

Análisis factorial exploratorio: cuestiones conceptuales y metodológicas

*Exploratory factor analysis:
conceptual and methodological issues*

Irini Mavrou

Universidad Antonio de Nebrija.

imavrou@alumnos.nebrija.es

Mavrou, I. (2015). Análisis factorial exploratorio: cuestiones conceptuales y metodológicas. *Revista Nebrija de Lingüística Aplicada* (2015) 19.

RESUMEN

El objetivo de este artículo es revisar las aportaciones de varios especialistas sobre la técnica multivariante del Análisis Factorial Exploratorio. En primer lugar, se discuten las diferencias entre los dos principales métodos del AFE, el Análisis de Componentes Principales y el Análisis Factorial Común. A continuación, se presentan los estándares que deben cumplirse para su aplicación, así como ciertos métodos de extracción y de rotación de los factores. Finalmente, se comentan algunos criterios para la determinación de los factores que van a ser retenidos y se proporcionan algunas directrices para la evaluación de la significancia de las cargas factoriales en función del tamaño muestral.

Palabras clave: análisis factorial exploratorio, extracción de factores, rotación de factores, número de factores que serán retenidos, significación de las cargas factoriales

ABSTRACT

The purpose of this article is to provide an overview of several important contributions regarding Exploratory Factor Analysis (EFA). After briefly addressing the differences between the two main methods of EFA, i.e. Principal Component Analysis and Common Factor Analysis, the standards that have to be met for its implementation, as well as certain methods of factor extraction and rotation are discussed in detail. Furthermore, the criteria for defining the number of factors to retain when conducting EFA are presented and some guidelines for evaluating the significance of factor loadings based on the sample size are offered.

Keywords: exploratory factor analysis, factor extraction, factor rotation, number of factors to retain, significance of factor loadings

Fecha de recepción: 15/4/2015

Fecha de aceptación: 10/9/2015

1. INTRODUCCIÓN

El Análisis Factorial Exploratorio (AFE) [Exploratory Factor Analysis; EFA] es una técnica estadística que permite explorar con mayor precisión las dimensiones subyacentes, constructos o variables latentes de las variables observadas, es decir, las que observa y mide el investigador. Si, por ejemplo, para los propósitos de un estudio se elabora o se adapta una batería de

pruebas (pruebas de medición de constructos cognitivos, cuestionarios de variables afectivas), el AFE se aplica con el objetivo de comprobar hasta qué punto estos instrumentos o los ítems que los conforman representan adecuadamente los constructos latentes de interés o diferentes dimensiones del mismo constructo. Asimismo, el AFE sería una valiosa técnica para examinar en qué medida ciertos indicadores cuantitativos utilizados en el campo de la Adquisición de Segundas Lenguas (ASL) (por ejemplo, índice de subordinación, longitud promedio de la cláusula y de la unidad terminal) representan adecuadamente diferentes constructos o componentes de la actuación y del desarrollo lingüístico de los aprendientes de lengua extranjera (LE) (por ejemplo, complejidad sintáctica, precisión lingüística, fluidez, etc.).

En la bibliografía especializada y en artículos dedicados a la revisión de las prácticas del AFE más frecuentemente utilizadas en la investigación empírica se ha constatado una preferencia por ciertos métodos del AFE (mayor uso del Análisis de Componentes Principales frente al Análisis Factorial Común), de rotación (ortogonal y, en menor medida, rotación oblicua) y de determinación de los factores (regla de Gutman-Kaiser y, por lo general, el uso de un solo criterio). Asimismo, se ha comprobado que en gran parte de esta investigación los autores proporcionan de forma limitada detalles sobre los procedimientos que han seguido o las decisiones conceptuales que han tomado (Fabrigar, Wegener, MacCallum & Strahan, 1999; Conway & Huffcutt, 2003).

Partiendo de estas reflexiones, el objetivo del presente artículo es exponer, de manera breve, las aportaciones de varios especialistas sobre una serie de cuestiones teóricas y metodológicas que conciernen a la aplicación del AFE, esperando de este modo realizar una contribución que pueda servir de referencia a aquellos investigadores de la ASL interesados en aplicar dicha técnica en sus estudios.

