anales de psicología, 2014, vol. 30, nº 3 (octubre), 1170-1175
http://dx.doi.org/10.6018/analesps.30.3.199991

ISSN edición impresa: 0212-9728. ISSN edición web (http://revistas.um.es/analesps): 1695-2294

El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales

Pere J. Ferrando* y Urbano Lorenzo-Seva

CRAMC (Centro de Investigación para la Medida de la Conducta); Departamento de Psicología; Universitat Rovira i Virgili (Tarragona, España)

Resumen: El presente artículo puede considerarse como una ampliación al trabajo de Lloret et al. (2014) en la que se discuten, de forma ampliada, dos tópicos de especial relevancia en análisis factorial de ítems: (a) la decisión acerca de la matriz de correlación más apropiada en cada caso, y (b) la determinación de soluciones finales semi-confirmatorias, que sean realistas, interpretables y que utilicen la información disponible por el investigador. La presentación de los dos tópicos no es neutral, sino que refleja las posiciones de los autores, por lo que debe ser también evaluada críticamente por parte del lector. En ambos casos se ofrecen recomendaciones prácticas. Sin embargo, el trabajo va especialmente dirigido a los investigadores aplicados con cierta orientación metodológica que quieren ir un poco más allá de las recomendaciones actualizadas propuestas anteriormente.
Palabras clave: correlaciones producto-momento; correlaciones policóricas; modelo de respuesta graduada; estructura simple; rotaciones semi-especificadas; matriz diana; rotaciones ortogonales y oblicuas.

Title: Exploratory Item Factor Analysis: Some additional considerations.
Abstract: This article can be considered to be an extension of a previous work by Lloret et al. (2014) in which we discuss in more depth two particularly relevant topics in item factor analysis: (a) how to chose the most appropriate correlation matrix, and (b) how to arrive at a realistic and interpretable semi-confirmatory solution by making use of the available information. The discussion of these topics strongly reflects the views of the authors. So, we encourage the interested reader to take it critically. For both topics we offer practical recommendations. However, the article is mainly intended for applied researchers with a certain methodological penchant and who want to go beyond standard recommendations.
Key words: Product-moment correlations; Polychoric correlations; Graded Response Model; Simple Structure; Semi-Specified rotations; Target Matrix; Orthogonal and Oblique Rotations.

Pocas técnicas psicométricas han generado tanta controversia y tenido tantos altibajos como el análisis factorial exploratorio (AFE). Con respecto al primer punto, ha sido y sigue siendo con mucho, la técnica más usada en análisis de ítems, pero también la más criticada. Con respecto al segundo, el uso cada vez más generalizado de los modelos de ecuaciones estructurales lo arrinconó hace unos años al limbo de las técnicas casi obsoletas. Un AFE podía tener cierta justificación en las etapas iniciales de un estudio. Sin embargo, a partir de aquí, la técnica 'correcta' era el más riguroso análisis factorial confirmatorio (AFC). La realidad, sin embargo, es tozuda: la estructura de la mayor parte de los tests es inherentemente compleja y no casa bien con las restrictivas hipótesis del AFC. Parece pues que el AFE sigue siendo necesario para analizar las repuestas a los ítems.

Los psicómetras, por su parte, no han dejado de investigar en AFE y de hacerlo evolucionar. Sin embargo, debido a la falta de interés de unos y a la inercia de otros, estas evoluciones no se han visto reflejadas en el campo aplicado. Los autores, por tanto, aplaudimos sin reservas tanto la iniciativa del editor de Anales de Psicología como el trabajo de Lloret et al. (2014) que precede al nuestro. Se trata de una guía actual, correcta, clara y didáctica acerca de cómo debe proceder el investigador aplicado que utiliza el AFE en el análisis de ítems y tests.

En esta ampliación pretendemos discutir dos tópicos que consideramos de interés central y que, por falta de espacio no se han tratado más a fondo en el artículo precedente: el tópico de la matriz de correlación más apropiada en cada caso (o mejor, como se verá, del modelo AF más apropiado), y el de las soluciones intermedias entre el AFE y el AFC, tema que, creemos, cobrará cada vez más importancia en el campo aplicado. El tratamiento sigue estando dirigido al usuario del AF, pero se orienta sobre todo a aquellos investigadores aplicados con interés en la metodología que pretenden ir un paso más allá en el conocimiento de las evoluciones de la técnica y las líneas que cobrarán más interés en el futuro.

