

Contribución de los Modelos Factoriales Confirmatorios a la Evaluación de Estructura Interna desde la Perspectiva de la Validez

Contribution of the Confirmatory Factor Models to the Evaluation of Internal Structure from the Validity Perspective

Daniel Ondé¹ y Jesús M^a Alvarado²

Resumen

En el presente trabajo se reexaminan algunas de las contribuciones más importantes del Análisis Factorial Confirmatorio (AFC) aplicado a la evaluación de estructura interna. Nuestro objetivo es mostrar la utilidad que tienen estas técnicas para conectar el modelo factorial con el modelo teórico (atendiendo al contenido de los ítems, a la finalidad de los tests, y al uso de las puntuaciones derivadas (total vs subescalas)). En la primera sección caracterizamos el tipo de aplicaciones que aparecen en la literatura científica. En las siguientes secciones a) discutimos las ventajas que tiene la aproximación confirmatoria sobre la exploratoria, b) valoramos las ventajas e inconvenientes del modelo bifactor confirmatorio, y c) introducimos el modelo bifactor S-1 como una prometedora alternativa. En la última sección ilustramos con un ejemplo empírico la utilidad de los distintos modelos AFC examinados en este trabajo.

Palabras clave: análisis factorial, bifactor, bifactor S-1, estructura interna, validez

Abstract

We re-examine some of the most important contributions of Confirmatory Factor Analysis (CFA) to internal structure evaluation. Our main goal is to show the usefulness of these techniques to connect the factor model with the theoretical model (taking into account the content of the items, the purpose of the tests, and the use of the scores derived (total vs subscales)). In the first section, we characterize the type of applications reflected in the scientific literature. In the following ones, we a) discuss the advantages of the confirmatory over the exploratory framework, b) assess the advantages and disadvantages of the confirmatory bifactor model, and c) introduce the S-1 bifactor model as a promising alternative. In the last section, we illustrate with an empirical example the usefulness of the different CFA models examined in this work.

Keywords: factor analysis, bifactor, bifactor S-1, internal structure, validity

Agradecimientos: Esta publicación es parte de los proyectos de I+D+i PID2019-105177GB-C21 y PID2019-105177GB-C22 financiado por MCIN/ AEI/10.13039/501100011033/

¹ Doctor en Psicología. Profesor Ayudante Doctor. Facultad de Psicología, Universidad Complutense de Madrid. Campus de Somosaguas, 28223, Pozuelo de Alarcón, Madrid, España. Tel. : 913942884. Correo: donde@ucm.es (Autor de correspondencia)

² Doctor en Psicología. Catedrático de Universidad. Facultad de Psicología, Universidad Complutense de Madrid. Campus de Somosaguas, 28223 Pozuelo de Alarcón, Madrid, España. Tel.: 913943055. Correo: jmalvara@ucm.es

Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica. RIDEP · Nº66 · Vol.5 · 5-21 · 2022

ISSN: 1135-3848 print /2183-6051online

This work is licensed under CC BY-NC 4.0. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Introducción

La evaluación de la estructura interna de los datos es un aspecto esencial en el proceso de construcción, validación y aplicación de tests psicométricos. En el presente trabajo se revisan algunas de las contribuciones más importantes del modelado estadístico en este campo, fundamentalmente compuesto por técnicas de análisis de la familia factorial (AF) y desde una aproximación confirmatoria. El objetivo de este trabajo es revisar el potencial de estas técnicas estadísticas desde la perspectiva de la validez, entendida como “grado en que la evidencia y la teoría respaldan las interpretaciones de las puntuaciones de los instrumentos para los usos propuestos de los tests” (AERA et al., 2014, p11), con especial énfasis en la evaluación de la fidelidad de la estructura de puntuaciones respecto a la estructura del dominio del constructo (Loevinger, 1957). Nuestra motivación proviene principalmente de nuestro trabajo como psicómetras, y como editores y revisores en revistas científicas.

Frente a la práctica convencional de comparar el ajuste de distintos modelos factoriales, se hace necesaria una mayor reflexión sobre la conexión que tienen estos modelos con la teoría y con los resultados de aplicaciones previas de los instrumentos. Este tipo de conexión implica dos aspectos fundamentales en el proceso de validación. Por un lado, conectar un modelo psicométrico (AF) con un modelo teórico conlleva una reflexión sobre la forma en la que se conceptualizaron los factores, sus interrelaciones y el contenido seleccionado para la recogida de información (i.e., elaboración y selección de ítems). Por otro lado, los modelos teóricos preexistentes no deben consistir solamente en descripciones de los ítems que forman parte del instrumento y de su relación con los distintos factores, es decir, deben incorporar algún tipo de predicción sobre otras variables, y dichas predicciones deben ser sometidas a prueba empíricamente. Además, no hay que perder de vista la finalidad de los instrumentos de medida, aspecto que nos va a permitir delimitar con mayor claridad las estrategias y aproximaciones de análisis a utilizar. Esta última cuestión incide en la importancia de valorar el uso más preciso que se

puede hacer de las puntuaciones derivadas de los instrumentos de medida.

En la primera sección de este trabajo, comenzamos con una revisión de todos los artículos publicados en la *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica* (RIDEP) entre 2018 y 2022. Esta revisión nos ha permitido realizar una caracterización del uso que se hace del AF en la investigación aplicada, así como identificar algunos aspectos que consideramos clave visitar. En base a esta caracterización, en la que podemos encontrar aplicaciones tanto del Análisis Factorial Exploratorio (AFE), del Análisis Factorial Confirmatorio (AFC), o ambas, hemos elaborado el resto de secciones de este trabajo. En la segunda sección discutimos sobre la utilidad que tiene el AFC (frente al AFE) para profundizar en la evaluación de propiedades psicométricas, especialmente en lo relativo al tipo de restricciones que se pueden imponer en estos modelos. En la tercera sección discutimos las ventajas que tiene el modelo bifactor confirmatorio como herramienta de evaluación frente a los modelos AFC clásicos (un factor, factores correlacionados y segundo orden). En la cuarta sección discutimos algunas limitaciones que tiene el modelo bifactor confirmatorio e introducimos el modelo bifactor S-1 confirmatorio, una alternativa que permite superar alguna de estas limitaciones y que puede servir para establecer una conexión más directa entre modelo AF y modelo teórico. Por la novedad del bifactor S-1 respecto al resto de modelos confirmatorios examinados en este trabajo, en la última sección se incluye un ejemplo de análisis con datos reales en el que se comparan los resultados de un modelo de un factor (unifactorial), un modelo de factores correlacionados, un modelo bifactor y un modelo bifactor S-1. Se discuten los resultados en favor del modelo bifactor S-1, mostrando que esta solución factorial puede ser más conveniente en términos estadísticos y tener un mejor encaje e interpretación desde el punto de vista de la validez.