2. ANÁLISIS DE COMPONENTES PRINCIPALES VS. ANÁLISIS FACTORIAL COMÚN

En la bibliografía especializada existe un gran debate con respecto a la idoneidad del Análisis de Componentes Principales (ACP) [Component Analysis; CA], método más frecuentemente utilizado en la investigación empírica, frente al Análisis Factorial Común (AFC) [Factor Analysis; FA] (Velicer & Jackson, 1990a, 1990b; Widaman, 1993). La gran popularidad del ACP se ha atribuido a su mayor simplicidad y al hecho de que constituye la opción por defecto de varios programas estadísticos, como el Statistical Package for the Social Sciences (SPSS) (Fabrigar et al., 1999; Conway & Huffcutt, 2003; Costello & Osborne, 2005). Por otro lado, varios autores advierten sobre la necesidad de crear en los investigadores una postura más crítica a la hora de llevar a cabo pruebas estadísticas y de considerar los supuestos que subyacen a los modelos estadísticos que se van a efectuar, más allá de la práctica común de basarse en las opciones default de los programas informáticos. Al respecto son muy ilustrativas las siguientes palabras de Gorsuch (2003) y de Steiger (2004).

In 1960 computers entered psychology, but they were simple and slow ... It was then Henry Kaiser at the University of Illinois introduced the simplest complete computer package, called "Little Jiffy". It was doable in those computers because it was CA [Component Analysis] and had no communality estimation procedure ... In 1970 (Kaiser, 1970) he introduced "A Second Generation Little Jiffy" but then it was too late. The computer packages had already picked up the runnable "Little Jiffy" and that is still often the default in major statistical packages. My personal opinion is that the rationales for CA developed as a post hoc explanation because so many used a computer package which had "Little Jiffy" as the default. BUT NOTE: the origin of any construct in science is not judged by its history but only by its merits. (Gorsuch, 2003:156)

We now have a much better idea what we should do in regression and ANOVA. What the majority of researchers *will* do it is, unfortunately, determined to a considerable extent by what SPSS, SAS, and other manufacturers of general-purpose statistical software are willing to implement in their programs. For better or worse, statistical practice is software-driven (Steiger, 2001). (Steiger, 2004:71)

De los fragmentos anteriores se derivan dos conclusiones: (a) los programas informáticos con los que se cuenta en la

actualidad son más potentes y efectúan en menor tiempo operaciones complejas que las computadoras y programas antiguos no eran capaces de hacer; (b) no es el programa estadístico el que decide qué tipo de análisis se va a efectuar o cómo se interpretarán los resultados; estas decisiones recaen en la responsabilidad del propio investigador. En palabras de Field (2009):

It is important to remember that SPSS may appear to be very clever, but in fact it is not. Admittedly, it can do lots of complex calculations in a matter of seconds, but what it can't do is control the quality of the model that is generated – to do this requires a human brain (and preferably a trained one). SPSS will happily generate output based on any garbage you decide to feed into the data editor and SPSS will not judge the results or give any indication of whether the model can be generalized or if it is valid. (Field, 2009:225)

En cuanto a los métodos del AFE anteriormente mencionados, si bien es cierto que ambos presentan varias similitudes e incluso han llegado a utilizarse los términos ACP y AFC de manera intercambiable, los modelos matemáticos de los que parten y los supuestos conceptuales subyacentes a su aplicación son diferentes (Rietveld & van Hout, 1993; DeCoster, 1998; Fabrigar et al., 1999; Conway & Huffcutt, 2003; Costello & Osborne, 2005; Hair, Black, Babin & Anderson, 2010).

Desde un punto de vista conceptual, el ACP asume que los puntajes individuales en una serie de medidas causan o definen el componente; según DeCoster (1998): component=composite of the observed variables. Por tanto, el ACP debería aplicarse cuando el objetivo es reducir la dimensionalidad inicial de los datos en un conjunto menor de componentes que expliquen la varianza total observada. Por otro lado, en el AFC las variables observadas se consideran como el resultado de un factor o constructo subyacente (DeCoster, 1998). Por tanto, mediante el AFC se intenta identificar las dimensiones o constructos latentes que den cuenta de la varianza común entre las variables, eliminándose la varianza no compartida, propia de una variable, así como la varianza de error de medición, ocasionada por factores como el cansancio o sesgos en la recolección de datos.¹

Dado que difícilmente se puede asumir que las variables estén libres de error, Gorsuch (1974, 1990, en Rietveld & van Hout, 1993:267-268) considera el AFC como una técnica más fundamentada teóricamente. Según el autor, las cargas de las variables sobre los componentes tienden a ser más altas que las cargas sobre los factores (efecto denominado inflación), lo cual no representa necesariamente la realidad. Sin embargo, Velicer y Jackson (1990a) consideran el término inflación como “peyorativo y desorientador”. Como apuntan estos autores:

It involves the implicit assumption that the factor loading is the *correct value*. It would be equally appropriate but similarly misleading to describe the factor loadings as *deflated* ... Arguments that the slightly larger loading are *better* or *worse* are not likely to convince anyone. (Velicer & Jackson, 1990a:9)