*** Dirección para correspondencia [Correspondence address]:**
Pere J. Ferrando, Departamento de Psicología; Universitat Rovira i Virgili; Carretera de Valls s/n; 43007 Tarragona (España).
E-mail: perejoan.ferrando@urv.cat

¿Correlaciones producto-momento o correlaciones policóricas?

El título del apartado refleja, de manera intencional, la forma en que se ha planteado tradicionalmente esta cuestión. Se trataría de decidir cuál es la matriz de asociación más adecuada para estimar el modelo lineal o común de AF. Sin embargo, al tomar esta decisión, se está decidiendo en realidad entre dos modelos factoriales distintos: Un modelo lineal y un modelo no lineal. Más específicamente, cuando se ajusta un AF basado en policóricas, lo que se ajusta en realidad es un modelo no lineal de la teoría de respuesta a los ítems: el modelo de respuesta graduada de Samejima (1969, véase Ferrando y Lorenzo-Seva, 2013).

En el caso unidimensional, que es el más claro desde un punto de vista ilustrativo, el modelo lineal (i.e., AF sobre Pearson) asume que la regresión de la puntuación del ítem sobre el factor común es lineal y con variancia constante. El modelo no lineal (i.e., AF sobre policóricas), asume, en cambio, que dicha regresión es una curva en forma de S (una ojiva) y que la variancia disminuye hacia los extremos. Dado que las respuestas a los ítems están limitadas (e.g., entre 0 y 1 ó entre 1 y 5) y el factor se concibe como una variable ilimitada, el modelo no lineal es más plausible y teóricamente más apropiado. El modelo lineal, en cambio, debe verse siempre como una aproximación.

El primer punto que vamos a discutir refiere a los determinantes que hacen que la aproximación lineal sea buena o no (asumiendo que el modelo no lineal es el correcto). Los primarios son dos: (a) la posición o dificultad de los ítems (o, de forma equivalente, el grado en que las distribuciones de los ítems son extremas) y (b) su capacidad discriminativa. Cuando los ítems son de dificultad media y discriminación moderada, las curvas de regresión tienen pendiente moderada y están centradas en la media de la distribución del factor. Por tanto, la regresión es esencialmente lineal para la mayor parte de los sujetos que responden y el modelo AF-Pearson es una buena aproximación. En cambio la aproximación será mala cuando los ítems sean a la vez extremos y muy discriminativos. El número de categorías de respuesta también influye, pero no es el determinante principal como suele creerse. Al aumentar el número de categorías, las distribuciones tienden generalmente a hacerse menos extremas y además se minimizan los problemas de atenuación que discutiremos después. Sin embargo, el AF lineal puede ser a veces una buena aproximación para ítems binarios y una mala aproximación para ítems continuos (Ferrando, 1994; McDonald y Ahlawat, 1974). Finalmente, desde un punto de vista práctico, aconsejamos evaluar los dos condicionantes primarios mediante la inspección de la medias y los coeficientes de asimetría de los ítems (posición) y la magnitud de las correlaciones inter-ítem (discriminación). Si las medias no son demasiado extremas, los coeficientes de asimetría no son mayores de 1 en valor absoluto, y las correlaciones inter-ítem se mueven, digamos en torno a .50 o menos, es de esperar que el modelo lineal funcione razonablemente bien. Estas recomendaciones concuerdan con los resultados obtenidos por Muthén y Kaplan (1985) utilizando simulación.

Si se ajusta el modelo lineal bajo condiciones en que no es apropiado, las principales consecuencias que cabe esperar son dos: (a) evidencia espuria de multidimensionalidad y (b) atenuación diferencial de los pesos factoriales (Ferrando y Lorenzo-Seva, 2013). Con respecto al primer punto, si se pretende ajustar una curva mediante un modelo lineal, harán falta términos o factores adicionales de curvatura para obtener un buen ajuste. Estos factores se conocen tradicionalmente como 'factores de dificultad' (Ferrando, 1994; McDonald y Ahlawat, 1974) y no tienen interpretación substantiva. Por el contrario, como el lector supondrá, su composición estará relacionada con la dificultad de los ítems y su capacidad discriminativa. En suma, si se pone a prueba un modelo con un adecuado número de factores substantivos, este modelo dará posiblemente un ajuste insuficiente, y será necesario estimar factores adicionales que no serán interpretables.