Aplicación del AF: Revisión de Publicaciones en RIDEP (2018-2022)

Para describir las aplicaciones del AF que suelen aparecer reflejadas en la literatura científica, hemos realizado una revisión de todos

Tabla 1. Distribución del tipo de aplicación AF identificado en RIDEP (2018-2022; N=208)

Tipo de aplicación AF (aproximación analítica / modelado estadístico)	Publicaciones		Aplicaciones	
	<i>n</i>	%	<i>n</i>	%
Solo Análisis de Componentes Principales (ACP)	32	15.4	n.e.	n.e.
Solo Análisis Factorial Exploratorio (AFE)	31	14.9	n.e.	n.e.
Análisis Factorial Confirmatorio (AFC)	145	69.7	216	100
Solo un factor (AFC unifactorial)	---	---	63	29.2
Factores correlacionados ^a	---	---	115	53.2
Promedio correlación entre factores <.50	---	---	36	31.3
Promedio correlación entre factores >.50	---	---	66	57.4
No aporta valores de correlación entre factores	---	---	13	11.3
AFC Segundo orden	---	---	22	10.2
AFC Bifactor	---	---	11	5.1
Comparación modelos AFC ^b	---	---	125	57.9
Evaluación de un solo modelo AFC ^b	---	---	91	42.1
AFC-MG (invarianza) ^c	35	24.1	---	---
Validación cruzada (ACP/AFE submuestra 1 y AFC submuestra 2) ^c	42	29.0	---	---

n.e.: No evaluado.

^a% de aplicaciones en función del valor de correlación entre factores (base = 115).

^b% de aplicaciones AFC (base = 216).

^c% de publicaciones en las que se utiliza AFC (base = 145).

los artículos publicados en la RIDEP entre el 2018 y el 2022 (número 46 (volumen 1) a número 64 (volumen 3), ambos incluidos). Esta descripción de tipologías nos ha permitido a) valorar el grado en el que las distintas aproximaciones y técnicas de modelado de la familia factorial se encuentran presentes en los estudios en los que se evalúan propiedades psicométricas de algún instrumento, y b) reflexionar sobre el uso que se hace de dichas técnicas de modelado desde la perspectiva de la validez (i.e., potenciar la conexión entre modelo AF y modelo teórico). Hemos elegido RIDEP por ser una de las revistas científicas sobre evaluación más importantes dentro del mundo hispanohablante, y por tratarse de una revista especializada en la publicación de trabajos empíricos en los que se analizan las propiedades psicométricas de diferentes instrumentos y en diferentes áreas. Los resultados de la revisión de la RIDEP de valoran en los sucesivos epígrafes de este trabajo.

En la primera fase del proceso de revisión se inspeccionaron uno a uno todos los artículos dentro del período de análisis (2018-2022), un total de 268 publicaciones. Los artículos fueron clasificados en grandes tipologías o grupos en función del tipo de aproximación o técnica de modelado utilizada para analizar las respuestas de los participantes a los distintos instrumentos evaluados en cada estudio: Análisis de Componentes Principales (ACP), AFE y AFC (ver

Tabla 1). En total, 208 publicaciones fueron clasificadas dentro de alguna de estas tipologías (*N* válido=77.5% del conjunto total de publicaciones en el período evaluado). El ACP no es un procedimiento para estimar un modelo factorial con error de medida (ver, por ejemplo, Ledesma et al., 2019) y por esta razón esta forma de modelar las respuestas no es objeto de reflexión teórica en este trabajo. El motivo de incluir ACP es que dentro de los estudios que utilizan validación cruzada (29%, ver Tabla 1) en varios casos se aplicó ACP en lugar de AFE en el análisis con la primera submuestra. A nuestro juicio, excluir estos estudios genera sesgos en la magnitud de los resultados. Se incluyeron publicaciones en las que se aplicó Teoría de Respuesta al Ítem (TRI) siempre y cuando incluyeran algún AF (por ejemplo, para testar el supuesto de unidimensionalidad).

En la segunda fase del proceso de revisión clasificamos las publicaciones que incluyen la aplicación de AFC (69.7%) en torno a los siguientes criterios: un modelo AFC de un solo factor (29.2%), modelos de factores correlacionados (53.2%) en función del grado de correlación entre factores (inferior o superior a .50 en promedio, no se informa), AFC de segundo orden (10.2%), y modelo AFC bifactor (5.1%). Estos resultados se calcularon teniendo en cuenta que en una publicación pueden aparecer reflejados los resultados de más de una aplicación AFC (por

ejemplo, cuando se evalúa más de un instrumento o cuando se evalúa más de un modelo factorial). Adicionalmente, se contabilizaron las publicaciones que mostraban evaluaciones de invarianza mediante AFC Multigrupo (AFC-MG; 24.1%).

Esta revisión no tiene como objeto cuantificar las denominadas prácticas cuestionables de investigación o PCI (Wicherts et al., 2016). Por esta razón, hemos preferido no profundizar en los aspectos técnicos del AF, que se pueden consultar en la extensa literatura al respecto (por ejemplo, Brown, 2015), ni tampoco en las prácticas de publicación, para lo que se puede consultar el trabajo de Izquierdo et al. (2014) sobre AFE y el trabajo de Jackson et al. (2009) sobre AFC, entre otros. Tampoco entramos a valorar si la elección de las técnicas (por ejemplo, AFE o AFC) es la más adecuada, por lo que tratamos de mantener nuestra posición lo más neutra posible. Nuestro interés es mostrar el potencial que tienen las técnicas de análisis para conectar el modelo psicométrico (AF) con la teoría y el uso de las puntuaciones de los tests, fomentando así la reflexión sobre la validez.

Ventajas del AFC sobre el AFE en la Evaluación de Estructura Interna

Jöreskog (1969) elaboró un procedimiento general de carácter confirmatorio, que pasó a partir de ese momento a denominarse como AFC, en el que cualquier número de parámetros del modelo factorial se puede fijar de antemano, dejando que se estimen los parámetros “liberados” restantes mediante MV. Más tarde, este procedimiento se generalizó al análisis de variables ordinales y categóricas, especialmente relevante en el análisis de ítems, utilizando como matrices de entrada las de correlaciones policóricas o tetracóricas, y desarrollando o adaptando distintos métodos de estimación (por ejemplo, Muthén, 1984).

Como el propio Jöreskog sugirió, este procedimiento es de interés cuando el/la investigador/a ya ha obtenido una cierta cantidad de conocimiento sobre las variables medidas y, por tanto, está en posición de formular una hipótesis que especifica algunos de los factores involucrados (Jöreskog, 1969, p. 183). En este sentido, frente a las aplicaciones puramente

exploratorias que predominaban en la década de 1960 (“blind factor analysis”; Mulaik, 2009), el desarrollo del AFC supuso un avance muy importante ya que permitía evaluar el patrón hipotético de relaciones entre ítems y factores, implicando más al investigador/a en la toma de decisiones en función de su conocimiento sobre el objeto de estudio y potenciando la generación de modelos explicativos (Pérez, 2020). Por tanto, un primer aspecto a valorar es si actualmente se puede considerar que el AFC se aplica cuando hay modelos teóricos preexistentes y el AFE se aplica de manera puramente exploratoria.

