Retain in EFA: An easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment Research & Evaluation*, 12(2). Recuperado de: <http://pareonline.net/getvn.asp?v=12&n=2>
- Lee, C.-T., Zhang, G., y Edwards, M.C. (2012). Ordinary least squares estimation of parameters in exploratory factor analysis with ordinal data. *Multivariate Behavioral Research*, 47, 314-339.
- Lorenzo-Seva, U. (1999). Promin: A Method for Oblique Factor Rotation. *Multivariate Behavioral Research*, 34, 347-365.
- Lorenzo-Seva, U. (2000). The weighted oblimin rotation. *Psychometrika*, 65, 301-318.
- Lorenzo-Seva, U. y Ferrando, P. J. (2006). FACTOR: a computer program to fit the exploratory Factor Analysis model. *Behavioral Research Methods, Instruments & Computers*, 38(1), 88-91.
- Lorenzo-Seva, U. (2013). *How to report the percentage of explained common variance in exploratory factor analysis*. Recuperado de: <http://psico.fcep.urv.es/utilitats/factor/>
- Lorenzo-Seva, U. y Ferrando, P. J. (2012a). TETRA-COM: a comprehensive SPSS program for estimating the tetrachoric correlation. *Behavioral Research Methods*, 44(4), 1191-1196.
- Lorenzo-Seva, U. y Ferrando, P. (2012b). *Manual of the program FACTOR v. 8.10*. Recuperado de: <http://psico.fcep.urv.es/utilitats/factor/>
- Lorenzo-Seva, U. y Ferrando, P. J. (2013). FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement*, 37-6, 497-498. Recuperado de: <http://apm.sagepub.com>
- Lorenzo-Seva, U., Timmerman, M. E. y Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340-364.
- Marsh, H. W., Richards, G. E., Johnson, S., Roche, S. y Tremayne, P. (1994). Physical Self-Description Questionnaire: Psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport and Exercise Psychology*, 16, 270-305.
- Marsh, H. W., Morin, A. J. S., Parker, P. D. y Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*. doi/abs/10.1146/annurev-clinpsy-032813-153700
- Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S. y Nagengast, B. (2011). Measurement fruitfulness of exploratory structural equation modelling (ESEM) in addressing key substantive issues in motivation and engagement research. *Journal of Psychoeducational Assessment*, 29, 322-346. doi:10.1080/09243453.2013.812568
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U. y Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471-491. doi:10.1037/a0019227
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S. y Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, 16, 439-476. doi:10.1080/10705510903008220
- Marsh, H. W., Hau, K. T. y Wen, Z. (2004). In search of Golden Rules: Comment on hypothesis-testing approaches to setting cutoff values for

- fit indexes and dangers in overgeneralizing Hu and Bentler's findings. *Structural Equation Modeling*, 11(3), 320-341.
- Martínez, J. A. y Martínez, L. (2009). El análisis factorial confirmatorio y la validez de escalas en modelos causales. *Anales de Psicología*, 25(2), 368-374. Recuperado de: <http://www.redalyc.org/>
- Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, 3 (1), 97-110. Recuperado de: <http://www.redalyc.org/>
- MacCallum, R. C., Widaman, K. F., Preacher, K. y Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36, 611-637.
- MacCallum, R. C., Widaman, K. F., Zhang, S. y Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- McDonald, R. P. (1985). *Factor Analysis and Related Methods*. Hillsdale, NJ: LEA.
- McDonald R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- McDonald R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement*, 24, 99-114.
- McDonald, R. P. (2005). Semiconfirmatory factor analysis: The example of anxiety and depression. *Structural Equation Modeling*, 12, 163-172.
- McDonald, R. P. y Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82-99.
- Morin, A. J. S., Marsh, H. W., y Nagengast, B. (2013). Exploratory structural equation modeling. En G. R. Hancock, y R. O. Mueller, (Eds.), *Structural equation modeling: A second course* (pp. 395-440). Charlotte, NC: Information Age Publishing, Inc.
- Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw Hill.
- Mundfrom, D. J., Shaw, D. G. y Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159-168.
- Muthén, B., y Asparouhov, T. (2010). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313-335.
- Muthén, B. y Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189. Recuperado de: <http://www.statmodel.com/>
- Muthén, B. y Kaplan D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the mode. *British Journal of Mathematical and Statistical Psychology*, 45, 19-30. Recuperado de: <http://www.statmodel.com/>
- Muthén, L. K. y Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K. y Muthén, B. O. (1998-2012). *Mplus User's Guide*. (7th ed.) Los Angeles, CA: Muthén & Muthén
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32, 396-402.
- O'Connor, B. P. (2001). EXTENSION: SAS, SPSS, and MATLAB programs for extension analysis. *Applied Psychological Measurement*, 25, p. 88.
- Park, H. S., Dailey, R. y Lemus, D. (2002). The Use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28 (4), 562-577.
- Peres-Neto, P. R., Jackson, D. A. y Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, 49, 974-997. Recuperado de: <http://labs.eeb.utoronto.ca/>
- Pérez, E. y Medrano, L. (2010). Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. *Revista Argentina de Ciencias del Comportamiento*, 2(1), 58-66.
- Preacher, K. J. y MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13-32.
- Revelle, W. (2014). *Psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, Illinois, USA. Recuperado de: <http://CRAN.R-project.org/>
- Sass, D. A. y Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, 45, 73-103.
- Satorra, A., y Bentler, E.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. En A. von Eye y C.C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage
- Seldmeier, P. y Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309-316.
- Steiger, J. H. y Lind, J. M. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA. Recuperado de: <http://www.statpower.net/>
- Tabachnick, B. G. y Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.
- Tanaka, J. S. y Huba, G. H. (1989) A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 42, 233-239.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thompson, B. y Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement*, 45, 203-209.
- Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: The University Chicago Press.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Velicer, W. F. y Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3(2), 231-251.
- Velicer, W. F. y Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25(1), 1-28.
- Velicer, W.F., Eaton, C. A., y Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. En R. D. Goffin y E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy*. (pp. 41-71). Norwell, MA: Kluwer Academic. Recuperado de: <http://link.springer.com/>
- Vigil-Colet, A., Canals, J., Cosí, S., Lorenzo-Seva, U., Ferrando, P. J., Hernández-Martínez C., ... Domenech, E. (2009). The factorial structure of the 41-item version of the screen for child anxiety related emotional disorders (SCARED) in a spanish population of 8 to 12 years-old. *International Journal of Clinical and Health Psychology*, 9(2), 313-327. Recuperado de: <http://scholar.google.com/citations>
- West, S. G., Finch, J. F. y Curran, P. J. (1995). Structural equation models with nonnormal variables. Problems and remedies. En R. H. Hoyle (Ed.). *Structural equation modeling: Concepts, issues and applications*. (pp. 56-75). Newbury Park, CA: Sage.
- Williams, B., Brown, T., y Onsmann, A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3), 1-13.
- Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany: State University of New York Press.
- Yuan, K.H., y Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51, 289-309.
- Zwick, R.W. y Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

(Artículo recibido: 8-2-2014; revisión recibida: 27-5-2014; aceptado: 28-5-2014)