Con respecto al segundo punto, las estimaciones de los pesos proporcionadas por el modelo lineal estarán sesgadas hacia abajo (i.e. atenuadas) con respecto a los pesos verdaderos y, además, esté efecto no será constante: la atenuación será mayor cuanto más extremos y discriminativos los ítems (Muthén y Kaplan, 1985). Dado que el peso factorial es el principal indicador de la calidad del ítem como medida (Ferrando, 2009), este problema tiene consecuencias importantes ya que llevará, posiblemente, a interpretaciones incorrectas.

A estas alturas, el lector podría plantearse la siguiente cuestión: Aunque el AF lineal sea una buena aproximación bajo determinadas condiciones, ¿por qué no utilizar siempre el más correcto modelo no lineal? Pues, en términos generales, porque no es inusual que un modelo aproximado pero simple y robusto funcione mejor que un modelo teóricamente más apropiado pero más complejo e inestable, y, en el caso del AF, esto es bastante frecuente (Ferrando, 2009; Ferrando y Lorenzo-Seva, 2013; Hofstee, ten Berge y Hendricks, 1998). El AF no lineal es, desde luego, bastante más complejo que el lineal y presenta potenciales problemas que conviene conocer y que discutiremos a continuación.

El primer grupo de problemas refiere a la precisión y estabilidad de las correlaciones policóricas que sirven de base al AF no lineal. La correlación policórica no es un estadístico obtenido directamente sobre los datos, sino un estimador de una correlación latente entre supuestas variables continuas de respuesta que se estima de forma iterativa y bastante compleja (e.g., Rigdon, 2010). Este estimador puede no converger, llegar a resultados implausibles, o simplemente ser muy impreciso, con errores típicos mucho mayores que los de una correlación de Pearson. Entre otros condicionantes, estos problemas dependen del tamaño de la muestra y del número de categorías de respuesta. Una forma de verlo es teniendo en cuenta que la correlación policórica se calcula a partir de la tabla de contingencia entre las puntuaciones de dos ítems y que para obtener estimaciones estables conviene que las celdas contengan un número razonable de casos (Mislevy, 1986). Cuantas más categorías de respuesta, mayor será el número de celdas y por tanto, más muestra se necesitará potencialmente para llenarlas. Estamos de acuerdo con la recomendación de que con muestras menores de 200 casos no es aconsejable utilizar el modelo no lineal. Sin embargo, muestras mayores por si solas no garantizan la estabilidad de la matriz de base. Y, aunque es obvio, no está de más recordar que si las correlaciones que sirven de base no son estables, mucho menos lo serán los pesos estimados cuando se factorizen dichas correlaciones.

El origen del segundo grupo de problemas es que cada correlación policórica es un estimador máximo-verosímil independiente de una supuesta correlación entre dos variables latentes pero la matriz de correlaciones policóricas en su conjunto no es un estimador homogéneo de la supuesta matriz de correlaciones latentes. Aún en el caso de que todas las estimaciones sean plausibles, algunas tendrán más error que otras. El símil más apropiado podría ser una matriz de correlaciones de Pearson en la que cada correlación se hubiese obtenido en función de un tamaño muestral distinto: las correlaciones estimadas en muestras grandes serían posiblemente más correctas y más fiables que las obtenidas en muestras pequeñas. Esta situación de partida puede generar varios problemas. De entrada, la matriz policórica puede no ser positiva definida y entonces algunos procedimientos AF (máxima verosimilitud en particular) son simplemente inaplicables. Aunque sea aplicable, sin embargo, la estimación factorial mediante máxima verosimilitud, no será generalmente apropiada. Más específicamente, los indicadores de bondad de ajuste derivados del estadístico ji-cuadrado serán incorrectos y estarán generalmente inflados. Por esta razón, nuestra recomendación es estimar el AF de policóricas mediante ULS, y evaluar el ajuste mediante indicadores que no dependan directamente del ji-cuadrado: el índice GFI y la RMCR (e.g. McDonald, 1999).