No obstante, es oportuno mencionar que argumentos como los de Gorsuch han ganado mayor apoyo. Bentler y Kano (1990), por ejemplo, señalan: “We would argue that the error structure of component analysis is not appropriate to most real applications, and that the search for appropriate models should typically be limited to models with uncorrelated additive errors” (Bentler & Kano, 1990:68). Conway y Huffcutt (2003) sostienen que la mayoría de los investigadores atribuyen significado más allá de las variables observadas, por lo que, según ellos, el AFC parece ser la mejor opción. Widaman (1993), por su parte, desaconseja el ACP cuando el objetivo es obtener estimadores de los parámetros poblaciones que reflejen los constructos latentes de interés, mientras que Fabrigar et al. (1999) opinan que, a pesar de que a veces se alude al ACP como si se tratara de un método del análisis factorial, “... it is not factor analysis at all” (Fabrigar et al., 1999:275).

Otra cuestión que debe considerarse tiene que ver con la terminología utilizada. Dadas las diferencias entre el ACP y el AFC, Rietveld y van Hout (1993) sugieren el uso del término ‘componentes’ para las dimensiones que resultan de la aplicación del ACP, mientras que, según ellos, el término ‘factores’ debería guardarse para las dimensiones que derivan de la aplicación del AFC. En palabras de estos autores:

Although different assumptions are made in factor analysis (FA) and principal component (PC) analysis about the composition of the variance that is to be accounted for, often the more general term ‘factor analysis’ is even used in those cases where principal component analysis is being discussed. The same goes for the use of the term ‘factors’ where in fact ‘components’ is

the proper term. (Rietveld & van Hout, 1993:251)

Evaluar de manera minuciosa las propiedades del ACP y del AFC excede los propósitos de este artículo. Asimismo, resulta complejo tomar una postura en el debate sobre qué método es el más adecuado. Como apuntan Trendafilov, Unkel y Krzanowski (2013), en ocasiones, puede ser especialmente informativo e incluso más aceptable aplicar ambos métodos, comparar directamente las soluciones factoriales y examinar si estas difieren considerablemente. A pesar de que la aplicación de ambos métodos a partir del mismo conjunto de datos supone un trabajo más laborioso por parte del investigador, parece ser la mejor opción y, probablemente, el camino más seguro para la obtención de resultados firmes y más fáciles de contrastar.

3. ESTÁNDARES PARA LA APLICACIÓN DEL ANÁLISIS FACTORIAL EXPLORATORIO

En la aplicación del AFE es necesario considerar tres estándares fundamentales que se relacionan con: (1) el tamaño muestral, (2) la ratio mínima de casos por variable y (3) la ratio de variables por factor.

Por lo general, el AFE no debería aplicarse cuando el tamaño muestral está constituido por menos de 50 casos. Además, es preferible que dicho tamaño sea mayor que 100 o, idealmente, 300-400 casos (Hair et al., 2010) a fin de minimizar la probabilidad de equivocación (valores inflados), aumentar la precisión de los estimadores poblacionales y, por tanto, la confianza en las inferencias elaboradas (Osborne & Costello, 2004). Arrindell y van der Ende (1985), por su parte, concluyeron que la obtención de una solución factorial estable es posible cuando el tamaño muestral se aproxima a 20 veces el número de factores.²

Por otro lado, es interesante mencionar que en estudios basados en simulaciones computacionales se ha constatado que incluso con muestras pequeñas puede tener sentido aplicar el AFE. Como apuntan de Winter, Dodou y Wieringa (2009), un tamaño muestral pequeño no debería constituir el único criterio para rechazar el AFE, ya que, aun bajo condiciones restrictivas, se podrían revelar patrones latentes valiosos. Por tanto, según recomiendan estos autores, “aplicar el AFE en una fase exploratoria es preferible que rechazarlo a priori” (de Winter, Dodou & Wieringa, 2009:171).

En lo que se refiere a la ratio de observaciones (casos) por variable, aunque no existe un acuerdo unánime, se sugiere que esta sea por lo menos de 10-15 casos por cada variable. Si las communalidades son bajas y los factores no bien determinados [poorly overdetermined], entonces se recomienda una ratio de al menos 20:1 (MacCallum, Widaman, Preacher & Hong, 2001), mientras que con tamaños muestrales grandes, la ratio requerida puede ser menor. Como puntualizan Osborne y Costello (2004): “total N matters (but more so when subject:item ratio is low), and the ratio of subjects to items matters (but more so when N is relatively low), and if you have a large N or large ratio, your results will be more reliable” (Discussion section, párr. 4).

