

Revisión del Concepto de Causalidad en el Marco del Análisis Factorial Confirmatorio

A Revision of the Causality Concept in the Confirmatory Factor Analysis Framework

Daniel Ondé Pérez¹

Resumen

En Psicología el Análisis Factorial Confirmatorio (AFC) es ampliamente utilizado en el proceso de elaboración de tests y escalas, siendo una técnica definida formalmente como potencial generadora de modelos causales de medida. No obstante, en numerosos estudios la aplicación del AFC se elabora a partir de diseños de investigación no experimental, en donde muchos investigadores realizan rutinariamente atribuciones sobre los modelos y los instrumentos que van más allá de una perspectiva estrictamente relacional o predictiva. En este trabajo se presenta una revisión del concepto de causalidad desarrollado dentro del marco de los Modelos de Ecuaciones Estructurales (MEE) y del AFC, con varias recomendaciones de carácter teórico y práctico dirigidas a los investigadores aplicados. Se discute sobre el estatus de las relaciones causales en los diseños no experimentales y sobre la necesidad de pensar en términos causales con el fin de potenciar el alcance explicativo de los modelos AFC en Psicología.

Palabras clave: causalidad, análisis factorial confirmatorio, modelos de ecuaciones estructurales, investigación no experimental, validez

Abstract

In Psychology, Confirmatory Factor Analysis (CFA) is widely used in the tests and scales development process, being a technique formally defined as a powerful tool for generating causal measurement models. However, in many studies, the application of the CFA is developed from a non-experimental research design, where many researchers routinely make attributions about the models and instruments that go beyond a strictly relational or predictive perspective. This work shows a review of the causality concept elaborated within the Structural Equation Models (MEE) and the CFA frameworks, with several theoretical and practical recommendations for applied researchers. The status of causal relationships in non-experimental designs and the need to think in causal terms to enhance the explanatory scope of CFA models in Psychology are discussed.

Keywords: causality, confirmatory factor analysis, structural equation modeling, no experimental research, validity

¹ Doctor en Psicología. Profesor Asociado de la Universidad Complutense de Madrid, Facultad de Psicología. Campus de Somosaguas, s/n, 28223, Pozuelo de Alarcón, Madrid, España. Tel.: 913942884. Correo: donde@ucm.es

Introducción

Desde que Jöreskog y Lawley (1967) propusieron el modelo general del AFC, la cantidad de aplicaciones que se han desarrollado en diversas áreas de investigación ha sido enorme, existiendo un abundante número de libros especializados que abordan sus fundamentos y utilidades (por ejemplo, Bollen, 1989; Brown, 2015; Kaplan, 2008; Kline, 2015; Mulaik, 2009; Thompson, 2004), así como una amplia variedad de programas de análisis (AMOS, EQS, LISREL, MPLUS y el paquete lavaan del programa R, entre otros). No obstante, es importante señalar desde el principio que autores como McDonald (1999) describen el AFC como una peligrosa técnica conjetural ya que, de ser usada superficialmente, permite generar mediciones aparentes, aunque posiblemente incorrectas. El problema es, precisamente, que esta técnica recibe con frecuencia un tratamiento superficial en la investigación aplicada (Jackson, Gillaspay, & Purc-Stephenson, 2009; MacCallum & Austin, 2000; McDonald & Ho, 2002; Shah & Goldstein, 2006).

En Psicología, el AFC se suele utilizar como parte del proceso de validación de escalas o instrumentos de medida, siendo su principal misión servir de herramienta para analizar la estructura interna de los datos como fuente de validez de constructo. Aunque no se invoque explícitamente el término causalidad, es habitual encontrar en la mayoría de los estudios que aplican AFC expresiones como “validar”, “confirmar” o “explicar”. Sin embargo, validez y causalidad son conceptos estrechamente relacionados (Markus & Borsboom, 2013), el término confirmar frecuentemente se utiliza de manera errónea puesto que el AFC es más útil como herramienta que permite eliminar o desconfirmar modelos teóricos (Bollen, 1989; Kline, 2015), y la idea de modelo explicativo invoca a un cierto nivel el establecimiento de causalidad. En consecuencia, cuando se aplica AFC como técnica para evaluar la estructura interna de un conjunto de indicadores, el concepto de causalidad está implícitamente presente. No obstante, no conviene idealizar la capacidad explicativa de los modelos AFC, puesto que una de las principales críticas que se puede hacer es que la mayoría de los modelos propuestos en

contextos aplicados descansa sobre diseños de carácter no experimental, con escaso control de variables extrañas, por lo que su alcance no debería ser de tipo causal, sino de carácter relacional-predictivo como máximo, tal vez incluso solo de carácter exploratorio. Utilizar con sentido expresiones del tipo validar o confirmar implica reflexionar sobre conceptos complejos que trata la ciencia, y también sobre las condiciones (diseño de la investigación) en las que se desarrolla cada aplicación. Implica pensar en términos de proyectos de investigación complejos, y no en estudios concretos. Esta reflexión pocas veces aparece reflejada en los estudios publicados, quedando reducida casi exclusivamente a unas pocas líneas cuando se comentan las limitaciones del estudio en el apartado de discusión o de conclusiones.

El concepto de causalidad está recibiendo una atención considerable en los últimos años, como refleja el intento por revitalizar la distinción clásica entre relaciones causales y estadísticas (West & Kotch, 2014), el manual de Mulaik (2009) sobre MEE causales lineales o la publicación en 2009 de la segunda edición del libro de Judea Pearl *Causality: Models, Reasoning, and Inference*. Esta revitalización del concepto de causalidad pone un mayor énfasis en el papel apriorístico de la teoría y del conocimiento sustantivo, en la independencia condicional, y en la evaluación del ajuste local, y resta importancia a la evaluación de los modelos mediante el uso de índices de ajuste global.

Este trabajo se ha realizado en favor de esta corriente revitalizadora. El objetivo es presentar una revisión del concepto de causalidad desarrollado desde la perspectiva de los MEE y del AFC, especialmente a partir de los trabajos de Mulaik (2009) y de Bollen (1989). Además, se presentan también algunas reflexiones sobre el alcance de los diseños no experimentales, planteando que la causación no es cuestión de todo o nada, pudiendo ser asumida (aunque sea débilmente) en estudios no experimentales de carácter explicativo. A lo largo del texto se ha contrapuesto el potencial explicativo del AFC con lo que podríamos denominar, siguiendo a Kaplan (2008), su práctica convencional (o más frecuente) en el marco de la Psicología, que deriva con frecuencia en un mal uso de la técnica. No

obstante, dada su complejidad, la causalidad no es un concepto cerrado, siendo necesario admitir que pueden existir zonas importantes de controversia y de debate. Por esta razón, no se ha pretendido elaborar un decálogo sobre causalidad, sino fomentar una mayor reflexión sobre aspectos básicos y fundamentales del AFC con el fin de favorecer una mayor comprensión del alcance y relevancia de los resultados que se obtienen a partir de datos empíricos en Psicología, y de las conclusiones y generalizaciones que se pueden llegar a establecer.