Los estimadores factoriales ULS son consistentes (Mislevy, 1986), por lo que, en condiciones razonables, el AF de la matriz policórica se espera que llegue a estimaciones correctas, y esto es lo que generalmente encuentran los estudios de simulación. Si se generan datos desde un modelo no lineal, entonces la solución ULS basada en la matriz policórica recupera mejor los pesos factoriales verdaderos que la solución ULS lineal (e.g., Lee, Zhang y Edwards, 2012). Esto es de esperar ya que el primer procedimiento corrige el problema de la atenuación diferencial discutido antes. La evaluación del ajuste, sin embargo, es otro tema. Nuestra experiencia con datos reales indica que el ajuste proporcionado por el AF policórico es generalmente peor que el obtenido en función de las correlaciones de Pearson, incluso cuando las condiciones sugieren que debería ser mejor. Además, este resultado se acentúa cuanto mayor el número de ítems (véase también Rigdon y Ferguson, 1991). En suma, en condiciones en las que resulta claramente más apropiado a priori, el AF no lineal lleva generalmente a estimaciones más correctas de los pesos factoriales que el AF lineal. Sin embargo, la correcta evaluación del ajuste sigue siendo un problema. Este problema podría quizás mejorar implementando procedimientos de mínimos cuadrados ponderados, como los que ha propuesto Muthén (1993) para el caso confirmatorio, que tengan en cuenta los diferentes grados de error de las correlaciones policóricas.

En conclusión, el problema de decidir cual es el enfoque AF más apropiado para evaluar las respuestas a un test es complejo, y admitimos que nuestra exposición del tema puede haber creado confusión. Nos resultaría más fácil dar una serie de recomendaciones simplistas o categóricas, pero esto sería engañar al lector, y son precisamente estas recomendaciones estereotipadas las que acertadamente critica el artículo de Lloret et al. (2014). Como guía, proponemos que el investigador considere o evalúe los siguientes aspectos: (a) posición y distribución de las respuestas, (b) magnitud de las correlaciones inter-ítem, (c) número de categorías de repuesta y (d) tamaño de la muestra. En algunos casos este examen llevará a decisiones claras. Así, un cuestionario formado por ítems con formato Likert de 7 puntos que tienen dificultades medias y no son excesivamente discriminativos, y que se administra a una muestra de 250 sujetos justifica el uso del AF lineal basado en las correlaciones de Pearson. Este escenario es relativamente frecuente en medidas de personalidad o actitud que evalúan rasgos no patológicos (Ferrando, 2009). En el otro extremo, un test que contiene ítems muy fáciles junto a ítems muy difíciles, todos ellos muy discriminativos, con formato de respuesta en 3 puntos y administrado a una muestra de 600 sujetos sugiere claramente el uso de la opción no lineal. Este segundo escenario no es tan habitual, pero se da en algunos tests de capacidad y algunos cuestionarios clínicos (Reise y Waller, 2009). En el terreno intermedio donde se situarán la mayoría de casos, nuestra recomendación es llevar a cabo los dos tipos de análisis y evaluar ambas soluciones. En un programa como FACTOR (Lorenzo-Seva y Ferrando, 2013) basta un clic del ratón para pasar de una opción a otra, y la información adicional que se obtiene vale siempre la pena.

Para terminar este apartado, nos gustaría indicar algunas acciones que podrían mejorar notablemente la opción no lineal. En primer lugar, convendría implementar métodos de estimación más robustos y, a ser posible, no iterativos, para las correlaciones policóricas. Hasta el momento, el procedimiento que nos ha funcionado mejor es la estimación Bayesiana basada en el criterio MAP y con una distribución a priori moderadamente restrictiva. Este método garantiza la estimación de todas las correlaciones de la matriz (no hay problemas de convergencia) y lleva además a estimaciones plausibles en todas ellas incluso en muestras pequeñas. En segundo lugar, habría que proporcionar los errores típicos de las correlaciones policóricas a fin de que el investigador evaluara la estabilidad y precisión de las mismas antes de someterlas al AF. Finalmente, valdría la pena estudiar y eventualmente implementar métodos de estimación AF basados en mínimos cuadrados ponderados que quizás permitieran evaluar de forma más correcta el ajuste de la solución propuesta. Los autores estamos trabajando en estos puntos para implementarlos en un futuro en el programa FACTOR (Lorenzo-Seva y Ferrando, 2013).