Desde la década de 1960 ha pasado mucho tiempo y el número de aplicaciones del AFE de tipo “blind” es menos frecuente. Por ejemplo, entre los artículos revisados de la RIDEP, encontramos algunas aplicaciones del AFE puramente exploratorias cuando el instrumento es de nueva creación, en donde se desconoce el número óptimo de factores sobre el que hipotetizar la estructura interna de la medida. También encontramos aplicaciones (la mayoría) en las que se fuerza la solución a un número concreto de factores y el tipo de interrelación entre factores (i.e., método de rotación) en función de modelos teóricos preexistentes, en general en situaciones en las que el instrumento de medida ya ha sido aplicado en estudios previos. Por tanto, se puede aplicar AFE puramente exploratorio, orientando la toma de decisiones de manera heurística, o se puede aplicar AFE con una orientación más sustantiva. La diferencia fundamental entre una aplicación y otra radica en el conocimiento existente en cuanto al número de factores que se deben extraer.

Dentro de las aplicaciones del AFC, un 29% aplican procedimientos de validación cruzada en los que se divide la muestra en dos submuestras, aplicando ACP/AFE en la primera y AFC en la segunda. No en todos los casos se pone a prueba un modelo teórico, aunque no es lo más frecuente. Por ejemplo, a veces se decide el número de factores mediante ACP/AFE en una submuestra y después se especifica el modelo AFC resultante del primer análisis y se evalúa en la segunda submuestra. Por otro lado, tanto si se sigue un procedimiento de validación cruzada como si no, en el 57.9% de las aplicaciones AFC se ponen a prueba distintos modelos (en mayor medida

unifactorial y factores correlacionados, en menor medida de segundo orden y bifactor). Este tipo de aplicación suele ser un tanto exploratoria, buscando el modelo que mejor ajusta de varios posibles, aunque la aproximación analítica sea confirmatoria y la evaluación gire en torno a algún modelo teórico preexistente evaluado en estudios previos. También en las aplicaciones AFE es relativamente frecuente encontrar varios modelos factoriales compitiendo entre sí, aunque la evaluación se suele centrar en los valores de las cargas factoriales en términos de estructura simple. De lo anterior se desprende que, en cierta medida, en cuanto a la toma de decisiones orientada por modelos teóricos y herramientas preexistentes, las aplicaciones AFE y AFC no difieren mucho. El elemento diferencial sigue siendo si la aplicación corresponde a herramientas de nueva creación y a la falta de orientación desde una teoría. Los extremos de este continuo podrían definirse entre los instrumentos de nueva creación y la aplicación AFC de un solo modelo en la que se apuesta por una propuesta teórica concreta. Por ejemplo, en un 29.2% de las aplicaciones AFC revisadas en RIDEP se pone a prueba un modelo unifactorial exclusivamente (esto es, sin comparar con otros modelos), orientado por un modelo teórico que plantea la unidimensionalidad estricta de la medida.

¿Dónde radica entonces el avance del AFC sobre el AFE? En el tipo de operaciones y de restricciones que se pueden establecer con esta técnica, lo que permite profundizar en la evaluación de propiedades psicométricas y, por consiguiente, en la reflexión sobre la validez. AFC permite obtener errores típicos de estimación y realizar pruebas de significación de los parámetros estimados, incluyendo las cargas factoriales, las cargas cruzadas, la correlación entre los términos error de los ítems y la correlación entre factores. También permite realizar pruebas de significación de las varianzas de los factores (este punto se desarrolla más adelante). Por otro lado, como señala Brown (2015), AFC permite evaluar tau-equivalencia en modelos unifactoriales y obtener la fiabilidad de las escalas a partir de la solución factorial (frente a alfa de Cronbach, ver Trizano-Hermosilla & Alvarado, 2016), realizar análisis multigrupo (AFC-MG) para evaluar la invarianza factorial,

analizar covariables (modelos MIMIC) y evaluar el poder predictivo de las variables latentes, entre otras posibilidades. Todo lo anterior implica imponer algún tipo de restricción sobre el modelo AFC, por lo que tradicionalmente se considera al AFC como una técnica más flexible y potente que el AFE.

Ahora bien, actualmente existen desarrollos dentro del AFE que permiten abordar varios de estos aspectos técnicos habitualmente tratados mediante AFC. Por ejemplo, actualmente el programa Mplus permite obtener errores típicos de estimación con los que elaborar tests de significación estadística sobre los parámetros estimados (también el programa R, mediante el paquete EFAutilities). Otro ejemplo son los procedimientos denominados como EFA Trees, propuestos por Sterner y Goretzko (2022) para evaluar invarianza. No obstante, conviene señalar que estos desarrollos son más recientes, con un corpus de textos técnico bastante más limitado que el que se encuentra disponible para el AFC. En este sentido, en la revisión de la RIDEP no hemos encontrado aplicaciones que utilicen este tipo de técnicas. Además, algunos desarrollos (por ejemplo, EFA Trees) se encuentran en plena fase de desarrollo en el momento de escribir este trabajo. Desde nuestro punto de vista, si bien existen desarrollos técnicos prometedores dentro del AFE, el modelado mediante AFC es a día de hoy la técnica a recomendar si se quiere profundizar en las propiedades psicométricas de los instrumentos, poniendo a prueba la estructura subyacente a los datos e introduciendo y poniendo a prueba nuevas restricciones sobre el modelo inicial.

Existen otras aproximaciones denominadas como intermedias (entre el AFE y el AFC) que pueden ser de utilidad para conectar modelo psicométrico y modelo teórico (Brown, 2015), como Análisis Factorial Exploratorio/Confirmatorio y Modelos Exploratorios de Ecuaciones Estructurales. Valorar estas técnicas excede los objetivos de este trabajo. Baste señalar que al tratarse de aproximaciones intermedias incluyen elementos formales del AFC por lo que pueden admitir formalmente el tipo de restricciones que estamos examinando en este trabajo. Estas aproximaciones cuentan con más tradición que las que

comentábamos más arriba sobre el AFE, aunque no parecen estar todavía tan implementadas en la investigación empírica como el AFC.

Nos centramos ahora en el análisis de invarianza (AFC-MG) y en el poder predictivo de los modelos AF por su relevancia desde el punto de vista de la evaluación de estructura interna. En las publicaciones de la RIDEP que aplican AFE es frecuente encontrar análisis estadísticos de diferencias de grupo en las puntuaciones de cada factor mediante T de Student o ANOVA, procedimientos que no son equiparables al análisis de invarianza mediante AFC-MG. No obstante, hay que tener en cuenta que no siempre se publican resultados relacionados con AFC-MG. Solamente el 24.1% de las publicaciones revisadas que aplicaron AFC mostraron el resultado de algún AFC-MG. Por otro lado, pueden existir escenarios de evaluación en los que no sea factible la equivalencia métrica entre diferentes grupos por cuestiones sustantivas relevantes (por ejemplo, falta de invarianza en adolescentes debida a efectos madurativos). Una alternativa ante la falta de invarianza puede ser la aplicación de modelos MIMIC (Brown, 2015) para el estudio de las diferencias inter-grupo como potencial covariable. En estos modelos se utilizan las variables latentes y los ítems como regresores de las covariables (i.e., variable agrupadora que se utilizaría en el AFC-MG). Efectos directos sobre los ítems estadísticamente significativos reflejarían falta de invarianza, si bien permiten estudiar el potencial efecto moderador o mediador de la covariable de forma más directa que con un AFC-MG.