Ahora bien, es oportuno señalar que ambos criterios mencionados anteriormente se han criticado de simplicidad, puesto que el tamaño muestral requerido no puede establecerse independientemente del grado de representatividad de los factores y el valor de las communalidades (Arrindell & van der Ende, 1985; Velicer & Fava, 1998; MacCallum, Widaman, Zhang & Hong, 1999; MacCallum et al., 2001). Una ratio de 3 o, idealmente, 4 variables por factor constituye un umbral necesario para la adecuación del modelo. Además, ciertas distorsiones pueden ocurrir, cuando las communalidades de las variables son relativamente bajas (< .60-.70), debido a errores de medición o problemas de fiabilidad de los instrumentos utilizados (Velicer & Jackson, 1990b; Velicer & Fava, 1998; Fabrigar et al., 1999; MacCallum et al., 1999; MacCallum et al., 2001).

Velicer y Jackson (1990a, 1990b), refiriéndose a las discrepancias entre las soluciones proporcionadas por el ACP y el AFC, explican que estas se hacen evidentes cuando el número de variables para la identificación de los factores es insuficiente, el tamaño muestral pequeño y las saturaciones muy bajas. Bajo estas condiciones, se producen soluciones factoriales inadecuadas, difícilmente interpretables y poco generalizables. Bentler y Kanon (1990), por su parte, sostienen que, con un pequeño número de variables, el AFC es más apropiado para la comprensión del patrón de intercorrelaciones entre las variables.

Conseguir soluciones estables y replicables y hacer estimaciones más precisas de los parámetros poblacionales es posible siempre que un factor esté comprendido por un mínimo de tres ítems fuertes (saturaciones > .60-.70) y las communalidades sean altas, independientemente del tamaño muestral, el nivel de determinación de los factores o la presencia de error (Fabrigar et al., 1999; MacCallum et al., 1999; MacCallum et al., 2001). Por otro lado, como muy acertadamente

señalan Osborne y Costello (2004), en las investigaciones empíricas en las Ciencias Sociales el rango de saturaciones suele ser moderado o bajo (.32 < ? < .50), de ahí que saturaciones por encima del .50 puedan considerarse, por lo general, saturaciones fuertes [strong item loading].

Fabrigar et al. (1999) realizan dos observaciones adicionales, las cuales resulta pertinente mencionar aquí. La primera tiene que ver con la naturaleza y la composición de la muestra. Como explican estos autores, cuando la muestra presenta mayor homogeneidad en comparación con la población de la que se ha extraído, aumenta la probabilidad de que el rango de valores de las variables medidas a partir de esta muestra sea muy restringido y, por tanto, aumenta también la probabilidad de obtener correlaciones atenuadas entre las variables; tal atenuación dará lugar a bajos estimadores de las cargas factoriales y de las correlaciones entre los factores (Fabrigar et al., 1999).

La segunda cuestión tiene que ver con la calidad y la naturaleza de las variables observadas. Según Fabrigar et al. (1999), estas variables deben ser relevantes en función del dominio que se estudia, pues, en caso contrario, existe el riesgo de que el AFE no revele factores importantes o que conduzca a la obtención de factores espurios y no bien definidos. Además, el número de estas variables debería considerarse en función del número de los factores esperados, al menos cuando se tienen algunas expectativas sobre el número de estos factores, y, efectivamente, atendiendo a los fundamentos teóricos del tema estudiado.

En resumen, a la hora de aplicar el AFE, el investigador debe considerar profundamente cuestiones como las siguientes:

- (a) los valores de las communalidades (proceder con el AFE si estos valores son altos);
- (b) el grado de determinación de los factores (número de variables que comprende cada factor; idealmente 3 o 4 variables por factor, aunque en la práctica incluso 2 variables podrían ser suficientes para la identificación e interpretación de un factor determinado);
- (c) la magnitud de las cargas factoriales considerando factores relacionados con el contexto del estudio y la disciplina en la que se trabaja (saturaciones por encima del .40 o .50 pueden ser satisfactorias en la investigación empírica de la ASL);
- (d) el hecho de que, en ciertas circunstancias, un aumento del tamaño de la muestra pueda compensar cargas factoriales bajas o un número limitado de variables (Velicer & Fava, 1998);
- (e) el grado de coincidencia de las estructuras factoriales derivadas de la aplicación de diferentes métodos del AFE (ACP y AFC).