Soluciones factoriales intermedias entre el AFE y el AFC

Como ya hemos apuntado, la aparición del AFC relegó el uso del AFE a una técnica de tipo *menor* de uso justificado sólo en estadios iniciales de una investigación. La yuxtaposición entre ambas aproximaciones era tal, que incluso se entendían como dos técnicas completamente diferentes. Hoy en día existe unanimidad en que, como Lloret et al. (2014) exponen, el AFE y el AFC son dos polos de un continuo. Cabe indicar que, pese a los nombres que reciben, el modo como habitualmente se utilizan una y otra técnica en investigación aplicada hace que ni la una sea totalmente *exploratoria*, ni la otra sea totalmente *confirmatoria*. La razón se expone a continuación.

La aplicación habitual del AFE implica calcular un procedimiento objetivo que permita decidir el número óptimo de factores a extraer, seguido de una rotación factorial que maximice el criterio de *estructura simple*. Esta aplicación hace que el AFE no sea una técnica puramente exploratoria: (1) el número de dimensiones se decide habitualmente utilizando procedimientos que favorecen los modelos con pocos factores (lo que en sí ya indica un tipo de modelo factorial que se favorece); y (2) el criterio de estructura simple es en si un modelo de solución factorial (un modelo que espera un patrón simple de saturaciones factoriales). Si bien es cierto que el criterio de estructura simple tal y como fue definido por Thurstone (1947, página 335) es imposible de ajustar, Kaiser (1974) lo redefinió como *simplicidad factorial*: un modelo que espera que cada variable muestre una única saturación marcadamente diferente de cero en un único factor. Dado que de forma implícita se está utilizando un modelo, la solución final no es puramente *exploratoria*.

En el polo opuesto, el AFC implica plantear de forma explícita un modelo poblacional que determina el número de factores, así como las saturaciones que cada variable ha de mostrar en cada factor. Hay que puntualizar que es habitual que los investigadores definan para cada variable cuáles son las saturaciones que han de ser cero en la población, dando libertad de magnitud a la saturación que se espera que sea diferente de cero. Ahora bien, en la aplicación práctica del AFC es muy habitual que una vez ajustado un modelo inicial utilizando los datos de una muestra concreta, los investigadores relajen *ad hoc* alguno de los parámetros del modelo (por ejemplo, permitiendo en el modelo que alguna de las variables muestre más de una saturación diferente de cero) para favorecer un mejor ajuste del modelo. En el momento en que se relajan de este modo los parámetros del modelo inicial, la solución final ya no es puramente *confirmatoria*.

Ahora bien, si el AFE y el AFC son dos polos de un continuo, la pregunta natural es: ¿cuáles son las opciones metodológicas que el investigador dispone entre un y otro polo? En el texto que continúa presentamos las opciones disponibles partiendo de las opciones más confirmatorias y desplazándonos hacia las opciones cada vez más exploratorias.