El texto se ha organizado en ocho puntos. En el primero se describen una serie de distinciones entre MEE, AFC y Análisis Factorial Exploratorio (AFE) que resultan convenientes para situar al lector. Sobre el uso de los términos MEE y AFC, prácticamente todos los conceptos, formalizaciones e ideas desarrolladas en el texto, aunque se han centrado en un tipo de práctica concreta (AFC en Psicología) son equivalentes en ambos casos. En el segundo punto se desarrolla el concepto de causalidad en el marco de los MEE, siguiendo principalmente el trabajo de Mulaik (2009) por ser uno de los autores que más en profundidad han trabajado este tema. En el tercer punto, se describen los requisitos formales para inferir causalidad en los modelos AFC (Bollen, 1989), prestando especial atención a la condición de pseudo-aislamiento. En el cuarto se reflexiona sobre el tipo de proyecto de investigación científica que deben seguir los modelos AFC cuando el objetivo es invocar causalidad. En el quinto se plantea la importancia de comparar entre modelos AFC que compiten por explicar de forma cualitativamente distinta el fenómeno objeto de estudio. En el sexto se describe la práctica convencional en la generación de modelos AFC, y se discuten los límites del carácter confirmatorio de la técnica. En el séptimo se detallan algunas consideraciones acerca del diseño de carácter no experimental de los estudios en los que se suelen elaborar modelos AFC en Psicología, tratando de reconciliar la ausencia de manipulación de variables y de aleatorización con el carácter causal a nivel formal de la técnica. En el octavo se recogen de manera sintética diversas recomendaciones dirigidas al investigador aplicado con el fin de evitar malos usos o prácticas en la aplicación del AFC.

AFC como caso particular de los MEE

Existe la creencia errónea de que el AFC no es más que una versión más potente y flexible del AFE. Aunque el AFE y el AFC son modelos matemáticamente equivalentes, existen importantes diferencias entre ambas técnicas que hacen más adecuado entender el AFC como un caso especial o particular de los MEE (Mulaik, 2009). A nivel conceptual, la principal diferencia es que en AFC el investigador parte de una hipótesis sustantiva acerca del efecto que tendrá un conjunto de variables latentes exógenas (VI o causales) sobre un conjunto de indicadores o variables endógenas (VD), y trata de testar o poner a prueba dicha hipótesis. Esta concepción del modelo factorial es formulada por el investigador antes de recoger los datos, y debe fundamentarse en un conocimiento previo acerca de las VD y de las causales (McDonald, 1999; Mulaik, 2009). Por su parte, la ausencia de restricciones de los modelos AFE no permite poner a prueba de forma genuina hipótesis sustantivas, por lo que los resultados previos de un AFE no deberían ser considerados superficialmente como evidencia de dimensionalidad (por ejemplo, en cuanto al número de factores). En este sentido, una práctica que debería ser abandonada es la aplicación de AFC sobre los resultados de un AFE en un mismo conjunto de datos.

Los MEE son técnicas de análisis multivariadas que se fundamentan en los modelos de regresión, aunque más flexibles (menos restrictivos) al permitir introducir en los modelos el error de medida de las variables (Ruiz, Pardo, & San Martín, 2010). Dentro de los MEE se distingue entre el modelo estructural y el modelo de medida o AFC, unificando en una sola técnica el análisis simultáneo de ecuaciones estructurales y la idea de variable latente del análisis factorial. En su definición formal, tanto los MEE como el AFC se presentan como potenciales modelos explicativos de fenómenos de la realidad, permitiendo poner a prueba teorías e hipótesis de carácter causal (Bollen, 1989; Mulaik, 2009), esto es, permitiendo al investigador intentar ir más allá de la descripción y análisis de patrones de relación entre variables. Tanto es así que los modelos AFC se denominan en ocasiones como modelos causales de medida. Los MEE/AFC tienen

limitaciones propias, si bien se puede entender que en la mayoría de los estudios las limitaciones en términos de razonamiento causal se deben a aspectos no considerados por el investigador, como la ausencia de control de variables extrañas o de confundido (estas cuestiones se abordan más adelante).

Antes de continuar, hay que diferenciar el tipo de causalidad que se puede estudiar en los modelos estructurales y en los modelos de AFC. Mientras que en el primer caso las relaciones de causalidad se establecen hipotéticamente al nivel de las variables latentes (se dice que una variable latente es exógena cuando es causa, mientras que la variable latente que recibe el efecto de otra variable latente se denomina como endógena), en el segundo caso son las variables latentes las que se hipotetizan como causa de los valores obtenidos en las variables observables o indicadores. En realidad, los modelos MEE permiten estudiar ambos tipos de relaciones, puesto que incluyen ambos modelos. Las covariaciones o correlaciones entre factores están fuera del razonamiento causal, ya que no implican direccionalidad entre una posible causa y su efecto.

Concepto de causalidad y MEE

Mulaik (1986, 1993, 2009) presenta una teoría de la causalidad que es central para los MEE (y, por extensión, para el AFC), bajo el argumento de que la causalidad se entiende mejor como relación funcional entre variables. Frente a la concepción empirista clásica de Hume, y que se materializó en el empirismo lógico de la escuela positivista de Viena, actualmente existe una actitud más tolerante con el concepto de causalidad entre los filósofos de la ciencia, fundamentada en el pensamiento contrafactual. Esta tolerancia con el concepto no debe confundirse con falta de rigor, lo que ocurre es que ahora la causalidad se entiende como un concepto más complejo, más difícil de determinar lógicamente. En este sentido, una de las diferencias fundamentales respecto al empirismo lógico es que la causalidad ha dejado de entenderse como un evento binario (si ocurre A entonces ocurre B, y si no ocurre A entonces no ocurre B).

Cuando se trabaja con MEE/AFC, Mulaik considera que es más importante pensar en

términos de variables, frente a nombres o etiquetas de constructos, forzando así a los investigadores a pensar de manera concreta en lugar de abstracta. En otras palabras, cada constructo debe ser una variable, esto es, un conjunto de estados, cualidades o atributos de tal manera que, a cualquier objeto/sujeto que pueda describirse en términos de estos atributos se le puede asignar un solo elemento del conjunto en cualquier momento. Normalmente, los atributos son representados por cantidades numéricas. Por ejemplo, cuando decimos que una persona tiene un CI de 150, el constructo al que hacemos referencia es la inteligencia, la representación numérica de los atributos se realiza mediante un determinado test, y dicha persona tendrá solo dicho valor (no podrá tener un CI de 150 y de 165 simultáneamente). El conjunto de valores que constituyen la variable inteligencia equivale al rango de atributos posibles que presentan las distintas personas en este constructo. Su representación numérica se operativiza mediante la aplicación de un determinado test, el cual no tiene por qué recoger todos los atributos posibles (a diferencia de lo que ocurre con otras variables como la altura de las personas, en donde el rango de atributos posibles puede ser representado numéricamente prácticamente en su totalidad mediante el instrumento de medida).

Inspirado en la idea de Herbert Simons de que las causas determinan distribuciones de probabilidad, Mulaik defiende una conceptualización probabilística de la causalidad. La causalidad debe expresarse entre variables y no entre valores o atributos concretos de las variables. Una relación funcional entre variables es una relación entre dos conjuntos, en donde a cada elemento o miembro del primer conjunto se le asigna uno y solo uno de los elementos del segundo conjunto. En términos de relación funcional, podemos designar la variable Y como una VD o efecto, y la variable X como una VI o causa, en donde un valor y de la variable Y es función del valor x de la variable X ($y=f(x)$). La causalidad probabilística implica que las VI o causas determinan la distribución de probabilidad en la que se manifiestan los valores de la VD o efecto. Una relación funcional causal probabilística entre variables implica, por tanto, que a cada valor de la VI o causa (X) se le asigna

una y solo una distribución de probabilidad del conjunto de distribuciones de probabilidad de la VD o efecto (Y). Pero para establecer causalidad (probabilística), además, se deben dar ciertas condiciones de fondo que establecen dirección y orden causal, mecanismos y conexiones mediadoras, criterios de relevancia, cierre (en el sentido de las operaciones matemáticas, por ejemplo, los números reales están cerrados (*closed*) mediante sumas, pero no mediante restas), estabilidad y (cuando corresponda) la forma de la distribución de probabilidad conjunta del sistema de variables (Mulaik, 2009, pp. 87-89).

