

RESUMEN DE MALAS PRÁCTICAS EN ANÁLISIS FACTORIAL¹

En este documento se indican las principales prácticas incorrectas al realizar un Análisis Factorial (ver tabla 1).

1. Realizar un análisis de componentes principales (ACP) en lugar de un Análisis Factorial Exploratorio para conocer la estructura factorial del constructo.

El Análisis de Componentes Principales (opción por defecto de SPSS) es conceptualmente diferente del Análisis Factorial Exploratorio (AFE). ACP se puede utilizar cuando **no se asume la existencia de una variable latente** (constructo) y el objetivo es encontrar un subconjunto de componentes que resuman la información de un amplio conjunto de indicadores (ítems). Por ejemplo, un investigador ha tomado muchos indicadores de algún fenómeno como la actividad cerebral de 256 canales (MEG) y necesita reducir toda esta información en unos pocos componentes que sigan patrones comunes de covariación. Para la medida de constructos psicológicos se asume que los ítems son reflejo de una variable latente (modelos reflectivos), por lo tanto, **no es correcto el uso de ACP** y debe utilizarse AF, evitándose así errores como pesos factoriales inflados o informar de una varianza explicada (autovalores) superior a la que realmente se tiene.

En la práctica incorrecta de aplicar primero un ACP (en lugar de un AFE) para a continuación realizar un AFC se observarían importantes discrepancias, como por ejemplo, valores más bajos para los pesos factoriales en el AFC (téngase en cuenta, además, que no es correcto utilizar la misma muestra para ambos análisis como se comenta a continuación).

2. Realizar un Análisis Factorial Exploratorio (AFE) y CON LOS MISMOS DATOS poner a prueba la estructura factorial con un Análisis Factorial Confirmatorio (AFC).

Si solo se dispone de UNA muestra debe elegirse entre realizar un AFE o un AFC, dependiendo de la evidencia disponible y del desarrollo de la teoría acerca del modelo que sustenta la medida: AFE se recomienda cuando apenas existe evidencia empírica sobre la estructura factorial, lo que sucede, por ejemplo, en las etapas previas de construcción de nuevos instrumentos. AFC se recomienda cuando ya se conoce la estructura y se trata de poner a prueba si se replica, por ejemplo, cuando el instrumento se aplica en otro país.

Cuando en un estudio se han tomado dos muestras y no se conoce bien la estructura factorial del instrumento, se puede aplicar en la primera muestra un AFE y en la segunda muestra un AFC para obtener una mayor evidencia de que la estructura inicial se replica: no se recomienda dividir aleatoriamente en dos una muestra, sino que es mejor separar en función de criterios temporales (datos tomados en distintos momentos) puesto que si se ha producido una contaminación en parte de la muestra que conduce a una solución factorial incorrecta, al aleatorizar la contaminación estará presente en las dos muestras. Es un error similar a realizar un AFE para replicar el resultado con un AFC en la misma muestra.

3. Decidir el número de factores mediante la regla de Kaiser

Cuando se realiza un AFE previamente hay que fijar el número de factores a extraer, el procedimiento por defecto en SPSS es la regla de Kaiser (autovalores >1). Esta regla se ha mostrado inadecuada y la recomendación actual es utilizar varios procedimientos y entre ellos, es más recomendado es el Análisis Paralelo.

¹ Jesús M^a Alvarado (2019). Resumen de Malas prácticas en Análisis Factorial. Asignaturas de Medición y Validez del Máster de Metodología de las Ciencias del Comportamiento y de la Salud. Facultad de Psicología. Universidad Complutense de Madrid.

4. Realizar Rotación Varimax, cuando teóricamente que los factores están relacionados.

En AFE, entre las prácticas en desuso está la de utilizar rotaciones ortogonales como Varimax. Puesto que difícilmente los constructos psicológicos multidimensionales estarán formados por factores no-correlacionados, el aplicar este tipo de rotación puede conducir a estimaciones incorrectas de los pesos factoriales. En su lugar se recomienda el uso de rotaciones oblicuas como por ejemplo oblimin,

5. No indicar el Método de Estimación o utilizar Máxima Verosimilitud en datos ordinales.

Se debe indicar siempre el método de estimación.

Máxima verosimilitud es el método recomendado para datos continuos con distribución normal multivariante. Este supuesto debe comprobarse mediante pruebas como el coeficiente multivariante de Mardia o en su defecto realizar análisis univariados para comprobar que la asimetría y curtosis en valor absoluto son menores a 2.

En caso de incumplimiento del supuesto de normalidad multivariante deben utilizarse estimadores robustos. Si los datos son dicotómicos u ordinales el estimador recomendado es WLSMV (weighted least squares mean and variance adjusted estimator) implementado en MPLUS y R o bien DWLS (diagonal weighted least squares) implementado en Lisrel. Adicionalmente, una buena práctica es incluir en el cuerpo del artículo o como anexo la matriz de covarianzas o de correlaciones utilizada para el AF.

6. No informar adecuadamente de la estructura factorial o de la bondad de ajuste

En AFE, no solo hay que informar de los pesos factoriales en el factor con mayor saturación, sino que es obligatorio informar de las saturaciones en los demás factores.