Como ya hemos apuntado, la estructura de la mayor parte de los tests psicológicos es compleja y difícilmente modelable en términos de hipótesis restrictivas propias del AFC. Quizás la hipótesis más restrictiva de los modelos factoriales es asumir que en la población cada ítem es un indicador puro de una única dimensión subyacente: es decir, proponer que cada ítem presentará en la población una única saturación factorial con valor absoluto 1, y todas las demás saturaciones serán exactamente cero. Una primera técnica disponible para relajar este modelo tan restrictivo es la *rotación procusteana* (en alusión al posadero de la mitología griega que cortaba o estiraba a sus huéspedes con la finalidad de que se ajustasen lo mejor posible a la cama de hierro donde les hacía dormir). La rotación procusteana ortogonal fue propuesta por Cliff (1977), mientras que la oblicua fue propuesta por ten Berge y Nevels (1977). La idea de esta rotación es proponer una matriz diana (o *target*) donde se indica con los valores 1 y 0 el patrón factorial esperado para cada ítem. Por ejemplo, supongamos un conjunto de 20 ítems de un test, de los que se hipotetiza que 10 ítems miden Extraversión y 10 ítems miden Responsabilidad. La matriz diana consistirá en una matriz de 2 columnas (una columna por cada factor) y 20 filas (una por cada ítem), y contendrá únicamente valores de unos y ceros. Los 10 ítems que se espera que estén relacionados con Extraversión presentarían en la matriz diana un valor de 1 en una columna (que representa al factor Extraversión), y cero en la otra columna (que representa al factor de Responsabilidad). El patrón de saturaciones factoriales esperado de los 10 ítems de Responsabilidad quedará igualmente representado (en este caso con el patrón de unos y ceros acorde al hipotetizado factor de Responsabilidad). Una vez propuesta la matriz diana, la rotación busca la posición que minimiza la distancia entre las saturaciones en el patrón factorial rotado respecto a la matriz diana. Cabe indicar que lo más probable es que aun en un buen ajuste del modelo, el patrón factorial rotado no muestre valores exactos ni de unos ni de ceros, sino más bien valores cercanos a unos y a ceros. Para evaluar hasta qué punto el ajuste de la rotación ha sido congruente con la matriz diana, es habitual calcular el índice de congruencia factorial (Tucker, 1951): un valor entre .85 y .94 indica una congruencia factorial aceptable, mientras que valores superiores a .94 indican una buena congruencia factorial (Lorenzo-Seva y ten Berge, 2006). Este tipo de rotación se conoce como *rotación procusteana completamente especificada* (ya que la matriz diana hipotetiza cómo deberían ser *todas* las saturaciones del patrón factorial rotado).

El siguiente paso hacía la relajación de las restricciones del modelo se conoce como *rotación procusteana semi-especificada* y fue propuesta por Browne tanto en la rotación ortogonal (1972a) como en la rotación oblicua (1972b). El aspecto del modelo que se relaja es el suponer que cada ítem ha de presentar una saturación perfecta (es decir, un valor de 1) en el factor con el que el ítem está relacionado. El aspecto del modelo que *no* se relaja es asumir que todos los ítems son buenos indicadores de uno u otro factor (y por tanto muestran saturaciones cercanas a cero en los factores con los que no se relacionan). Una vez más se propone una matriz diana, si bien sólo se especifica el valor de algunos de los elementos de las matriz diana, dejando sin especificar el resto. Los valores de la matriz diana que se especifican corresponden a los valores que se espera que sean cero en el patrón factorial rotado. Siguiendo con nuestro ejemplo de 20 ítems, la matriz diana consistirá igualmente en una matriz de 2 columnas (una columna por cada factor) y 20 filas (una por cada ítem). Los 10 ítems que se espera que estén relacionados con Extraversión presentarán en la matriz diana un valor libre en una columna (que representa al factor Extraversión), y cero en la otra columna (que representa al factor de Responsabilidad). El patrón de saturaciones factoriales esperado de los 10 ítems de Responsabilidad quedará igualmente representado (en este caso con el patrón de valores libres y ceros acorde al hipotetizado factor de Responsabilidad). Una vez más cabe indicar que lo más probable es que aun en un buen ajuste del modelo, el patrón factorial rotado no muestre valores exactos ni de unos ni de ceros, sino más bien valores cercanos a ceros (para las saturaciones especificadas) o marcadamente diferentes de cero (para las saturaciones no especificadas). La rotación procusteana semi-especificada es menos restrictiva por el hecho de que no se hipotetiza que todos los ítems son *indicadores puros* de un factor u otro: más bien se considera que todos los ítems son *buenos marcadores* de uno u otro factor del test.

El siguiente paso hacia la relajación de las restricciones del modelo se conoce como el *ajuste de clusteres independientes* (McDonald, 2000, 2005). Como Lloret et al. (2014) exponen, se trata de identificar los factores rotados en función de un número reducido de buenos marcadores (entre dos o tres marcadores por factor). Desde un punto de vista computacional, se puede obtener el ajuste a este modelo mediante una rotación semi-especificada en la que sólo unas cuantas filas de la matriz diana presentan valores de cero. En nuestro ejemplo de 20 ítems, 2 ítems de extraversión se seleccionarían como los marcadores del factor de Extraversión, mientras que otros 2 ítems se seleccionarían como los marcadores del factor de Responsabilidad. Por tanto la matriz diana sólo especificaría los valores de estos 4 ítems que se espera que sean cercanos a cero en el patrón rotado. La posición concreta de los ceros en la matriz diana dependerá del factor que se hipotetiza que marca cada uno de los 4 ítems. Por otra parte, los otros 16 ítems no presentarán ninguna especificación en la matriz diana. El ajuste de clusteres independientes es menos restrictivo por el hecho de que sólo se hipotetiza que algunos ítems son buenos marcadores de uno u otro factor del test.