Mulaik ha desarrollado el concepto de causalidad probabilística como el marco en el que se pueden entender actualmente las relaciones de causalidad en ciencia, si bien los MEE/AFC pueden entenderse más fácilmente a partir de la definición de relación funcional entre variables del tipo $y=f(x)$.

Requisitos para la inferencia causal en AFC

Bollen (1989) describe los tres requisitos formales que deben cumplir simultáneamente los modelos AFC para poder inferir causalidad: 1) asociación, 2) dirección de la causación y 3) aislamiento (o pseudo-aislamiento).

A partir del análisis de las correlaciones o de las covarianzas entre todos los pares de VD, en AFC la asociación (estadística) entre variables se expresa en forma de relaciones lineales: cargas o pesos factoriales (λ_{ij}) que expresan la relación funcional existente entre las VI (los factores) y los indicadores. Los pesos factoriales representan el cambio que se produce en las VD por cada unidad de cambio en los atributos de las VI (normalmente, la interpretación se realiza a partir de los pesos estandarizados). No obstante, este no es más que el funcionamiento por defecto que tiene la aplicación de cualquier modelo AFC. Basta con introducir unos datos en uno de los programas de análisis existentes y se obtendrá en mayor o menor medida información que pone en relación unas variables con otras, porque estadísticamente las variables latentes pueden ser cualquier cosa. Por tanto, el requisito de asociación, si bien es necesario, es el más débil de los tres.

Siguiendo la máxima de que correlación no implica causación, para ir más allá en el pensamiento causal se requiere del conocimiento del investigador para definir qué es VI y qué es VD, y para identificar las variables extrañas que pueden mediar en la relación funcional estudiada. Las causas deben preceder en el tiempo a los efectos, ya que si dos sucesos son concurrentes no se puede establecer que uno sea causa del otro. Las variables latentes (VI) son variables aleatorias que representan conceptos unidimensionales. Son variables hipotéticas que se asumen teóricamente a un cierto nivel de abstracción. Es decir, se asume que operan en la realidad a pesar de no ser directamente observables, por lo que se utilizan para representar formalmente diferentes conceptos (Bollen & Hoyle, 2012; Borsboom, Mellenbergh, & Heerden, 2003). En este sentido, cuando aplicamos AFC en Psicología se puede asumir a priori que las variables latentes anteceden a las VD, en tanto que serán las causantes de las respuestas de las personas evaluadas a los reactivos. No obstante, este enfoque puede ser incorrecto cuando se trabaja con determinados constructos, en donde puede tener más sentido considerar la relación funcional desde los indicadores hacia el factor. Cuando la relación es del tipo [factor \rightarrow indicadores] se dice que los indicadores del modelo AFC son reflectivos, mientras que cuando es del tipo [indicadores \rightarrow factor] los indicadores se denominan como formativos (ver, por ejemplo, Brown, 2015). Podsakoff, MacKenzie, Lee y Podsakoff (2003) señalan que, desde hace unos años, se está empezando a reconocer que un buen número de los constructos evaluados podrían ser más precisos si se representaran mediante indicadores formativos. Además, la especificación de indicadores reflectivos o formativos en un modelo AFC tiene implicaciones importantes en la interpretación de las puntuaciones del instrumento de medida (Markus & Borsboom, 2013), y la especificación errónea de indicadores reflectivos en lugar de formativos puede derivar en importantes sesgos en la estimación de la relación entre constructos (Law & Wong, 1999). Por todo ello, el uso de indicadores reflectivos o formativos debería ser explicitado y justificado en cualquier estudio en el que se aplique AFC.

En AFC, el pseudo-aislamiento tiene que ver con la capacidad para evaluar el efecto que producen las variables latentes o factores sobre las respuestas de los sujetos a los indicadores en ausencia de otras fuentes de variación. El aislamiento puro es una abstracción salvo en sistemas de ecuaciones determinísticos, ya que no resulta posible tener la absoluta certeza de que una variable latente determinada influya causalmente sobre las respuestas de los sujetos excluyendo toda posible influencia externa (se puede consultar también una revisión del paralelismo entre la condición de pseudo-aislamiento y el supuesto de independencia local en Mulaik, 2009). Por tanto, desde un enfoque probabilístico de la causalidad solamente se pueden asumir condiciones de pseudo-aislamiento (Bollen, 1989; Mulaik, 2009). Tómese como ejemplo la ecuación fundamental del AFC $x_1 = \lambda_{11}\xi_1 + \delta_1$, que representa un modelo de medida simple compuesto por una sola variable observable (x_1) y una sola variable latente o factor (ξ_1). En términos causales se debe asumir que los factores no tenidos en cuenta (δ_1) y la variable latente ξ_1 no están relacionados ($\text{COV}_{\xi_1\delta_1} = 0$), lo que posibilita la condición de pseudo-aislamiento. En otras palabras, el supuesto $\text{COV}_{\xi_1\delta_1} = 0$ es la condición de posibilidad que nos permite asumir la influencia de ξ_1 sobre x_1 en términos de causalidad, asumiendo que dicha influencia no se ve afectada por efectos no tenidos en cuenta (δ_1). No obstante, y por definición, para cada observación en x_1 el término δ_1 interfiere en la relación $\xi_1 \rightarrow x_1$, por lo que el aislamiento es una condición imperfecta y solamente se puede asumir la condición de pseudo-aislamiento.

Para terminar, conviene señalar que una cosa es asumir matemáticamente que $\text{COV}_{\xi_1\delta_1} = 0$ y otra muy distinta es aportar evidencias de que la relación entre VI y VD se analiza en condiciones de pseudo-aislamiento, como veremos en el siguiente apartado. En realidad, se debe operar en sentido contrario, esto es, se podrá asumir que $\text{COV}_{\xi_1\delta_1} = 0$ en la medida en la que se aporten evidencias de que otras posibles causas han sido controladas.

Modelos causales de medida

Formalmente, el modelo general del AFC puede expresarse como un modelo estructural

generalizado a partir de la ecuación $x_1 = \lambda_{11}\xi_1 + \delta_1$, el cual conecta una o más variables latentes, VI o factores ($\xi_1, \xi_2, \xi_3, \dots, \xi_j$) con, al menos, una variable observable o VD (x_{ij}) por factor. En formato matricial, el modelo general del AFC se expresa como $\Sigma_0 = \Lambda_x \Phi \Lambda_x' + \Theta_\delta$, en donde Σ_0 es la matriz simétrica $i \times i$ que contiene las varianzas-covarianzas de las variables observables, Λ_x es la matriz $i \times j$ de cargas factoriales λ_{ij} , Φ es la matriz simétrica $j \times j$ de varianzas-covarianzas entre factores (ϕ), y Θ_δ es la matriz diagonal $i \times i$ que contiene el término error asociado a cada variable observable (δ_i). En la práctica, tras la formalización de un determinado modelo AFC (Σ_0), y asumiendo la existencia de un modelo "real" o poblacional (Σ), se recoge información empírica de las variables observables especificadas en el modelo (S), se estima la matriz Σ_0^* derivada del modelo, y se analiza el grado de discrepancia o de aproximación entre Σ_0^* y S .