En AFC se recomienda presentar un gráfico en el que se informe de las cargas o pesos “estandarizados”. En AFC es obligatorio informar de la bondad de ajuste de los modelos indicando el valor de chi cuadrado (con grados de libertad y valor p), RMSEA (con su IC), CFI, TLI (NNFI) y SRMS. RMSEA y SRMS mostraran buen ajuste si sus valores están por debajo de 0.08, CFI y TLI deben tomar valores de 0.95 o superior.

7. Eliminar ítems para mejorar la bondad de ajuste.

Cuando un test se aplica a una muestra diferente, el investigador puede verse tentado a eliminar ítems que no muestran adecuadas propiedades psicométricas (por ejemplo, pesos <0.30). Esta práctica es muy peligrosa ya que al eliminar ítems se está atentando contra la integridad del instrumento. Solo si se puede justificar desde la teoría es posible introducir cambios en la estructura del test. La eliminación de un solo ítem puede afectar a la representación del constructo (infrarrepresentación que afectará a la validez). Toda prueba acertada, debe mostrar que el acortamiento no afecta a la medida que se realiza, para ello se debe recurrir a pruebas externas u otros constructos de interés para mostrar que la medida no se ve afectada.

8. Eliminar ítems y continuar los análisis con los mismos datos.

Cuando está justificado eliminar ítems y por tanto proponer una prueba acertada (eliminando uno o más ítems) es preciso aplicar el “nuevo” instrumento a otra muestra para poner a prueba que la prueba acortada mantiene las propiedades psicométricas que tenían los ítems seleccionados en la prueba inicial.

9. Muestras pequeñas y número insuficiente de ítems por factor

Se recomienda aplicar el AF a muestra de tamaño medio (500) o grande (1000). En caso de usar el AF en muestras pequeñas (200) debe tenerse especial cuidado ya que las estimaciones

no serán precisas, salvo que se observe que los ítems son de media o alta calidad (pesos factoriales superiores a 0.70).

El número de indicadores por factor, es un elemento crítico y se recomienda tener un número suficiente de cuatro o más.

10. Considerar validado el instrumento por haber realizado un AFE o un AFC

En primer lugar, hay que recordar que según los Standards, no se validan los test sino las puntuaciones obtenidas en una determinada aplicación del instrumento. Por esta razón en todo nuevo estudio es preciso analizar tanto la fiabilidad como la estructura factorial.

En segundo lugar, el AF en sus variantes (AFE, CFA y ESEM) es una técnica supeditada a la teoría sobre el constructo. Si desde la teoría se determina una estructura factorial mediante AF se puede evaluar si la evidencia empírica permite aceptar el modelo teórico o si se debe rechazar y plantear un nuevo modelo. El hecho de que los índices de bondad de ajuste para una determinada estructura factorial sean buenos, no excluye que haya otras estructuras con mejor ajuste o incluso puede ocurrir que el modelo correcto presente peor ajuste en unos datos concretos. Por ello, es muy arriesgado decidir por una estructura factorial cuando no hay un modelo respaldado por la teoría sobre el constructo o descartar una determinada estructura factorial por la bondad de ajuste en una muestra.

En tercer lugar, como recomiendan los Standards (AERA, APA, NCME, 2014), es preciso que la validación de las puntuaciones se realice a partir de varias fuentes de evidencia. Es decir, además del análisis de la estructura interna mediante AF, deben aportarse otras evidencias, entre las que destacan:

1. El instrumento tiene la propiedad de invarianza (realizar un CFA multigrupo, para poner a prueba que el instrumento mide de forma “justa” a los integrantes de los distintos grupos que forman la muestra: hombres/mujeres, grupos de edad, etc)
2. Relación del test con otras medidas relevantes, otros constructos o criterios de interés.
 - a. Debe aportarse evidencia de que las puntuaciones del instrumento correlacionan con constructos que según la teoría están relacionados.
 - b. Si se utiliza el test para predecir algún criterio (ej. rendimiento) debe mostrarse que las puntuaciones del test correlacionan con la variable criterio.

En consecuencia, una mala práctica que debe conducir al rechazo de un manuscrito en revistas de prestigio, es considerar validado el instrumento por haber realizado un análisis factorial. Para que una medida se considere potencialmente relevante en psicología, debe mostrarse que la medida se ajusta al modelo teórico (ej. mediante AF), que tiene la propiedad de invarianza y que se muestra las correlaciones que la teoría sobre el constructo establece sobre otras medidas de interés.

Nunca un test estará universalmente validado, solo sus puntuaciones en muestra concretas, pero si es posible mediante técnicas de meta-análisis obtener evidencia sobre la generalizabilidad de la medida y de sus relaciones con las medidas de otros constructos y criterios.