4. MÉTODOS DE EXTRACCIÓN DE LOS FACTORES

Los métodos de obtención de los factores más ampliamente utilizados son el método de Componentes Principales (véase apartado 2), la Factorización de Ejes Principales [Principal axis factoring] y el método de Máxima Verosimilitud [Maximum likelihood], siendo los dos últimos propios del AFC (Conway & Huffcutt, 2003; de Winter & Dodou, 2012).³

La Factorización de Ejes Principales (FEP) es un método iterativo basado en la extracción sucesiva de aquellos factores que explican la mayor parte de la varianza común y, también, robusto a violaciones del supuesto de normalidad (Fabrigar et al., 1999; de Winter & Dodou, 2012). Como sostienen de Winter y Dodou (2012), la FEP tiene la ventaja de recuperar factores débiles y es recomendable para soluciones factoriales para las que se cuenta con pocos indicadores (variables). Beauducel (2001) se muestra de acuerdo al señalar que, con tamaños muestrales pequeños y correlaciones moderadas entre las variables, la FEP ofrece soluciones más estables. Sin embargo, la principal limitación de este método es que no proporciona índices de ajuste [goodness of fit], tampoco permite la computación de test de significancia y de los intervalos de confianza (Fabrigar et al., 1999).

El método de Máxima Verosimilitud, aunque no presenta los inconvenientes anteriores, presupone distribución normal multivariada y ausencia de casos atípicos, lo cual con cierto tipo de datos es difícil de conseguir. Para dar un ejemplo, en la evaluación del grado de complejidad sintáctica del discurso oral o escrito de un grupo de aprendientes de LE que poseen un nivel de dominio lingüístico bajo, se esperarían, por lo general, valores bajos en medidas lingüísticas como el índice de subordinación o la longitud promedio de la cláusula. Una asimetría positiva, como en el ejemplo anterior, implicaría desviaciones del supuesto de normalidad univariada y, en consecuencia, violación del supuesto de normalidad multivariada.

Problemas como los anteriores se podrían superar aumentando el tamaño de la muestra. Sin embargo, el método de

Máxima Verosimilitud y, en concreto, el likelihood statistic ratio puede verse afectado por el tamaño muestral. Como explican Fabrigar et al. (1999):

When N is large, even trivial discrepancies between the model and the data are likely to give rise to rejection of the model with any reasonable number of factors (Hakstian et al., 1982; M. L. Harris & Harris, 1971; Humphreys & Montanelli, 1975; MacCallum, 1990). In situations where N is small, even large discrepancies between the model and the data may not be statistically significant, thereby leading to underfactoring (Humphreys & Montanelli, 1975). (Fabrigar et al., 1999:280)

En todo caso, se ha de reconocer que la Máxima Verosimilitud constituye el método más puro del AFC, por lo que parece ser la mejor opción siempre que se cumplan los supuestos necesarios para su aplicación (tamaño muestral suficientemente grande, distribución normal multivariada). Por otro lado, la FEP resulta una alternativa adecuada y, probablemente, más viable en la investigación empírica de la ASL donde se trabaja con tamaños muestrales, en ocasiones, muy pequeños y con variables para las que, por multitud de razones, es difícil conseguir distribuciones que se aproximen a la distribución normal.⁴

5. MÉTODOS DE ROTACIÓN DE LOS FACTORES

Sin bien los modelos del AFE de más de un factor pueden conducir a un gran número de soluciones, el objetivo del investigador es encontrar aquella solución que proporcione una estructura simple (Fabrigar et al., 1999).

Thurstone used the term *simple structure* to refer to solutions in which each factor was defined by a subset of measured variables that had large loadings relative to the other measured variables (i.e. high with-in-factor variability in loadings) and in which each measured variable loaded highly on only a subset of the common factors (i.e., low factorial complexity in defining variables). (Fabrigar et al., 1999:281)

Existen varios métodos de rotación de los factores, aunque estos se pueden clasificar bajo dos categorías más amplias: las rotaciones ortogonales y las rotaciones oblicuas. Las rotaciones ortogonales (varimax, quartimax y equimax) producen factores no correlados, por lo que podrían considerarse apropiadas cuando se asume independencia de los factores, al menos desde un punto de vista conceptual (Fabrigar et al., 1999; Conway & Huffcutt, 2003; Costello & Osborne, 2005; Beavers et al., 2013).

En la investigación empírica las rotaciones ortogonales se han implementado con mayor frecuencia (Fabrigar et al., 1999; Conway & Huffcutt, 2003; Costello & Osborne, 2005), hecho que no sorprende, ya que la varimax constituye la opción default en el SPSS. Sin embargo, Costello y Osborne (2005) realizan una observación de suma relevancia.