Un modelo hipotético no logra captar los procesos del "mundo real" si no es capaz de mostrar consistencia empírica (Bollen, 1989). Por tanto, la mayor virtud de los modelos AFC es la posibilidad de rechazar modelos que son inconsistentes con los datos. No obstante, su mayor debilidad es que el modelo real o verdadero es solamente uno de los modelos posibles que pueden ser ajustados a partir de los datos. En consecuencia, el ajuste del modelo no es suficiente para establecer consistencia con el mundo real. En otras palabras, no es posible asumir Σ a partir de $S \approx \Sigma_0^*$, (Bollen, 1989). Cualquier modelo AFC es una aproximación a la realidad. Los modelos son representaciones formales de teorías acerca de los fenómenos del mundo real, y su construcción se realiza mediante sucesivas aproximaciones hipotético-deductivas a dichos fenómenos. Cada modelo debe ser depurado a través de investigaciones sucesivas, a partir de re-aplicaciones con nuevos datos ($\Sigma_{0,0}^* \rightarrow \Sigma_{0,1}^* \rightarrow \Sigma_{0,2}^* \rightarrow \dots \rightarrow \Sigma_{0,C(0)}^*$), asumiendo que este proceso de investigación deriva en un correcto conocimiento de la realidad (el límite de $\Sigma_{0,C(0)}^*$ cuando $C(0)$ tiende a infinito es Σ). El sufijo $C(0)$ se utiliza en las expresiones anteriores para representar el proceso de investigación en torno a un modelo específico Σ_0^* , reflejando la cantidad de aplicaciones y de avances que pueden

producirse en un contexto determinado. Idealmente, a medida que aumenta $C(0)$ el modelo propuesto se aproxima con mayor seguridad y concreción a la realidad. $\Sigma_{0,0}^*$ representa el modelo inicialmente propuesto, el modelo estimado en sus fases iniciales, que ira ganando y consolidando su consistencia empírica durante las fases del proceso ($C(0)=1, 2, 3, \dots$) o será finalmente desechado. Generalmente, es en estas fases de la construcción del modelo en donde se ponen a prueba los requisitos de asociación y de direccionalidad causal de las variables, y en donde el modelo suele recibir las re-especificaciones más gruesas por parte del investigador respecto a la selección e identificación de variables latentes y observables.

La asunción de pseudo-aislamiento es tentativa en las fases iniciales del proceso, y solamente a medida que aumenta $C(0)$ se pueden acumular evidencias de influencia causal de las variables latentes sobre las observables. Pero, ¿cuánto debe aumentar $C(0)$ para invocar causalidad en el modelo de medida? Idealmente, cuanto mayor sea $C(0)$ mayor aproximación habrá de Σ_0 y Σ_0^* a Σ , si bien este razonamiento, llevado al extremo, resulta poco razonable desde un punto de vista práctico, además de suponer una involución al pensamiento inductivo. Lo importante es entender que cada aumento en $C(0)$, cada nueva aplicación, debe añadir algo que no estaba incluido en los datos originales. Mulaik (2009) incluye algunos ejemplos, como hacer constante el valor de ciertas variables extrañas (por ejemplo, trabajar solamente con determinado rango de edad, a pesar de la pérdida de validez externa), además de exponer la conveniencia en términos científicos de que los modelos sean evaluados por distintos investigadores. También ha argumentado en favor de imponer nuevas restricciones en los modelos a partir de la información que proviene de aplicaciones previas (por ejemplo, fijar el valor de algunas cargas factoriales por factor en base al valor estimado en análisis anteriores). Cada nueva aplicación permite poner a prueba la hipótesis original variando los contextos y el tipo de muestra, al tiempo que se pueden añadir nuevos controles de distintas fuentes de variación y evaluar los cambios que puede producir el paso del tiempo en

las distintas variables (y que, por tanto, pueden afectar a las relaciones funcionales analizadas).

Comparación de modelos

A un nivel de abstracción mayor, la aproximación de modelos teóricos a la realidad también debe tener en cuenta la existencia de modelos alternativos ($\Sigma_1^*, \Sigma_2^*, \Sigma_3^*, \dots, \Sigma_B^*$) que compiten por reflejar Σ de la manera más plausible. El sufijo B de la expresión anterior representa la cantidad de modelos teóricos que tratan de explicar causalmente determinada realidad. Cada uno de ellos trata de obtener el mayor número de evidencias en su propio proceso de investigación (por ejemplo, para Σ_1^* : $\Sigma_{1,0}^* \rightarrow \Sigma_{1,1}^* \rightarrow \Sigma_{1,2}^* \rightarrow \dots \rightarrow \Sigma_{1,C(1)}^*$). Lo deseable en cualquier campo científico es que B sea menor que $C(B)$, aunque con un número suficiente de modelos teóricos compitiendo por explicar un fenómeno determinado y, al mismo tiempo, el mayor nivel de desarrollo y de avance posible en la investigación (el mayor grado de control de fuentes externas). B y C reflejan la asimetría consistencia empírica-realidad, ya que cuanto mayor es C más se concreta el modelo teórico que refleja la realidad, al menos idealmente, mientras que cuanto mayor es B no necesariamente el nivel de desarrollo es mayor.

La comparación de modelos alternativos va más allá de la mera comparación entre modelos anidados con los mismos datos, en donde los parámetros de un modelo son un subconjunto de los parámetros de otro de los modelos analizados (por ejemplo, modelo AFC de un factor, dos factores, tres factores, con y sin correlación entre factores, etc.) en donde lo que se busca simplemente es aquel modelo AFC que tenga un mejor ajuste. Es cierto que la comparación de modelos anidados puede tener cierta utilidad, pero su principal limitación es que no permite identificar ni evaluar la posible existencia de modelos cualitativamente distintos que podrían ser alternativas más adecuadas para explicar los datos (Jackson et al., 2009; MacCallum & Austin, 2000; Shah & Goldstein, 2006). Las alternativas a comparar deben estar fundamentadas a nivel substantivo. De esta cuestión también advirtieron McDonald y Ho (2002) refiriéndose a la generación mecánica de modelos alternativos post

hoc a partir de cambios sugeridos por los índices de modificación. Los modelos alternativos (anidados o no) deberían ser propuestos a priori por el investigador, y las re-especificaciones de los modelos evaluados deberían estar justificadas teóricamente.

El sesgo confirmatorio resulta especialmente problemático en los estudios en los que se aplica AFC (MacCallum & Austin, 2000), puesto que lo razonable es aceptar la existencia de modelos alternativos que pueden tener similares, o incluso mejores, niveles de consistencia empírica que el modelo evaluado ($B=1, 2, 3, \dots$). De hecho, en muchos contextos de investigación en los que se miden y evalúan constructos psicológicos existen varias aproximaciones que implican la posibilidad de utilizar modelos AFC teóricamente distintos, con distinto número de factores, VD e interpretaciones posibles, y que derivan en distintos instrumentos de medida. Por ejemplo, para medir el constructo procrastinación en adultos las escalas más utilizadas son la Escala de Procrastinación General, el Inventario de Procrastinación para Adultos, y la Escala de Indecisión, y entre estudiantes la Escala de Evaluación de la Procrastinación, el Inventario de Procrastinación de Aitken, y la Escala de Procrastinación de Tuckman (Díaz-Morales, 2019). Por supuesto, cualquier investigador puede apostar por una u otra de estas escalas cuando quiere aplicar AFC en un estudio sobre procrastinación. No obstante, si el objetivo es utilizar con propiedad expresiones como proceso de validación, una potente estrategia sería poner a prueba los modelos AFC resultantes obtenidos a partir de la utilización de alguno más de estos instrumentos en el mismo estudio (por ejemplo, dividiendo la muestra en subgrupos por instrumento utilizado). Si el modelo teórico y el instrumento preferido funciona mejor que los modelos alternativos, el investigador tendrá un mayor nivel de evidencia a su favor. Por otro lado, en la elección de los modelos e instrumentos se deberían evitar las justificaciones del tipo “es la prueba más utilizada” o dar por sentada la validez de los estudios previos, y tratar de argumentar (entre otras cosas) por qué se utilizan tanto, o el nivel de desarrollo actual ($C(0)$) en el que se encuentra la medición de los constructos evaluados.