En general, tanto si se aplica ACP/AFE como si se aplica AFC, el siguiente paso en la evaluación del modelo factorial es la correlación de las puntuaciones de las escalas y/o subescalas con otras variables, valorando alguna forma de validez predictiva (en la RIDEP lo más frecuente es valorar validez concurrente, convergente y discriminante, existiendo pocas publicaciones en las que se valore la validez referida a criterio). Ahora bien, como señalan Blanco-Canitrot et al. (2018) hay que tener en cuenta que la falta de invarianza métrica puede tener consecuencias negativas sobre la validez predictiva (i.e., relación con otras variables). Por esta razón, parece necesario combinar resultados sobre invarianza con resultados de validez predictiva en una misma

publicación. Si no se informa de invarianza métrica, los resultados de validez predictiva solo pueden ser considerados de forma provisional, cuestión que debería quedar claramente reflejada en la sección de discusión de resultados. En caso de falta de invarianza se puede combinar un modelo MIMIC con un modelo de regresión que incluya las variables a predecir objeto de estudio con el fin de controlar las diferencias encontradas a nivel de grupo (Brown, 2015). No hemos encontrado ningún estudio en la RIDEP que realice este tipo de aplicación.

Por último, aunque no es un procedimiento recomendado aplicar AFC sobre los resultados de un AFE previo con la misma muestra (i.e., no realizar validación cruzada), puede tener sentido evaluar primero la estructura mediante AFE y especificar después un AFC para valorar el tipo de resultados en mayor profundidad, gracias a las restricciones impuestas (por ejemplo, aplicando un AFC-MG). Lo que debería quedar claro en las publicaciones es este objetivo de carácter técnico-estadístico cuando se justifica el uso combinado de AFE y AFC, y no argumentar (erróneamente) que la solución AFC confirma o aumenta el estatus de validez de la estructura evaluada respecto a la solución AFE previa.

Contribución del Modelo Bifactor Confirmatorio

El modelo bifactor fue propuesto por Holzinger y Swineford (1937) en el contexto de la teoría clásica de la inteligéncia, y aparece descrito y desarrollado en profundidad en varios trabajos relevantes (Chen et al., 2006; Maydeu-Olivares & Coffman, 2006; Reise, 2012). En este modelo cada variable observada o ítem depende de dos factores: un factor general y un factor más pequeño (group factor) que caracteriza un subconjunto específico de ítems (ver Figura 1d). Una característica de estos modelos es que todos los factores (general y group factors) no pueden estar correlacionados. En las últimas décadas ha crecido el interés por el modelo bifactor gracias a esta forma que tiene de descomponer las respuestas a los ítems. En las publicaciones de la RIDEP que aplican AFC encontramos un 5.1% de modelos bifactor confirmatorios (AFC bifactor).

El modelo bifactor permite evaluar si las respuestas obtenidas en un determinado instrumento de medida son esencialmente

unidimensionales. El término unidimensionalidad esencial se utiliza para definir estructuras en las que domina el factor general (elemento de varianza común a todos los ítems) en presencia de cierto grado de multidimensionalidad reflejada en las distintas facetas de contenido o group factors (Reise, 2012). La lógica es la siguiente: en contextos aplicados es poco probable encontrar modelos que sean puramente unifactoriales (unidimensionalidad estricta, modelos en los que las respuestas a los ítems se deban a un solo factor; ver Figura 1a), y es poco probable también encontrar modelos que sean estrictamente multidimensionales, en donde las correlaciones entre factores sean nulas (lo habitual es encontrar modelos como el de la Figura 1b, asumiendo cierto grado de relación significativa entre los factores). La utilidad del modelo bifactor radica en su capacidad para evaluar simultáneamente unidimensionalidad en presencia de multidimensionalidad, aspecto que no puede ser abordado mediante modelos AFC clásicos como el unifactorial o el de factores correlacionados. El modelo AFC de unifactorial presenta en numerosas ocasiones términos error correlacionados, lo que implica una fuente de variación no especificada por el modelo y, por tanto, se puede alejar sensiblemente de la unidimensionalidad estricta que pretende evaluar. En el caso del modelo AFC de factores correlacionados está el problema de las cargas cruzadas significativas, lo que refleja en cierta medida que la multidimensionalidad no es perfecta ni unívoca, aspecto que no se puede evaluar directamente interpretando la correlación entre factores. Todas estas cuestiones se pueden detectar mediante los índices de modificación (IM), si bien no permiten evaluar unidimensionalidad esencial, además de tener el peligroso potencial de plantear re-especificaciones atóricas en los modelos para mejorar el ajuste (McDonald & Ho, 2002).

Existen índices que permiten evaluar el grado de unidimensionalidad esencial (Reise et al., 2013). Entre los índices más utilizados están la varianza común explicada (VCE) y el coeficiente omega jerárquico (ω_H). La VCE refleja la varianza explicada por el factor general descontando la varianza que se debe a los group factors. ω_H es una medida de fiabilidad que controla la presencia de group factors, a diferencia de α de Cronbach o

de ω (McDonald, 1999). Para asumir unidimensionalidad esencial (i.e., presencia de un factor general fuerte) se recomienda obtener valores de VCE superiores a .60 y de ω_H superiores a .70 (Reise et al., 2013). Conviene señalar que estos índices también se pueden calcular para los group factors, indicando el porcentaje de varianza explicada por cada uno de ellos (VCE_S) y la fiabilidad de cada subescala (ω_S) controlando la presencia del factor general.

Centrándonos en el objetivo de este trabajo, a nivel de modelado el bifactor es una excelente herramienta de evaluación psicométrica que nos va a permitir valorar hasta qué punto es más adecuado utilizar una sola puntuación total (asumiendo unidimensionalidad esencial) o es preferible utilizar puntuaciones de subescala. En las publicaciones de la RIDEP, por regla general, tanto si los estudios aplican ACP/AFE como si aplican AFC, la fiabilidad de los instrumentos evaluados se calcula a partir de α de Cronbach y, en menor medida, de ω . Una vez evaluada la estructura interna mediante AF, llama la atención que en muchos de estos estudios se calculan correlaciones con otras variables tanto para la puntuación total como para la de las subescalas. Utilizar una puntuación u otra no debe ser una cuestión arbitraria y aquí el modelo bifactor es donde puede ser de mayor utilidad gracias a la obtención de índices como VCE, VCE_S , ω_H , y ω_S . Una vez decidido cuál es el uso más adecuado de las puntuaciones (a nivel total o a nivel de subescala), lo recomendable sería utilizar las puntuaciones factoriales correspondientes en el análisis correlacional posterior.