Conventional wisdom advises researchers to use orthogonal rotation because it produces more easily interpretable results, but this is a flawed argument. In the social sciences we generally expect some correlation among factors, since behavior is rarely partitioned into neatly packaged units that function independently of one another. (Best Practice, Rotation section, párr. 1)

El mismo punto de vista mantienen Fabrigar et al. (1999) al señalar que para muchos constructos psicológicos como los rasgos de personalidad y las actitudes -objetos de estudio también en la ASL- existe suficiente evidencia teórica y empírica para esperar una correlación entre los factores que resultan de la aplicación del AFE. De este modo, las rotaciones oblicuas (oblimin direct, promax, orthoblique, procrustes) probablemente constituyen la mejor opción, al menos en las Ciencias Sociales, y, por tanto, en la ASL, especialmente cuando, en base a los fundamentos teóricos, se asume una asociación conceptual entre las variables latentes examinadas.

Según varios autores (Fabrigar et al., 1999; Conway & Huffcutt, 2003; Costello & Osborne, 2005; Beavers et al., 2013), las rotaciones oblicuas permiten obtener soluciones precisas, simples, reproducibles y más realistas. Fabrigar et al. (1999), por ejemplo, apuntan que si no existe correlación entre los factores, la rotación oblicua proporcionará estimadores cercanos a cero y similares a los de la rotación ortogonal; sin embargo, si tal correlación realmente existe, la rotación oblicua la revelará y

permitirá al investigador profundizar sobre la naturaleza conceptual de dichos factores. Incluso se ha sostenido que no existe ninguna razón para utilizar las rotaciones ortogonales. Como señalan Conway y Huffcutt (2003):

If oblique rotations sometimes produce better solutions (i.e., when the constructs are in reality correlated) and the rest of the time produce essentially equivalent solutions (i.e., when constructs are really uncorrelated or nearly so), then there seems no reason to use an orthogonal rotation. (Conway & Huffcutt, 2003:153)

Por otro lado, habría que considerar la advertencia de Hair et al. (2010) sobre la necesidad de tener mucho cuidado a la hora de validar los factores rotados oblicuamente, porque, como explican estos autores, la solución factorial puede llegar a ser específica de la muestra y no generalizable, particularmente con muestras pequeñas o bajas correlaciones entre las variables examinadas.

6. DETERMINACIÓN DEL NÚMERO DE LOS FACTORES QUE SERÁN RETENIDOS

En lo que concierne a la determinación del número de los factores que serán retenidos, en la bibliografía especializada se recomienda el uso de múltiples criterios atendiendo tanto a los fundamentos conceptuales como a la evidencia empírica (Fabrigar et al., 1999; Conway & Huffcutt, 2003; Hair et al., 2010).

Cuando se tienen algunas expectativas sobre la manera en que se agruparán las variables, se puede utilizar un criterio a priori, es decir, solicitar la extracción de un número de factores determinado. Según Hair et al. (2010), este criterio puede ser útil para probar una teoría o hipótesis o para replicar los resultados de investigaciones previas. En este caso, se adopta lo que Hair et al. (2010) denominan ‘una aproximación confirmatoria’, es decir, una aproximación mediante la cual se persigue evaluar el grado en que los datos proporcionarán la estructura esperada.

Otro criterio, probablemente el más utilizado, es la regla de Gutman-Kaiser, esto es, conservar aquellos factores con valores propios mayores que 1. No obstante, este criterio se ha criticado por proporcionar resultados imprecisos y, en ocasiones, por comprometer la capacidad explicativa de la solución factorial, especialmente cuando el número de variables es muy grande o muy pequeño (Zwick & Velicer, 1986; Velicer & Jackson 1990b; Fabrigar et al., 1999; Hair et al., 2010). Como lo ejemplifican Fabrigar et al. (1999:278): “For instance, it is not really meaningful to claim that a common factor with an eigenvalue of 1.01 is a ‘major’ factor whereas a common factor with an eigenvalue of 0.99 is not”.

Para resolver esta última cuestión se podría aplicar la regla de Gutman-Kaiser utilizando un límite más pequeño que la unidad (por ejemplo, 0.80). En este caso, resultaría útil consultar el gráfico de sedimentación [scree plot] y, además de aquellos factores que se encuentren antes del punto de inflexión del gráfico, retener algún factor más, siempre que su varianza se pueda considerar relativamente alta o próxima al límite establecido.