La comparación de modelos no tiene por qué sustentarse exclusivamente en la aplicación de modelos AFC alternativos. Una buena estrategia puede consultarse en el trabajo sobre procrastinación de Díaz-Morales (2019), en donde se hace una revisión crítica del marco conceptual y de los diversos instrumentos de medida utilizados por distintos autores. Tras la aplicación de diversos modelos de análisis por diferentes investigadores, en distintas muestras, todavía se debate sobre si este constructo es unidimensional o multidimensional. Estos trabajos son especialmente relevantes, al permitir situar a los investigadores sobre el nivel de desarrollo que tienen determinados modelos AFC.

Práctica convencional del AFC y carácter confirmatorio

Kaplan (2008) sostiene que los problemas sustantivos conducen idealmente a desarrollos metodológicos y estadísticos más avanzados. Con el tiempo, estos nuevos desarrollos deberían promover un mayor conocimiento sobre los problemas sustantivos y un mayor avance en el desarrollo de teorías (Σ_{B,C^*}). No obstante, para alcanzar un conocimiento más preciso de las asociaciones sustantivas o causales entre variables, los nuevos desarrollos metodológicos deben superar su práctica convencional. Dicha práctica convencional puede resumirse de la siguiente forma: 1) se especifica un modelo AFC (Σ_0) como representación formal de un determinado modelo teórico (Σ); 2) se selecciona una muestra y se recoge información numérica sobre los indicadores; 3) se utiliza algún procedimiento de estimación para obtener los parámetros del modelo (Σ_0^*); 4) se evalúa la consistencia empírica o aproximación entre Σ_0^* y S mediante pruebas e índices de ajuste; 5) en función del resultado de la fase anterior (grado de aproximación) se modifica el modelo mediante algún tipo de re-especificación. Normalmente, este último paso implica una serie de ciclos que consisten en sucesivas modificaciones del modelo hasta que se decide que su consistencia es adecuada, cumpliendo un conjunto de estándares y recomendaciones sobre medidas de ajuste. Por último, una vez se considera que el modelo muestra ajuste suficiente se discuten los hallazgos,

generalmente tratando de confirmar el modelo teórico que se quiere evaluar.

La práctica convencional del AFC dirige su atención directamente hacia la interpretación teórica de los modelos, con confirmación aparente de modelos explicativos (asumiendo su generalizabilidad). No obstante, esta práctica presenta importantes limitaciones en la fase relacional-predictiva y explicativa de la investigación (ver las diferencias existentes entre las etapas de abducción, deducción e inducción; Mulaik, 2019), de gran importancia en la elaboración secuencial de la investigación científica ($C=1, 2, 3, \dots$). Los resultados de un modelo AFC rara vez se utilizan en estudios predictivos. Tampoco se suele prestar demasiada atención al control de fuentes extrañas de variación (condición de pseudo-aislamiento), o al efecto que puede tener el paso del tiempo sobre las variables analizadas y su relación funcional. A pesar de su elevado potencial explicativo, la práctica convencional del AFC omite con demasiada frecuencia la obtención de nuevas evidencias una vez los datos se aproximan a los modelos propuestos, confiando excesivamente en las medidas de ajuste y, generalmente, a partir de un solo estudio (MacCallum & Austin, 2000).

Sobre el carácter confirmatorio del AFC, Sewall Wright desarrolló la herramienta conceptual path analysis con el fin de estimar las magnitudes de los efectos directos cuando el modelo causal verdadero es conocido de antemano (modelos de causas genéticas). No obstante, en ciencias sociales y de la salud el investigador no conoce el modelo verdadero. Por esta razón, desde una perspectiva falsacionista es más razonable asumir que el investigador procede eliminando modelos por falta de consistencia que probándolos o confirmándolos. En este sentido, el AFC es más útil para los investigadores como método para desconfirmar modelos “falsos” ya que resulta prácticamente imposible confirmar un modelo como “verdadero” o “real” (Bollen, 1989; Kline, 2015). Por otro lado, existe la creencia de que el modelo AFC estimado en forma de cargas factoriales (los parámetros estimados que expresan relación funcional entre variables) es lo que se pone a prueba como hipótesis, lo que se quiere confirmar (además del número de factores). Mulaik (2009) ha señalado que lo que hay que

entender es que lo que se pone a prueba en un modelo son las restricciones impuestas por el investigador y no los parámetros libres. Generalmente, el investigador desconoce cuánto valen estos parámetros, por lo que deja que el modelo los estime libremente. Dado que son valores desconocidos, no es posible contrastar hipótesis acerca de ellos (tal vez solamente que son distintos de cero, aunque esta hipótesis es muy débil). En AFC el investigador especifica que ciertas VD reciben cargas factoriales iguales a cero (“la ausencia de flechas” en un path diagram). Son este tipo de restricciones las que realmente se están poniendo a prueba.

Por otro lado, hay que tener en cuenta que diferentes modelos pueden presentar un buen ajuste, siendo solamente uno el modelo “verdadero” (incluso ninguno de ellos). En consecuencia, la cuestión de la consistencia de Σ_0^* con la realidad no puede ser resuelta de manera analítica (estadísticamente) o confirmatoria. Lo que sí puede resolverse confirmatoriamente, mediante la puesta a prueba de hipótesis, es la obtención de ciertas evidencias de consistencia empírica y de pseudo-aislamiento (aumentos significativos de C), como la asociación entre variables y la dirección de la causalidad, el ajuste del modelo replicado en diferentes muestras, la relación con variables criterio, o la invarianza de los parámetros. También hay que destacar la importancia que están cobrando en los últimos años los Modelos de Diagnóstico Cognitivo o CDM, que son modelos de variable latente discreta (ver, por ejemplo, Cáceres & Alvarado, 2019).