¿Cuándo puede ser conveniente aplicar el modelo bifactor con este objetivo de evaluación psicométrica? La aproximación más frecuente es cuando existe un alto grado de correlación entre factores. En las publicaciones que aplican AFC de factores correlacionados, un 57.4% de los modelos evaluados presentan niveles promedio de correlación entre factores superiores a .50 (es frecuente encontrar correlaciones por encima de .70 – .80, incluso superiores a .90). Este escenario es prototípico de una evaluación mediante bifactor, ya que estos niveles de correlación entre factores pueden reflejar la existencia de un factor general fuerte, lo que puede implicar la recomendación de

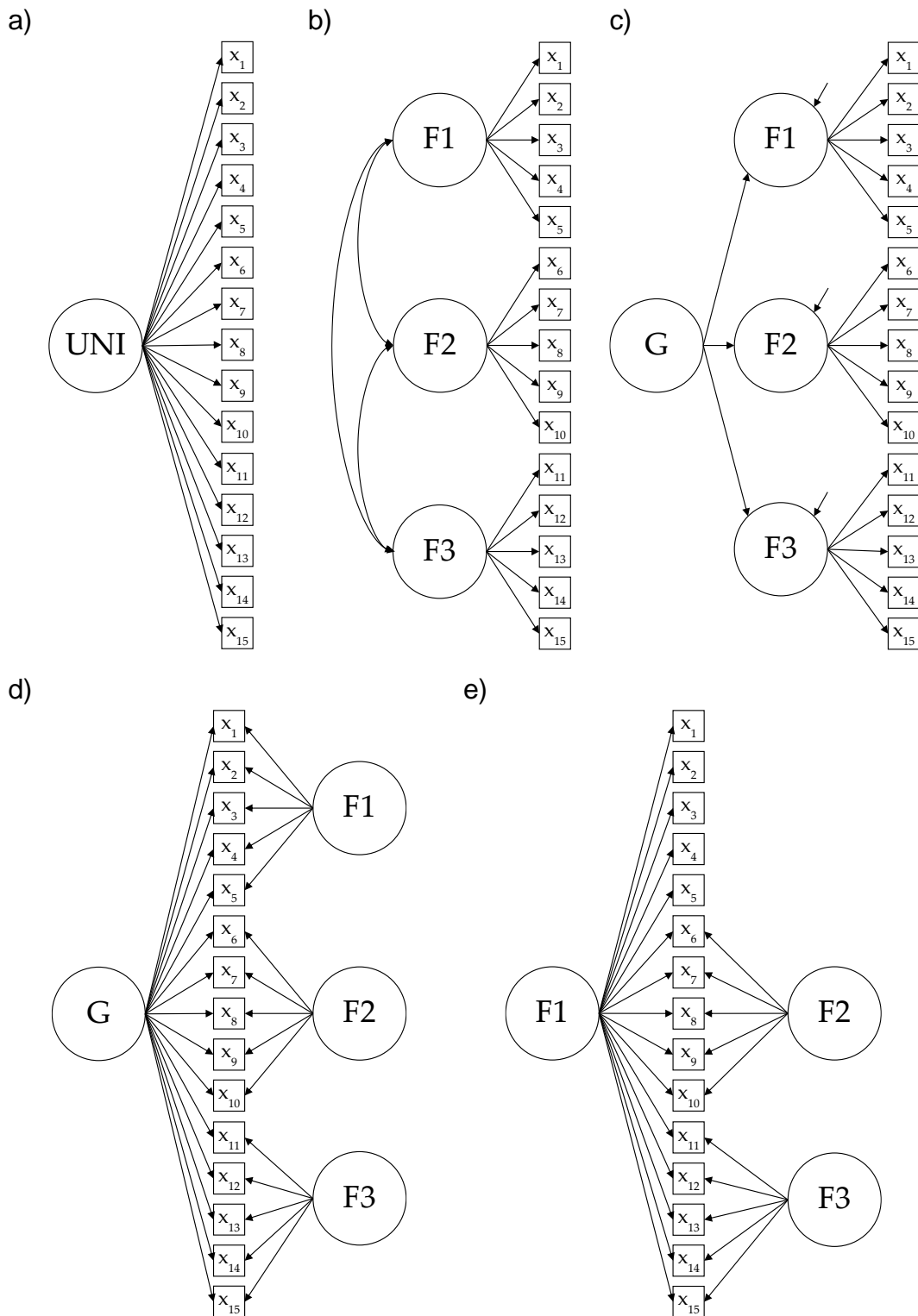


Figura 1. Esquema de representación (path diagram) de los modelos AFC clásicos y bifactor.

Nota. Modelos AFC a) unifactorial, b) factores correlacionados, c) segundo orden, d) bifactor y e) bifactor S-1. Por simplicidad, se omiten las flechas de los términos error.

utilizar una sola puntuación factorial total bajo la asunción de unidimensionalidad esencial en lugar de las puntuaciones de las subescalas. Esto no significa que el modelo subyacente deba ser un modelo unifactorial, ya que las facetas o group factors (multidimensionalidad) pueden tener sentido desde

el punto de vista del contenido de los ítems y del modelo teórico preexistente. En definitiva, el modelo teórico puede ser unifactorial, de factores correlacionados o, incluso, bifactor, pero la evaluación de la unidimensionalidad esencial nos sirve para valorar cual es el mejor uso que se

puede hacer de las puntuaciones del instrumento. Por tanto, el modelo bifactor, entendido como herramienta de evaluación psicométrica, puede ser una herramienta estadística complementaria al modelado AFC clásico, y no un modelo competidor en términos de ajuste estadístico o en términos teóricos.

En un 10.2% de las aplicaciones AFC revisadas se han analizado modelos de segundo orden, generalmente comparando el ajuste con otros modelos como el de factores correlacionados. Los modelos de segundo orden suponen una aproximación AFC clásica a la evaluación de un factor general (factor de segundo orden) y su relación con factores de varianza residual (i.e., varianza no explicada por el factor general) o factores primarios, por lo que pueden ser una buena alternativa a los modelos bifactor cuando la correlación entre factores es elevada.

No obstante, el modelo bifactor presenta ciertas ventajas que lo pueden hacer más aconsejable como herramienta de evaluación. En primer lugar, en los modelos de segundo orden cada ítem recibe un efecto directo de los factores primarios y un efecto indirecto de los factores del factor o factores de segundo orden (ver Figura 1c). En el caso del modelo bifactor, los ítems reciben dos efectos directos (factor general y group factor), lo que resulta más fácil de interpretar. En segundo lugar, los modelos bifactor son más tratables matemáticamente que los de segundo orden, lo que facilita el análisis de invarianza y del poder predictivo de los modelos (para profundizar en estas cuestiones, se puede consultar el trabajo de Chen et al., 2006).