El porcentaje de la varianza total explicada es otro criterio que se puede utilizar para decidir el número de factores que serán retenidos. Dicho porcentaje no se ha determinado de manera precisa, aunque en las Ciencias Sociales se recomienda que continúe la extracción de factores hasta alcanzar un porcentaje satisfactorio de la varianza total explicada. Como umbral para la extracción de los factores se suele establecer un mínimo de 60% (Hair et al., 2010) o de 70-80% (Rietveld & van Hout, 1993).

Sumado a lo anterior, cabe señalar que se considera como una práctica aceptable variar el número de los factores antes de optar por la estructura más adecuada (Costello & Osborne, 2005; Beavers et al., 2013). Asimismo, retener un mayor número de factores parece ser preferible que eliminar factores potencialmente necesarios, lo cual llevaría a una pérdida de información importante (Zwick & Velicer, 1986; Fava & Velicer, 1996). Beauducel (2001), por ejemplo, apunta que la extracción de un número de factores ligeramente mayor con el uso de métodos como la FEP ayuda a reproducir factores disueltos. No obstante, hay que tener en cuenta que esta práctica podría conducir a saturaciones factoriales débiles (Beavers et al., 2013), a la construcción de factores secundarios en perjuicio de los factores principales o de factores con muchas saturaciones bajas y pocas saturaciones altas y, por tanto, a estructuras complejas y difícilmente interpretables (Comrey 1978, en Zwick & Velicer, 1986:432).5

7. SIGNIFICACIÓN DE LAS CARGAS FACTORIALES

Una última cuestión que se debe tomar en cuenta a la hora de efectuar el AFE es la relativa a la significación de las cargas factoriales. Como se señaló anteriormente, Osborne y Costello (2004; Osborne & Costello, 2005) opinan que cargas de .50 pueden considerarse, por lo general, fuertes. Otros autores sugieren evaluar la magnitud de las cargas factoriales en función del tamaño muestral. Stevens (2002, en Field, 2009:644), por ejemplo, sostiene que para un nivel de significación $\alpha = .01$ (bilateral) y $n=50$, cargas factoriales de .722 pueden considerarse significativas, mientras que para $n=100$ pueden ser suficientes cargas de aproximadamente .512. Según las directrices propuestas por Hair et al. (2010), para $n=70$ y $n=85$ las cargas factoriales deberían alcanzar el .65 y el .60 respectivamente. Field (2009), por su parte, sugiere elevar al cuadrado las cargas factoriales a fin de obtener una estimación de la cantidad de varianza de la variable en cuestión que se explica por un factor determinado. Además, se recomienda interpretar solo aquellas cargas factoriales con un valor absoluto superior a .40, es decir, aquellas que dan cuenta del 16% de la varianza de la variable (Stevens, 2002, en Field, 2009:645).

En función de lo expuesto y teniendo en cuenta que la aplicación del AFE requiere tamaños muestrales relativamente grandes, se podría concluir que la identificación de un factor por medio de 3 variables que presentan en dicho factor cargas de .60 (o idealmente cargas mayores que .71) sería suficiente para asumir que estas variables son buenos indicadores del constructo latente de interés, y ello, a su vez, implicaría la validez convergente del constructo.⁶

8. CONCLUSIONES

La utilidad del AFE como herramienta estadística en ciertos campos de estudio es indiscutible. En la práctica, sin embargo, el investigador se ha de hacer frente a una serie de cuestiones conceptuales y metodológicas y debe considerar y elegir cuidadosamente, entre la cantidad de opciones y métodos de análisis, los que mejor se adecuen a los objetivos y al contexto de su investigación.

Fabrigar et al. (1999) advierten que decisiones pobres respecto al diseño del estudio y la metodología utilizada conducirán a resultados también pobres y confusos. Los mismos autores añaden:

Researchers sometimes base their analyses on studies with less than optimal features, commonly make questionable choices when selecting analytic procedures, and do not provide sufficient information for readers to make informed judgments of the soundness of the EFA being reported. (Fabrigar et al., 1999:295)

Entre las causas de las pobres y cuestionables decisiones y prácticas con respecto al AFE, Fabrigar et al. (1999) aluden al insuficiente entrenamiento en los programas de estudio ofrecidos, así como al hecho de que los artículos que abordan el AFE son de carácter técnico, publicados en revistas especializadas, a las que difícilmente recurrirá el estudiante de Psicología, y mucho menos el estudiante de Lingüística Aplicada.