AFC en el marco de la investigación no experimental

Los requisitos de asociación, dirección de la causación y pseudo-aislamiento presentan un fuerte paralelismo con los tres requisitos necesarios para inferir relaciones causales en la investigación de carácter experimental (la VI y la VD covarían, la VI precede temporalmente a la VD, y se pueden descartar hipótesis alternativas mediante el control de variables extrañas). Pero, a pesar de este paralelismo, una de las mayores limitaciones que presentan los modelos AFC en Psicología es que suelen desarrollarse dentro de algún diseño de carácter no experimental. En estos

diseños solamente se recoge información acerca del fenómeno objeto de estudio, sin intervención o manipulación por parte del investigador, y sin asignación aleatoria. La manipulación y la asignación aleatoria, características de los diseños experimentales, son las técnicas de control de posibles causas externas más potentes, por lo que su ausencia en el diseño de la investigación supone una seria limitación para inferir pseudo-aislamiento y, por tanto, para invocar causalidad. La pura asociación estadística (asumiendo que existe direccionalidad o antecendencia temporal) puede esconder una relación de tipo causal, o puede ser parcial o totalmente espuria. El problema es que el investigador no tiene forma de saber cuál de estas opciones es la correcta, o la más probable, si no se han realizado los controles necesarios sobre terceras variables o variables extrañas (“the omitted variable problem”; ver Bollen, 1989). Esta situación lleva a pensar a algunos investigadores que el alcance de los diseños no experimentales solo permite establecer asociaciones estadísticas entre variables (asumiendo direccionalidad o no). Desde esta perspectiva, ¿qué sentido tiene plantear la posibilidad de inferir causalidad a partir de modelos AFC en contextos de investigación en los que el requisito de pseudo-aislamiento no se puede cumplir por falta de manipulación o aleatorización? ¿Debería, por tanto, abandonarse el uso del AFC en estudios no experimentales en favor de estrategias más exploratorias como el AFE? El objetivo de este apartado es señalar que asumir que solamente se puede inferir causalidad cuando existe manipulación y asignación aleatoria supone un cierto reduccionismo metodológico (Bollen, 1989; Johnson, 2001; Mulaik, 2009), y que para valorar la potencialidad y alcance de los AFC en el marco de los diseños no experimentales conviene revisar algunas concepciones sobre el diseño de investigación.

Johnson (2001), en su clasificación de los diseños no experimentales, señala en referencia a los estudios de carácter explicativo que no pueden aportar evidencias de causalidad tan fuertes como los diseños experimentales o los cuasi-experimentales. En ningún caso plantea que exista una dicotomía causal – no causal en función del diseño aplicado. Por su parte, Mulaik (2009) ha puesto en entredicho la necesidad de que exista

manipulación de VI para inferir causalidad, argumentando que la causalidad está en la naturaleza de las cosas en el mundo y no solo en las manipulaciones humanas, que también son causas en el mundo. De hecho, parece razonable pensar que un factor o VI asumido como concepto que existe en la realidad puede influir sobre los indicadores (reflectivos) independientemente de la manipulación humana. Mulaik no plantea que la manipulación sea ineficaz, solo que no es infalible. Algo similar plantea Bollen (1989) en relación a la asignación aleatoria, exponiendo mediante ejemplos que no siempre resulta eficaz para potenciar la condición de pseudo-aislamiento. Asimismo, los diseños denominados como cuasi-experimentales se describen como potenciales estrategias para la inferencia causal, a pesar de no ejercer control mediante asignación aleatoria (Shadish, Cook, & Campbell, 2002).

La manipulación y la asignación aleatoria suponen una importante ventaja frente a la investigación no experimental, aunque no son la única manera de potenciar la condición de pseudo-aislamiento. Algunos ejemplos alternativos, menos eficaces pero válidos, son la selección o restricción de la muestra a partir de alguna categoría particular de las variables extrañas, el control estadístico de terceras variables, o la realización de estudios longitudinales. Para potenciar la condición de pseudo-aislamiento, un aspecto fundamental que se debe evaluar es la invarianza, es decir, el grado en el que la medida del constructo se puede generalizar a distintos grupos o submuestras y a distintos momentos temporales (siguiendo la lógica expuesta $\Sigma_{0,C(0)^*}$). Algunas técnicas que permiten evaluar la invarianza o estabilidad de los modelos AFC son los estudios con medidas longitudinales, el análisis multi-grupo, estudios de invarianza parcial, y modelos con múltiples indicadores-múltiples causas o MIMIC (sobre algunos problemas que pueden aparecer en los estudios de invarianza puede consultarse a Millsap, 2005).

Por otro lado, entender que un AFC puede ser parte de un modelo más complejo (MEE) permite al investigador manejar múltiples opciones para potenciar la condición de pseudo-aislamiento. Un ejemplo es el trabajo de Carlson y Mulaik (1993; Mulaik, 2009) sobre cómo las descripciones de

Tabla 1. Principales recomendaciones dirigidas al aumento de la capacidad explicativa de los modelos AFC en contextos aplicados

1. Clarificar cuáles son los objetivos de la investigación para qué se aplica AFC?, evitando justificaciones vagas o retóricas del tipo “con el fin de validar...”. La validación y la inferencia causal son un proceso (Σ_{B,C^*}).
2. Especificar el modelo antes de la recogida de información. El principal objetivo de la especificación es poner a prueba la teoría, no el modelo. Pensar en términos concretos (variables), en lugar de en abstracciones (nombres o etiquetas de constructos). Dotar de un nombre a los factores es una herramienta de trabajo, no una explicación causal (evitar la cosificación o reificación de los factores).
3. Si existen, tener en cuenta modelos alternativos cualitativamente distintos a la hora de explicar el fenómeno objeto de estudio. Evitar la comparación mecánica de modelos anidados con los mismos datos para buscar el mejor ajuste.
4. Describir la fase en la que se encuentra el desarrollo del modelo ($\Sigma_{0,C(0)^*}$) en base a la investigación previa (cuántas veces se ha aplicado, convergencia de resultados, hipótesis puestas a prueba en relación a la condición de pseudo-aislamiento).
5. Justificar las nuevas hipótesis puestas a prueba, valorando el carácter desconfirmatorio del AFC. Discutir las carencias del modelo en términos de pseudo-aislamiento, y plantear a futuro líneas y estrategias concretas de trabajo.
6. No depositar excesiva confianza en la validez de los estudios previos, valorar su alcance con sentido crítico.
7. Valorar la posibilidad de que el paso del tiempo y cambios de contexto produzcan cambios en las variables que afecten a la relación funcional analizada.
8. Especificar al menos 4 variables observables por variable latente o factor con el fin de garantizar su identificabilidad.
9. Considerar cuidadosamente la cuestión de la direccionalidad (medida reflectiva vs. formativa). En este sentido, el Análisis de Componentes Principales (ACP) implica la elaboración de modelos formativos, no interpretables como modelos de variable latente.
10. Las relaciones entre variables no tienen por qué ser lineales, pueden existir efectos curvilíneos y/o de interacción.
11. Justificar adecuadamente el método de estimación empleado en base a análisis preliminares, de carácter exploratorio. La teoría estadística indica que es más apropiado utilizar métodos robustos alternativos a Máxima Verosimilitud (ML) con datos no normales y con datos categóricos.
12. No interpretar estimaciones relativamente elevadas de efectos directos ($\xi_1 \rightarrow x_1$) como evidencia de causalidad, aun cuando las estimaciones de los efectos directos sean elevadas. Es más razonable asumir que los modelos podrían no corresponderse con secuencias causales en el mundo real. Con muestras pequeñas, se pueden producir importantes sobreestimaciones de algunas cargas factoriales.
13. La estimación de casos Heywood u otros valores ilógicos o aberrantes indica problemas de consistencia en el análisis.
14. No confiar excesivamente en las pruebas estadísticas (especialmente en relación a no interpretar la significación estadística como evidencia del tamaño del efecto o de significación práctica).
15. Sopesar el análisis de variables observables altamente relacionadas (por ejemplo, $r_{x_1x_2} > 0,85$). La presencia de colinealidad puede producir inestabilidad en la recuperación de parámetros y afectar gravemente a la estructura del modelo.
16. No interpretar $S \approx \Sigma_0^*$ como una buena aproximación a Σ .
17. No interpretar un buen ajuste del modelo como una buena predicción de las variables observables. En modelos correctamente especificados el ajuste no se verá afectado por la presencia de variables latentes que explican una pequeña proporción de la varianza de las variables observables. En otras palabras, las medidas de ajuste o de aproximación indican hasta qué punto el modelo reproduce la matriz S con precisión, no hasta qué punto la varianza de las variables observables puede ser explicada.
18. La sobreidentificación excesiva e injustificada del modelo mejora artificialmente el ajuste de los datos.
19. No especificar correlación entre errores de medida sin una razón sustantiva.
20. En caso de re-especificar el modelo, no utilizar criterios exclusivamente estadísticos (buscar/analizar fuentes de evidencia teórica, validez de contenido, validez predictiva, etc.).
21. En caso de eliminar variables observables (falta de comunalidad, cross-loading, etc.), repetir el análisis en su ausencia ya que pueden producirse importantes cambios respecto a la solución inicial. El resultado será tentativo hasta que no sea puesto a prueba en nuevas muestras.
22. En caso de especificar cross-loading en algunos indicadores, justificar teóricamente antes de poner a prueba el modelo.
23. Tratar de replicar los modelos en nuevas muestras, y tratar de analizar la relación que tiene el modelo con variables externas.
24. Sería deseable que las publicaciones suministrasen las matrices de varianzas-covarianzas o de correlaciones con las que se han realizado los análisis, así como los parámetros estimados y los errores estándar. La posibilidad de que otros investigadores puedan re-analizar la estructura factorial a partir de los mismos datos es especialmente relevante para evitar sesgos confirmatorios.