Tabla 1. Elementos básicos a tener en cuenta cuando se realiza un Análisis Factorial en un estudio de validación de un instrumento psicométrico

PRÁCTICAS CORRECTAS E INCORRECTAS DEL AF	Si	No	No aplica
Se asume la existencia de una variable latente (constructo psicológico) que se define y enmarca dentro de un modelo teórico.			
Se realiza un ACP en lugar de un AFE para obtener la estructura factorial del constructo, los pesos factoriales y la varianza explicada.			
Previo al AFE se investiga el número de factores a extraer por distintos procedimientos, en especial análisis paralelo			
Se utilizan LOS MISMOS datos para el AFE y el AFC (en lugar de utilizar muestras diferentes).			
Se realiza rotación ortogonal a pesar de asumir que los factores probablemente estarán correlacionados			
Se informa del método de estimación y se justifica su elección. Se hacen análisis del supuesto de normalidad multivariante (ej. Mardia) o se informa de la asimetría y apuntamiento de los ítems.			
En datos ordinales (especialmente, ítems de dos a cuatro alternativas de respuesta) se utiliza como estimador máxima verosimilitud.			
En AFE, se informa de los pesos factoriales en todos los factores			
En AFC se informa de los pesos estandarizados en un gráfico y se informa de todos los índices importantes: chi- cuadrado, RMSEA (CI 90%), CFI, TLI (NNFI) y SRMR			
Se incurre en la práctica incorrecta de eliminar ítems, tanto en AFE como AFC, para continuar los análisis sin tomar nuevos datos para el test “acortado”.			
Se eliminan ítems basándose solo en los pesos factoriales sin evaluar el contenido de los ítems eliminados ni los efectos que pueda tener sobre la integridad de la medida (problema de subrepresentación del constructo).			
Se utilizan muestras pequeñas ($n < 200$) o el número de ítems por factor es inferior a 4 en algún caso.			
Se utiliza como única fuente de evidencia de validez la bondad de ajuste a una estructura factorial, no se hace análisis de la propiedad de invarianza ni se recoge evidencia en relación con otras variables o criterios de interés			
Se utiliza el coeficiente alpha de cronbach para estimar la fiabilidad en escalas multidimensionales y no se informa del omega jerárquico (ver Green & Yang, 2015; Trizano-Hermosilla, I., & Alvarado, 2016).			

Las casillas de color verde indican prácticas correctas y adecuadas del AF, mientras que las casillas de color naranja identifican prácticas erróneas o no recomendadas en la actualidad.

Referencias

- AERA, APA, NCME (2014). Standards for educational and psychological testing. Washington, DC: AERA.
- Aparicio-García, M. E., & Alvarado-Izquierdo, J. M. (2019). Is there a “Conformity to Feminine Norms” Construct? A Bifactor Analysis of two Short Versions of Conformity to Feminine Norms Inventory. *Current Psychology*, 38, (5), 1110-1120. <https://doi.org/10.1007/s1214>
- Antino, M., Alvarado, J. M., Asún, R. A., & Bliese, P. (2018). Rethinking the Exploration of Dichotomous Data: Mokken Scale Analysis Versus Factorial Analysis. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124118769090>
- Asún, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2016). Developing multidimensional Likert scales using item factor analysis: The case of four-point items. *Sociological Methods & Research*, 45(1), 109-133. <https://doi.org/10.1177/0049124114566716>
- Blanco-Canitrot, D., Alvarado, J. M., & Ondé, D. (2018). Consequences of Disregarding Metric Invariance on Diagnosis and Prognosis Using Psychological Tests. *Frontiers in psychology*, 9, 167. <https://doi.org/10.3389/fpsyg.2018.00167>
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.
- 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. <http://dx.doi.org/10.1037/1082-989X.4.3.272>
- Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement: Issues and Practice*, 34(4), 14-20. <https://doi.org/10.1111/emip.12100>
- Ledesma, R. D., Ferrando, P. J., & Tosi, J. D. Uso del Análisis Factorial Exploratorio en RIDEP. Recomendaciones para Autores y Revisores. <https://www.aidep.org/sites/default/files/2019-07/RIDEP52-Art13.pdf>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología/Annals of Psychology*, 30(3), 1151-1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, 50(9), 741.
- Ondé, D., & Alvarado, J. M. (2018). Scale Validation Conducting Confirmatory Factor Analysis: A Monte Carlo Simulation Study With LISREL. *Frontiers in psychology*, 9, 751. <https://doi.org/10.3389/fpsyg.2018.00751>
- Ondé-Pérez, D., & Alvarado-Izquierdo, J. M. (2018). Principios básicos en el proceso de validación de escalas mediante Análisis Factorial. *Variables Psicológicas y Educativas*,

241.

<https://formacionasunivep.com/Vcice/files/libro%20variables%20psicologicas%20y%20educativas.pdf#page=241>

Ovejero-Bruna, M., Brabete, A. C., & Alvarado-Izquierdo, J. M. (2019). Reliability generalization as a seal of quality of substantive meta-analyses: the case of the VIA Inventory of Strengths (VIA-IS) and their relationships to life satisfaction. *Psychological reports*, 122(3), 1167-1188. <https://doi.org/10.1177/0033294118779198>

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements. *Frontiers in psychology*, 7, 769. <https://doi.org/10.3389/fpsyg.2016.00769>