Efectivamente, en el contexto de los programas de máster y doctorado en Lingüística Aplicada a la LE puede ser de mayor utilidad y, probablemente, más viable enseñar al estudiante técnicas estadísticas como la prueba de χ^2 , la t de Student o el análisis de varianza (ANOVA), ya que en una gran parte de la investigación en este ámbito se emplea un diseño inter o intrasujeto (nativos vs. no nativos, hablantes de diferente LE, actuación de los mismos informantes en diferentes momentos o a partir de diferentes tareas) y se trabaja con variables categóricas (lengua materna, LE, sexo, nivel de dominio lingüístico en la LE, por mencionar algunas). Por otro lado, se ha de reconocer que la insuficiente preparación e instrucción respecto a técnicas estadísticas como el AFE resta al estudiante o al investigador de la ASL la oportunidad de descubrir las opciones y las posibilidades que tiene en términos de análisis estadístico de sus datos o incluso del tipo de estudio que pueda emprender.

Otro punto señalado por Fabrigar et al. (1999) tiene que ver con la tendencia a 'hacer lo que hicieron otros investigadores' y a seguir las opciones default de los programas estadísticos como el SPSS. Aunque esta podría ser una práctica justificable,

al menos en las primeras etapas de la carrera académica e investigadora, podría convertirse en uno de los principales obstáculos para futuras perspectivas laborales del investigador.

El objetivo del presente artículo ha sido sintetizar una serie de cuestiones relacionadas con la técnica del AFE, esperando que las aportaciones aquí recogidas sean de utilidad para los investigadores de la ASL en lo referente a la justificación apropiada de sus decisiones y de los métodos por los que opten al efectuar el AFE. Finalmente, se insta a los investigadores sobre la necesidad de intentar, en la medida de lo posible, estar al tanto de diferentes fuentes y referencias estadísticas actuales y de proporcionar todos los detalles en cuanto a la aplicación del AFE en sus manuscritos.

REFERENCIAS BIBLIOGRÁFICAS

- Arrindell, W. A. & 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(2), 165-178. doi:10.1177/014662168500900205
- Beauducel, A. (2001). On the generalizability of factors: The influence of changing contexts of variables on different methods of factor extraction. *Methods of Psychological Research Online*, 6(1), 69-96.
- 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). Disponible en: <http://pareonline.net/pdf/v18n6.pdf>
- Bentler, P. M. & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, 25(1), 67-74. doi:10.1207/s15327906mbr2501_8
- Conway, J. M. & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147-168. doi:10.1177/1094428103251541
- Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7). Disponible en: <http://pareonline.net/pdf/v10n7.pdf>
- DeCoster, J. (1998). Overview of factor analysis. Disponible en: <http://www.stat-help.com/notes.html>. [Acceso Febrero 17, 2015]
- de Winter, J. C. F. & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, 39(4), 695-710. doi: 10.1080/02664763.2011.610445
- de Winter, J. C. F., Dodou, D. & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44, 147-181. doi: 10.1080/00273170902794206
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. doi:10.1037/1082-989X.4.3.272
- Fava, J. L. & Velicer, W. F. (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement*, 56(6), 907-929. doi:10.1177/0013164496056006001
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: SAGE Publications.
- Gorsuch, R. L. (2003). Factor analysis. En J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology, Vol. 2: Research methods in psychology* (pp. 143-164). Hoboken, NJ: John Wiley & Sons.
- Hair, J. F. Jr., Black, W. C., Babin, B. J. & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

- MacCallum, R. C., Widaman, K. F., Preacher, K. J. & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36(4), 611-637. doi:10.1207/S15327906MBR3604_06
- MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4 (1), 84-99. doi:10.1037/1082-989X.4.1.84
- Osborne, J. W. & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation*, 9(11). Disponible en: <http://pareonline.net/getvn.asp?v=9&n=11>
- Rietveld, T. & van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Berlin, Germany: De Gruyter Mouton.
- Steiger, J. H. (2004). Paul Meehl and the evolution of statistical methods in psychology. *Applied and Preventive Psychology*, 11, 69-72. doi:10.1016/j.appsy.2004.02.012
- Trendafilov, N. T., Unkel, S. & Krzanowski, W. (2013). Exploratory factor and principal component analyses: Some new aspects. *Statistics and Computing*, 23(2), 209-220. doi:10.1007/s11222-011-9303-7
- Velicer, W. F. & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3(2), 231-251. doi:10.1037/1082-989X.3.2.231
- Velicer, W. F. & Jackson, D. N. (1990a). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25(1), 1-28. doi:10.1207/s15327906mbr2501_1
- Velicer, W. F. & Jackson, D. N. (1990b). Component analysis versus common factor analysis: Some further observations. *Multivariate Behavioral Research*, 25(1), 97-114. doi:10.1207/s15327906mbr2501_12
- Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28(3), 263-311. doi:10.1207/s15327906mbr2803_1
- Zwick, W. R. & Velicer, W. F. (1986).%2