Existe otro tipo de aplicación del modelo bifactor que puede ser de interés cuando el nivel de correlación entre factores es bajo o moderado (inferior en promedio a .50, aproximadamente, como ocurre en el 31.3% de las aplicaciones del modelo AFC de factores correlacionados identificadas en la RIDEP; el 11.3% restante no aporta este tipo de información). Este escenario se aleja de la unidimensionalidad esencial en favor de la multidimensional, y el interés puede estar en obtener valores solventes para las subescalas (VCE_S y ω_S), así como obtener puntuaciones factoriales de subescala más depuradas una vez se ha controlado la varianza común a todos los ítems (i.e., factor general). Este tipo de aplicación es

más novedosa y está recibiendo atención, por ejemplo, en contextos en los que se asume la presencia de un factor general de engaño en las respuestas (Hendy et al., 2020).

Un comentario final sobre el modelo bifactor. En ocasiones, este modelo se utiliza para evaluar modelos teóricos que se hipotetizan como jerárquicos (existencia de un factor general teórico). No obstante, este tipo de aplicación no está exenta de fuertes críticas, sobre todo fuera del contexto de evaluación de la inteligencia y de los instrumentos cognitivos o de rendimiento máximo. Por ejemplo, se han realizado importantes críticas en el campo de la personalidad (Revelle & Wilt, 2013). Excede el alcance de este trabajo profundizar sobre este aspecto, además de que este tipo de justificaciones teóricas corresponden a los y las expertos/as en cada área de evaluación. Baste señalar que ajustar un modelo bifactor no es suficiente como para establecer una hipótesis teórica de factor general. Además, de todos los modelos que estamos valorando, el modelo bifactor es el que libera mayor número de parámetros, ya que para cada ítem se estiman dos cargas factoriales. Esta situación deriva en una fuerte tendencia al sobreajuste de la solución factorial una vez se estima el modelo bifactor. En otras palabras, liberar tantos parámetros hace que el modelo bifactor ajuste mejor que los modelos unifactorial, de factores correlacionados e, incluso, de segundo orden. Como señaló McDonald (1999), valorar de manera superficial este tipo de resultados puede convertir el modelado estadístico en una peligrosa técnica conjetural.

Facetas como Factor General de Referencia: el Modelo Bifactor S-1 como Alternativa al Modelo Bifactor

El modelo bifactor S-1 surge como un caso especial de modelo AFC bifactor en el contexto de evaluación multirasgo-multimétodo (Eid, 2000; Maydeu-Olivares & Coffman, 2006), extendiéndose años más tarde a otros contextos, por ejemplo, en la evaluación de la inteligencia (Eid et al., 2018). Esta forma de modelar las respuestas a los ítems consiste en especificar uno de los group factors como factor general de referencia. En la Figura 1e se muestra un ejemplo de especificación S-1. Uno de los group factors, el

factor 1 (F1) en este caso, se conceptualiza como factor general de referencia, con cargas factoriales en todos los ítems. Este factor general no es equivalente al del modelo bifactor (factor G, Figura 1d), no es un modelo bifactor parcial en el que desaparece una faceta o group factor por falta de varianza común (Chen et al., 2006), sino que puede servir para conectar de forma más directa la estructura interna de los datos con la fase previa de redacción y selección de ítems en determinados instrumentos. Por ejemplo, el instrumento Trait Meta-Mood Scale (TMMS-24) fue conceptualizado inicialmente a partir de tres factores relacionados con la autopercepción de las emociones: atención, claridad y reparación. Ondé et al. (2021) aplicaron un modelo S-1 con los datos recogidos mediante la TMMS-24, bajo la hipótesis de que el factor de atención podía servir como factor general de referencia (parece razonable pensar que se debe producir un cierto nivel de atención a las emociones para que los factores de claridad y reparación tengan sentido). La alternativa bifactor partiría de la hipótesis de que existe un factor general de inteligencia emocional (*e*-factor), en donde la atención sería una de las facetas de contenido. Esto supondría especificar en un mismo modelo un factor general (*e*-factor) y un group factor que es de carácter general (atención), lo que puede resultar difícil de justificar. Cabe suponer que la fase de redacción y selección de ítems no se desarrolló teniendo en mente un modelo bifactor. Parece más claro que sí se conceptualizó parte del contenido como un elemento atencional de carácter general, lo que conlleva (en este caso) una mayor conexión del modelo S-1 con el modelo teórico propuesto inicialmente y con el proceso de elaboración del instrumento.

Por otro lado, el modelo S-1 se ha propuesto como una alternativa al modelo bifactor por presentar ciertas ventajas. En primer lugar, si bien también libera más parámetros que los modelos clásicos, es un modelo más restrictivo que el bifactor. Esto implica que el ajuste de estos modelos es algo más realista y, por tanto, puede tener mayor utilidad práctica. En segundo lugar, en una revisión de publicaciones entre 2013 y 2014 en la que se evaluaron 82 estudios en los que se aplicó AFC bifactor, se encontró que en un 61% de los casos se obtenían resultados anómalos

(Eid et al., 2017). De hecho, los autores de esta revisión consideraron que este porcentaje supone un límite inferior, dado que en buena parte de las aplicaciones faltó información necesaria para realizar la valoración. Entre los resultados considerados como anómalos, de especial interés es la aparición de problemas de identificación de los modelos, la presencia de casos Heywood en forma de varianzas negativas y la aparición de factores con varianza estadísticamente nula. En las publicaciones revisadas de la RIDEP en las que se aplica bifactor (5.1%) encontramos un 72.7% de resultados anómalos. Solamente en una publicación se informa de la aparición de una varianza negativa, siendo el resultado anómalo más frecuente la aparición de group factors con escasa varianza explicada. Esta última valoración se ha realizado a partir de las cargas factoriales relacionadas con los group factors, encontrando casos con falta de significación estadística en la mayoría de los ítems que forman un determinado group factor o, simplemente, cargas factoriales bajas o próximas a cero. En AFC, para valorar si la varianza de los factores es estadísticamente significativa se debe fijar la métrica de los factores fijando a uno el valor de alguno de sus ítems. Esta técnica suele ser la que se utiliza por defecto (Brown, 2015), frente a fijar la varianza del factor a uno, pero no se suele informar del resultado estadístico (*p*-valor) relacionado con las varianzas de los factores, de especial importancia cuando se aplican estos modelos.

También se han planteado otras ventajas del modelo bifactor S-1 relacionadas con el análisis de invarianza y del poder predictivo de los modelos (Eid et al., 2018). El modelo bifactor presenta más resultados anómalos de los que puede parecer en un principio (por ejemplo, casos Heywood), pero aun así es un modelo que puede obtener adecuados valores de ajuste, incluso los mejores al comparar con otros modelos. El ajuste puede enmascarar la presencia de anomalías, y esto tiene un impacto negativo cuando se analiza invarianza mediante AFC-MG o la relación con otras variables, ya que se arrastran los problemas iniciales. El modelo S-1, al imponer un mayor número de restricciones (no especifica uno de los group factors), previene la aparición de estos problemas, por lo que resulta de mayor utilidad en estas situaciones como herramienta de evaluación

psicométrica para profundizar en los análisis relacionados con estructura interna. No hemos encontrado ninguna aplicación del bifactor S-1 en las publicaciones revisadas de la RIDEP.