Hasta aquí, la relajación del modelo ha pasado por definir una matriz diana sobre la que sucesivamente hemos ido reduciendo el número de restricciones. Sin embargo, el investigador ha tenido que proponer, cuando menos, un cierto número de marcadores por factor. El siguiente paso hacía la relajación pasa por evitar tener que proponer cuáles son los marcadores para cada factor. Este tipo de ajuste se consigue mediante procedimientos de rotación que construyen ellos mismos una matriz diana, y le asignan los valores que van a ser especificados. Un ejemplo de este tipo de rotaciones es Promin (Lorenzo-Seva, 1999). Promin se basa en: (1) utilizar los pesos propuestos por Cureton y Mulaik (1975) para identificar los ítems potencialmente más simples (es decir, los mejores marcadores) antes de iniciar la rotación para que sean precisamente esos ítems los que dirijan la rotación; (2) construir una matriz diana semi-specificada (donde se especifica los valores que se esperan que sean cercanos a cero); y (3) calcular la rotación procusteana semi-especificada oblicua. Podríamos decir que Promin toma las decisiones difíciles en la propuesta del modelo (desde el grado de especificación del modelo, hasta los valores particulares especificados). De este modo Promin está muy cerca del polo puramente exploratorio por lo que se refiere al investigador (que no ha de avanzar ningún parámetro del modelo, excepto el número de dimensiones).

Dado el continuo entre el polo exploratorio y el confirmatorio que acabamos de describir, resulta interesante preguntarse donde quedarían los procedimientos clásicos de rotación (como por ejemplo, Varimax, Oblimin o Promax). Estos métodos de rotación pretenden obtener un patrón factorial que se acerque al máximo al criterio de *simplicidad factorial*. Para que se cumpliese este criterio perfectamente, haría falta que todos los ítems fuesen indicadores perfectos de uno u otro factor (aunque es cierto que el investigador no ha de decidir para ningún ítem de qué factor sería un indicador el ítem), resultando un patrón factorial rotado con valores exactos de unos o ceros únicamente. Aunque no es fácil decidir donde deberían situarse estos procedimientos clásicos en el continuo que hemos descrito, lo que sí parece claro es que proponen un modelo con restricciones (la propia simplicidad factorial para todos los ítems sin excepción) que les aleja del polo puramente exploratorio. Una consecuencia de esto es que aplicar, por ejemplo, Oblimin para rotar una solución factorial donde algunos de los ítems no son indicadores puros de un factor u otro (es decir, ítems que son de carácter complejo), tiene la consecuencia negativa de que su complejidad es distribuida en mayor o menor medida a todas las saturaciones del patrón rotado (como consecuencia de intentar minimizar la estructura compleja de algunos de los ítems). Dicho de otro modo, el procedimiento de rotación no nos permitiría identificar correctamente hasta qué punto algunos de los ítems sí son buenos marcadores de alguno de los factores.

Finalmente, nos falta posicionarnos sobre una restricción habitualmente utilizada en el contexto del AFE de test psicológicos: la ortogonalidad de los factores rotados. Izquierdo et al. (en prensa) constatan que una proporción elevada de estudios en el contexto del AFE utiliza el procedimiento de rotación Varimax (es decir, un procedimiento que impone la ortogonalidad de los factores). Desde nuestro punto de vista, imponer la ortogonalidad de los factores rotados tiene como consecuencia ocular la posible relación de dependencia entre los factores. Por otra parte, cuando se permite la oblicuidad de los factores no se está imponiendo: es decir, que si se utiliza un procedimiento de rotación oblicuo, pero los factores son en realidad de naturaleza independiente entre ellos, la solución factorial rotada mostrará correlaciones entre factores cercanas a cero (y por tanto despreciables). Desde este punto de vista, nuestro consejo es aplicar sistemáticamente rotaciones oblicuas, como ya aconsejaba Browne en su revisión de la rotación factorial del 2001.