personas pueden producir juicios acerca de rasgos de personalidad y estos, a su vez, provocar las valoraciones que hacen los entrevistados sobre estas personas. Estos autores expanden un modelo de medida (AFC) formado por tres variables latentes (amigable, capaz y abierto), elaborando un MEE con dirección causal a partir de los estímulos presentados (variables latentes exógenas) sobre los juicios de los entrevistados (variables latentes endógenas), en donde se aleatorizó la presentación de los estímulos. Este estudio no presenta la forma de un experimento al uso, si bien en él se utilizan potentes herramientas

dirigidas a la inferencia causal. Más adelante, en estudios con nuevas muestras, estos autores aumentaron el número de restricciones sobre el modelo, fijando algunas cargas factoriales al valor obtenido en el estudio inicial, lo que les permitió poner a prueba nuevas hipótesis.

Lamentablemente, la utilización de este tipo de estrategias cuando se elaboran modelos AFC no son frecuentes en la práctica convencional en Psicología. En realidad, en la mayoría de estas aplicaciones el establecimiento de pseudo-aislamiento se encuentra en sus fases iniciales (por ejemplo, en cuanto al número de factores), y

la direccionalidad no tiene por qué estar totalmente especificada a priori (modelos reflectivos vs. formativos). Pero es esta situación, y no la falta de manipulación ni de asignación aleatoria per se, las que dificultan la inferencia causal. En esta situación, hay que reconocer que el alcance de la investigación no debe ir más allá del estudio de los patrones de interrelación entre variables ni de la elaboración de ciertas predicciones, y las conclusiones deben limitarse a las muestras, variables y espacio temporal analizados (MacCallum & Austin, 2000). El elevado grado de control estadístico, así como las posibilidades para potenciar la condición de pseudo-aislamiento colocan a esta herramienta por encima de otro tipo de técnicas de análisis. Pero es misión de los investigadores elaborar todos estos avances respecto a los modelos inicialmente propuestos (aumentos significativos de C), lógica que se puede aplicar también a los experimentos. Es necesario examinar todas las posibles evidencias, ya que tanto la investigación experimental como la no experimental requieren que el investigador aproveche todo su conocimiento para, en la medida de lo posible, considerar y controlar los factores externos que atenten contra la condición de pseudo-aislamiento (Bollen, 1989).

Recomendaciones para la elaboración de modelos AFC

En relación a la condición de pseudo-aislamiento y a la aproximación de modelos AFC explicativos, y frente a la práctica convencional descrita en este trabajo (Kaplan, 2008), existen varias cuestiones y recomendaciones que deberían ser abordadas por los investigadores en la comunicación y publicación de resultados y conclusiones (ver, por ejemplo, Kline, 2015), y que se recogen de manera sintética en la Tabla 1. Por otro lado, no conviene obviar otro tipo de recomendaciones relacionadas con la importancia de realizar análisis preliminares de los datos, de carácter exploratorio, con el fin de justificar adecuadamente el método de estimación de parámetros a utilizar. Por ejemplo, en un buen número de estudios los indicadores se definen a partir de variables discretas (con mayor o menor número de categorías ordenadas y con mayor o menor grado de asimetría), lo que puede impactar

enormemente en los valores estimados de los parámetros del modelo evaluado si no se utilizan estimadores adecuados. Dicho impacto puede adoptar la forma de infraestimaciones o sobreestimaciones importantes de los valores estimados, lo que puede derivar en una interpretación inadecuada de los modelos por elementos ajenos al objeto de estudio (sobre la aplicación de métodos de estimación robustos ver, por ejemplo, Brown, 2015; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Mulaik, 2009; Yang-Wallentin, Jöreskog, & Luo, 2010). Otras cuestiones de interés a valorar en los análisis preliminares son la presencia de valores perdidos (con pérdida de sujetos aleatoria o no aleatoria), la colinealidad, los valores atípicos o extremos, y la potencia del tamaño muestral, entre otros (Kline, 2015).

Discusión

El estatus y utilidad que tiene actualmente el AFC en Psicología resulta confuso. En primer lugar, a nivel formal esta técnica se describe como potencial generadora de modelos causales de medida (Bollen, 1989; Mulaik, 2009), bajo la asunción general de que los factores determinan los valores de los indicadores en condiciones de pseudo-aislamiento (en el caso de medidas reflectivas). En segundo lugar, es frecuente asumir este potencial en la investigación aplicada sin aportar evidencias, lo que ha derivado en una práctica convencional caracterizada por una excesiva generalización de las conclusiones que se pueden alcanzar en un determinado estudio (Kaplan, 2008; MacCallum & Austin, 2000; McDonald & Ho, 2002). En tercer lugar, existe un cierto reduccionismo metodológico que plantea que la ausencia de manipulación y de asignación aleatoria, propia de los diseños en los que se suele aplicar AFC, desaconseja su interpretación a nivel causal (Bollen, 1989; Johnson, 2001; Mulaik, 2009; Shadish, Cook, & Campbell, 2002). En este escenario la aplicación del AFC en Psicología no sale bien parada, bien sea por un exceso de confianza en los modelos (idealización del alcance de los resultados), bien por la imposibilidad de escapar a las condiciones de aplicación con las que se trabaja en referencia al diseño de investigación.

En este trabajo se han abordado varios aspectos y conceptos fundamentales para entender los mecanismos de elaboración de modelos causales de medida. Con ello, se pretende animar a los investigadores aplicados a que reflexionen en términos causales sobre el alcance de los modelos AFC elaborados, dejando de confiar excesivamente en los índices de ajuste obtenidos en aplicaciones transversales. El pensamiento causal es una poderosa herramienta para la generación de conocimiento científico, por lo que potenciar el carácter explicativo de los modelos AFC debería ser uno de los retos a los que hacer frente en los próximos años. Es verdad que la inmensa mayoría de los diseños de investigación en los que se fundamentan los estudios que aplican AFC son de carácter no experimental, y que el tratamiento superficial que se hace de los resultados y conclusiones que se extraen de estos estudios han convertido a esta poderosa herramienta de trabajo en una peligrosa técnica conjetural (McDonald, 1999). En estos casos, y a pesar del estatus explicativo que, pretendidamente o no, reciben los resultados y conclusiones de los modelos AFC, su alcance real solo puede ser de carácter relacional-predictivo, y en relación a las muestras, contextos y momentos analizados (MacCallum & Austin, 2000).