Un Ejemplo Práctico: Evaluación de la Estructura Interna del Metacognitive Awareness of Reading Strategies Inventory (MARSÍ)

A continuación, ilustramos la utilidad del modelo bifactor S-1 a partir de una aplicación sobre un conjunto de datos recogidos mediante el Metacognitive Awareness of Reading Strategies Inventory – MARSÍ (Mokhtari & Reichard, 2002). El MARSÍ es un instrumento de autoinforme compuesto por 30 ítems y 5 categorías de respuesta diseñado para evaluar el conocimiento de las estrategias de lectura de estudiantes de 11 a 18 años. El MARSÍ se conceptualizó a partir de los tres tipos de estrategias de lectura siguientes: estrategias de lectura global (GRS – 13 ítems, enfocadas a un análisis global del texto), estrategias de resolución de problemas (PSS – 8 ítems, utilizadas en situaciones en las que partes del texto parecen difíciles de leer) y estrategias de apoyo a la lectura (SRS – 9 ítems, como usar materiales de referencia o tomar notas). Este instrumento se acortó años más tarde a una versión de 15 ítems (MARSÍ-R; ver Mokhtari et al., 2018), con 5 ítems por factor, manteniendo la misma estructura conceptual de tres factores propuesta en la versión original. Los autores del instrumento señalan que se puede utilizar tanto la puntuación (suma) total como las de las tres subescalas, según el interés de los y las investigadores/as que utilicen este inventario.

A nivel sustantivo, el factor de estrategias de lectura global (GRS) puede considerarse como un factor general sobre el que analizar y valorar las otras dos facetas en ambas versiones (30 y 15 ítems), por lo que tiene sentido utilizarlo como factor de referencia. Animados por esta situación, en un estudio previo sobre la versión acortada (MARSÍ-R) evaluamos la estructura interna mediante S-1 (Ondé et al., 2022). Además, en el estudio en el que se propone la forma acortada del instrumento (Mokhtari et al., 2018) se observaron correlaciones muy elevadas entre los tres factores (entre .618 y .840), lo que planteaba la existencia

de un factor general fuerte (hipótesis de unidimensionalidad esencial). En lugar de especificar un factor general de meta-comprensión lectora subyacente a los tres dominios de estrategias de lectura propuestos, más difícil de justificar a nuestro juicio, nos planteamos evaluar el grado en el que GRS funcionaba como factor general de referencia controlando el efecto de las dos facetas o dominios restantes (PSS y SRS). Encontramos que el modelo S-1 producía una solución factorial válida, con un adecuado nivel de determinación factorial para GRS ($VCE_{GRS}=.728$, $VCE_{PSS}=.359$, $VCE_{SRS}=.370$, $\omega_{H-GRS}=.779$, $\omega_{S-PSS}=.177$, y $\omega_{S-SRS}=.127$; ver Ondé et al., 2022). Estos resultados reflejaron la conveniencia de utilizar la puntuación factorial de GRS en los análisis correlacionales subsiguientes en el proceso de validación (frente a las puntuaciones de PSS y SRS).

En el presente trabajo extendemos estos resultados a una muestra obtenida a partir de la versión sin acortar del instrumento (MARSÍ). Dado que ambas versiones mantienen la misma estructura a nivel conceptual, tiene sentido mostrar un adecuado nivel de determinación factorial de GRS como factor general de referencia en ambas, valorando la posibilidad de que los resultados obtenidos en la versión reducida no sean más que un artefacto estadístico. La muestra analizada está compuesta por 250 participantes, con edades comprendidas entre los 18 y los 52 años ($M=21.2$, $DT=3.3$) y un 85.2% de mujeres. Nuestra intención al recoger información con este instrumento era la de tratar de extender la evaluación del constructo a población mayor de 18 años en un contexto universitario.

Para ilustrar la utilidad del modelo S-1 se han estimado varios modelos AFC: unifactorial, tres factores correlacionados, bifactor y bifactor S-1. Nótese que en el trabajo de Ondé et al. (2022) no se aplicó el modelo bifactor por falta de encaje teórico. En el presente trabajo se incluye el modelo bifactor con fines ilustrativos (por ejemplo, para mostrar resultados anómalos). Analizamos la matriz de correlaciones policóricas como matriz de entrada y usamos Diagonal Weighted Least Squares (DWLS) como método de estimación con el paquete lavaan del programa R (Rosseel, 2012). En la Tabla 2 se muestran los

Tabla 2. Índices de ajuste de los modelos evaluados (matriz de correlaciones policóricas y estimador DWLS)

Modelo	χ^2 (gl)	CFI	TLI	RMSEA (90% CI)	Close fit (RMSEA)	SRMR
Unifactorial	898.01 (405)*	.902	.895	.080 (.074–.086)	< .001	.090
3 factores corr.	875.54 (402)*	.907	.900	.078 (.072–.084)	< .001	.088
Bifactor ^a	733.83 (375)*	.768	.731	.062 (.056–.069)	.001	.070
Bifactor S-1(GRS)	779.07 (388)*	.930	.921	.069 (.063–.076)	< .001	.081

* $p < .001$.^a Se muestra la solución estimada mediante MV y matriz de correlaciones de Pearson.

resultados obtenidos en términos de ajuste. Cabe señalar que el modelo bifactor no converge, por lo que se cambió la matriz de entrada y el método de estimación (correlaciones de Pearson y Máxima Verosimilitud-MV) con el fin de obtener una solución factorial que valorar. Salvo por los valores de CFI y de TLI, sensiblemente más bajos al utilizar la matriz de correlaciones de Pearson y ML en un tamaño muestral limitado, el modelo bifactor es el que refleja mejor ajuste (χ^2 , RMSEA, SRMR). No obstante, en la solución obtenida se identifican dos resultados anómalos: en primer lugar, se ha estimado una varianza negativa (caso Heywood) para el group factor GRS. En segundo lugar, dicha varianza no es estadísticamente distinta de cero (un resultado similar se obtiene aplicando un modelo de segundo orden, omitido por simplicidad). En el resto de modelos no se producen estos problemas, además de que todas las varianzas de los factores son estadísticamente distintas de cero, siendo el modelo S-1 el que obtiene mejor ajuste.