Nuestro programa FACTOR (Lorenzo-Seva y Ferrando, 2013), incorpora todos los procedimientos de rotación no restrictos que hemos comentado en este trabajo. La responsabilidad del investigador aplicado es decidir en función del test psicológico que pretende analizar dónde quiere ubicarse él mismo en el continuo existente entre el polo exploratorio y el polo confirmatorio.

Discusión

En los últimos años, y especialmente en nuestro país, se han publicado una serie de guías y revisiones actualizadas del AFE dirigidas a investigadores aplicados. Parece claro pues que la técnica sigue viva, y sigue despertando interés. Si además (como esperamos) estas recomendaciones acaban calando en el campo aplicado, el nivel de las investigaciones que utilizan AFE en nuestro país (que son muchas) mejorarán substancialmente.

El discurso central de nuestra aportación, por otra parte, es que es posible dar aún un paso más. El investigador con inclinación a hacerlo puede profundizar en las bases y en la lógica de algunos puntos claves del AFE, y este conocimiento adquirido, a su vez, le permitirá tomar las mejores decisiones en cada caso. Decisiones que, como hemos visto, van más allá de las recomendaciones generales, requieren una evaluación a fondo por parte del investigador, y refieren a puntos clave tales como la elección del modelo más apropiado –lineal o no lineal- dadas las características de sus datos o la transformación más acorde con las hipótesis de su investigación y la información de que dispone.

Acknowledgements.- The research was partially supported by a grant from the Catalan Ministry of Universities, Research and the Information Society (2009SGR1549) and by a grant from the Spanish Ministry of Education and Science (PSI2011-22683).

Referencias

Browne, M. W. (1972a). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 115-120.

Browne, M. W. (1972b). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, *25*, 207-212.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111-150.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, *31*, 33-42.

Cureton, E. E., y Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. *Psychometrika*, *40*, 183-195.

Ferrando, P. J. (1994). El problema del factor de dificultad: una revisión y algunas consideraciones prácticas.*Psicológica.* 15(2), 275-283.

Ferrando, P. J. (2009). Difficulty, discrimination and information indices in the linear factor-analytic model for continuous responses. *Applied Psychological Measurement*, 33, 9-24.

Ferrando, P. J. y Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory. Technical Report.* Department of Psychology, Universitat Rovira i Virgili. Retrieved from http://psico.fcep.urv.cat/utilitats/factor/

Hofstee, W. K. B., Ten Berge, J. M. F., y Hendricks, A. A. J. (1998). How to score questionnaires. *Personality and Individual Differences, 25*, 897-910.

Izquierdo, I., Olea, J. y Abad, F.J. (en prensa). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema.*

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36.

Lee, C. H., 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.

Lloret, S., Ferreres, A., Hernández, A. y Tomás, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología, 30* (3)*,* 1151-1169.

Lorenzo-Seva, U. (1999). Promin: a method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347-356.

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*, 497-498.

Lorenzo-Seva, U.y ten Berge, J. M. F. (2006). Tucker's Congruence Coefficient as a Meaningful Index of Factor Similarity. *Methodology*, *2*, 57-64.

McDonald, R. P. (1999). *Test theory: a unified treatment.* Mahwah (NJ): LEA.

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.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of educational statistics, 11*(1), 3-31.

Muthén, B. (1993). Goodness of fit with categorical and other nonnormal variables. En K.A. Bollen y J.S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park: Sage.

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.

Reise, S. P. y Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5,* 27-48.

Rigdon, E. E. (2010). Polychoric correlation coefficient. En N. Salkind (Ed.), *Encyclopedia of research design.* (pp. 1046-1049). Thousand Oaks, CA: SAGE Publications, Inc.

Rigdon, E. E. y Ferguson, C. E. (1991) The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research, 28*, 491-497.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17.* Iowa City: Psychometric Society.

Ten Berge, J.M.F., y Nevels, K. (1977). A general solution to Mosier's oblique Procrustes problem. *Psychometrika*, *42*, 593-600.

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

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report No.984*, Washington D.C.: Department of the Army.

(Artículo recibido: 27-5-2014; revisión recibida: 5-6-2014; aceptado: 8-6-2014)