Esta situación se puede mejorar ya que, frente a las técnicas de control características de los experimentos (manipulación de variables y asignación aleatoria), los investigadores pueden poner en marcha numerosas estrategias y técnicas encaminadas a potenciar la condición de pseudo-aislamiento. Lo importante es entender que para ser capaces de invocar causalidad (explicación de fenómenos, validez, ...) se debe superar la práctica convencional y las aproximaciones fundamentadas en un solo estudio (o conjunto limitado de estudios). Potenciar la condición de pseudo-aislamiento en los estudios AFC no es tarea fácil, requiriéndose de proyectos de investigación (Σ_{B,C^*}) más complejos que los actuales, en donde cada estudio añada nuevas hipótesis sobre el control de variables extrañas, en la medida de lo posible, y en donde se pongan a competir realmente distintos modelos teóricos. Esta lógica no es exclusiva de la aplicación del AFC, ni de los diseños no experimentales, sino de cualquier investigación.

Sobre los planteamientos y argumentos desarrollados en este trabajo, una cuestión importante que se puede plantear es la siguiente: ¿cuándo se puede llegar a afirmar que un determinado modelo AFC es un modelo causal de medida? No es sencillo tratar de contestar a esta pregunta, y excede los objetivos de este trabajo, si bien para aproximar una respuesta se plantean dos perspectivas complementarias, una relacionada con la propia elaboración de modelos AFC y otra relacionada con la filosofía y la sociología de la ciencia, y con el concepto de validez. En primer lugar, ¿se puede decir que exista un conjunto determinado o cerrado de condiciones o de controles que, una vez superados, permitan clasificar un modelo AFC como un modelo causal? En principio no. En este trabajo se ha argumentado en favor de elaborar aplicaciones AFC que produzcan aumentos significativos en C, permitiendo así elaborar mejores predicciones y explicaciones de los fenómenos objeto de estudio. Pero si entendemos las distintas fases en la elaboración de modelos como un proceso de investigación complejo (Σ_{B,C^*}), no resulta fácil determinar a priori qué significa “superar” un control determinado, o qué forma debe tener un aumento significativo de C. Así, por ejemplo, si encontramos evidencias sobre la invarianza de los parámetros en un estudio multi-grupo, bajo la hipótesis de que los atributos que definen la variable agrupadora pueden influir sobre la relación funcional entre factores e indicadores, habremos avanzado un paso más hacia la inferencia causal al descartar una fuente extraña de variación. Pero también puede ocurrir que la ausencia de invarianza permita a los investigadores detectar y controlar otras variables no tenidas en cuenta hasta el momento, o replantear su apuesta hacia otros modelos teóricos, tal vez más consistentes como modelos explicativos pasado cierto tiempo.

En segundo lugar, diremos que cualquier modelo de medida que sea considerado como causal solo lo será de manera provisional, puesto que establecer causalidad en ciencia es un proceso que debe hacer frente a los diferentes cambios que se producen en la realidad y en nuestra forma de entender el mundo. También es importante señalar que el establecimiento de causalidad depende del juicio humano, a partir de lo que podríamos

denominar como *expert commites*, así como de la finalidad y consecuencias que tengan los modelos (e instrumentos de medida) elaborados. El área de conocimiento y el tipo de constructo evaluado también son importantes. Por ejemplo, en el campo de la psicopatología se hipotetiza que varios constructos tienen cierta estructura psicobiológica, por lo que el funcionamiento de dicha estructura debería ser capaz de producir cambios en las variables latentes como fuente de evidencia (Borsboom, Mellenbergh, & van Heerden, 2003). En este sentido, la neurociencia y la biología del comportamiento pueden jugar un papel muy importante (Bonifay, Lane, y Reise, 2017). Esta lógica hace referencia a la red nomológica del constructo, en donde se pueden encontrar relaciones no solamente con otros constructos medidos con test y escalas, sino también con conductas, procesos cognitivos, psicobiológicos, etc. (Embretson, 1983). Por todo lo anterior, aunque parece que no resulta posible dar una respuesta cerrada a la cuestión de cuándo un modelo es un modelo causal de medida, lo que parece claro es que todavía queda bastante margen de mejora.

Referencias

- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.
- Bollen, K. A., & Hoyle, R.H. (2012). Latent variables in structural equation modeling. In R.H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 56-67). Guilford Press.
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184-186. <https://doi.org/10.1177/2167702616657069>
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203-219. <https://doi.org/10.1037/0033-295X.110.2.203>
- Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.
- Cáceres, P., & Alvarado, J. M. (2019). Evaluación de la comprensión lectora mediante un modelo de diagnóstico cognitivo. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 51(2), 149-162. <https://doi.org/10.21865/RIDEP51.2.11>
- Carlson, M., & Mulaik, S.A. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research*, 28(1), 111-159. https://doi.org/10.1207/s15327906mbr2801_7
- Díaz-Morales, J. F. (2019). Procrastinación: Una revisión de su medida y sus correlatos. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 51(2), 43-60. <https://doi.org/10.21865/RIDEP51.2.04>
- Embretson, S. E. (1983): Construct validity. Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16, 625-641. <https://doi.org/10.1080/10705510903203573>
- Jackson, D. L., Gillaspay, Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6-23. <https://doi.org/10.1037/a0014694>
- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, 30, 3-13. <https://doi.org/10.3102/0013189X030002003>
- Jöreskog, K. G., & Lawley, D. N. (1967). New methods in maximum likelihood factor analysis. *ETS Research Bulletin Series*, 1967(2), i-24. <https://doi.org/10.1002/j.2333-8504.1967.tb00703.x>
- Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions*. Sage Publications.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.
- Law, K. S., & Wong, C. S. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job

- perception and job satisfaction constructs. *Journal of Management*, 25, 143-160. <https://doi.org/10.1177/014920639902500202>
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51(1), 201-226. <https://doi.org/10.1146/annurev.psych.51.1.201>
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. Routledge.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64-82. <https://doi.org/10.1037/1082-989X.7.1.64>
- Millsap, R. E. (2005). Four unresolved problems in studies of factorial invariance. In A. Maydeu-Olivares, & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 153-172). Psychology Press.
- Mulaik, S. A. (1986). Toward a synthesis of deterministic and probabilistic formulations of causal relations by the functional relation concept. *Philosophy of Science*, 53(3), 313-332. <http://doi.org/10.1086/289320>
- Mulaik, S. A. (1993). Objectivity and multivariate statistics. *Multivariate Behavioral Research*, 28(2), 171-203. http://doi.org/10.1207/s15327906mbr2802_2
- Mulaik, S. A. (2009) *Linear causal modeling with structural equations*. Boca Raton FL, CRC Press, Taylor and Francis Group.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Ruiz, M. A., Pardo, A., & San Martín, R. (2010). Modelos de ecuaciones estructurales. *Papeles del psicólogo*, 31(1), 34-45.
- Shadish, W., Cook, T. D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management*, 24(2), 148-169. <https://doi.org/10.1016/j.jom.2005.05.001>
- Thompson, B. (2004): *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.
- West, S. G., & Koch, T. (2014). Restoring causal analysis to structural equation modeling. review of causality: Models, reasoning and inference, by Judea Pearl: New York, NY: Cambridge University Press, 484 pp. *Structural Equation Modeling*, 21(1), 161-166. <https://doi.org/10.1080/10705511.2014.856700>
- Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, 17(3), 392-423. <https://doi.org/10.1080/10705511.2010.489003>