En cuanto a la determinación factorial de los factores de la solución bifactor, se han obtenido los siguientes resultados: $VCE_G = .741$, VCE_{GRS} – no se puede calcular, $VCE_{PSS} = .143$, $VCE_{SRS} = .116$, $\omega_{H-G} = .831$, ω_{S-GRS} – no se puede calcular, $\omega_{S-PSS} = .227$, y $\omega_{S-SRS} = .112$. Estos resultados, con sus limitaciones, reflejan la conveniencia de utilizar una sola puntuación factorial global estimada a partir del factor general (unidimensionalidad esencial), siguiendo las recomendaciones sobre los valores de VCE y ω_H (Reise et al., 2013). El problema es que esta evaluación se realiza a partir de un modelo que presenta problemas estadísticos graves. De hecho, los valores de VCE y ω_H para el group factor GRS no se pueden calcular debido a la presencia de varianza negativa. Por su parte, el modelo S-1 ha obtenido valores similares a los obtenidos en el estudio previo sobre la versión acortada: $VCE_{GRS} = .763$, $VCE_{PSS} = .128$, $VCE_{SRS} = .109$, $\omega_{H-GRS} = .847$, $\omega_{S-PSS} = .180$, y ω_{S-}

$SRS = .095$. Estos resultados llevan a la misma conclusión sobre la unidimensionalidad esencial y el uso de las puntuaciones del instrumento, pero en una solución factorial que no presenta los problemas del modelo bifactor.

Estos resultados también suponen una ventaja sobre los modelos clásicos. Los valores de ω para el modelo unifactorial y para el modelo de tres factores correlacionados son: $\omega_{unifactorial} = .867$, $\omega_{GRS} = .753$, $\omega_{PSS} = .700$, y $\omega_{SRS} = .693$. A partir de estos valores, ¿cuál podría ser la recomendación respecto al uso de las puntuaciones? Los modelos clásicos no controlan la fiabilidad en función del resto de factores identificados, por lo que la estimación de la fiabilidad puede estar sesgada. En nuestro caso, las correlaciones entre factores obtenidas al aplicar el modelo de tres factores oscilaron entre .739 y .954. Estas correlaciones son muy elevadas, lo que plantea la presencia de unidimensionalidad esencial y sugiere que el modelo unifactorial es más adecuado que el de factores correlacionados. No obstante, $\omega_{unifactorial}$ es ligeramente mayor que ω_{H-GRS} (.867 > .847), pudiendo considerarse este último un nivel de fiabilidad más preciso, al controlar la varianza que se debe a PSS y SRS. Por tanto, sería preferible utilizar la puntuación factorial de GRS de la solución S-1 en el análisis correlacional con otras variables, más depurada que la puntuación factorial del modelo de un solo factor al controlar el efecto (significativo como hemos visto) de las facetas. En otras situaciones en las que las correlaciones entre factores no sean tan elevadas la diferencia entre estos valores podría ser sensiblemente mayor.

A partir de los resultados anteriores se deriva que la evaluación de la invarianza factorial mediante AFC-MG no se puede realizar para el modelo bifactor, ya que el modelo de partida no resulta estadísticamente adecuado. Por otro lado, en el modelo S-1 hemos podido constatar mediante AFC-MG que no se puede asumir

invarianza factorial o equivalencia métrica entre hombres y mujeres ni en función de la edad (a partir de la división en torno al 50% de la muestra de menores de 21 años y 21 años o más). Esta situación abre la posibilidad de estudiar el efecto de estas variables sobre el modelo mediante la técnica MIMIC. Por ejemplo, utilizando la variable sexo como covariable (en este caso solamente sobre los factores, no sobre los ítems) se obtienen los siguientes coeficientes de regresión: GRS ($\beta_{GRS}=-.058$; $p=.767$), PSS ($\beta_{PSS}=.407$; $p=.104$), y SRS ($\beta_{SRS}=.684$; $p=.009$). Solamente es significativo el coeficiente de regresión para el group factor de estrategias de apoyo (SRS), lo que podría indicar algún tipo de sesgo u otras cuestiones de interés. Por último, también se podría utilizar el modelo S-1 para modelar el poder predictivo de los factores del modelo sobre otras variables (Ondé et al., 2022), incluso combinar este tipo de análisis con la inclusión de covariables mediante MIMIC. El modelo S-1 nos permite en este caso seguir profundizando en la evaluación de las propiedades psicométricas del modelo más allá de lo que podemos profundizar con el modelo bifactor. En línea con el objetivo principal de este trabajo, el ejemplo expuesto sirve para ilustrar la potencialidad del modelado para evaluar estructura interna y conectar modelo psicométrico con modelo teórico.

Discusión

En el presente trabajo se han revisado varias de las contribuciones más importantes del AFC para evaluar estructura interna desde una perspectiva centrada en la reflexión sobre validez. Como hemos ido trazando a lo largo del texto, la modelización mediante AFC puede ser una poderosa herramienta de evaluación psicométrica para profundizar en la conexión entre modelo teórico y psicométrico.

Para evaluar estructura interna *per se* puede servir tanto una aproximación AFE como una AFC (incluso una aproximación ACP, si se justifica adecuadamente; ver, por ejemplo, Lloret et al., 2014). La diferencia principal puede estar en la fase de evaluación en la que se encuentra el instrumento evaluado, si es de nueva creación o

presenta resultados de estudios previos bajo hipótesis sustantivas. En los casos en los que el instrumento es de reciente creación y no exista demasiada orientación teórica, hay que reconocer que todavía no se está en disposición de establecer conexiones claras entre modelo psicométrico y teórico.

Actualmente, la principal contribución del AFC respecto al AFE provienen del tipo de restricciones que admiten estos modelos, lo que permite a investigadores/as poner a prueba más propiedades psicométricas de los instrumentos, como invarianza (AFC-MG), análisis de covariables (modelos MIMIC) y análisis del poder predictivo del modelo factorial (Brown, 2015), entre otras. Es cierto que existen ciertos desarrollos dentro del AFE (por ejemplo, estimación de errores típicos y evaluación de invarianza mediante EFA Trees), pero todavía no tienen el alcance que tiene el AFC y, en general, son poco conocidas por los investigadores aplicados. Habrá que esperar para ver cómo evolucionan estas líneas de trabajo en los próximos años. La revisión de la RIDEP muestra que las aplicaciones del AFE en las que se evalúan diferencias entre grupos, las estrategias de análisis mayoritariamente incluyen técnicas de análisis como T de Student, ANOVA o correlaciones con otras variables, recursos más limitados que los que ofrece el AFC. Esta situación sirve como muestra del limitado grado de implementación que existe actualmente de los nuevos desarrollos dentro del marco del AFE, al menos en comparación con la aplicación del AFC. Por otro lado, aunque existen alternativas intermedias como Modelos Exploratorios de Ecuaciones Estructurales, la revisión de la RIDEP refleja comparativamente una escasa presencia en relación con aplicaciones AFE y AFC.

También hemos reflexionado sobre la conveniencia de utilizar procedimientos de validación cruzada en los que se aplica primero un AFE sobre una submuestra y se pone a prueba el modelo resultante mediante AFC en otra submuestra. A nuestro juicio, tanto si se aplica validación cruzada como si no, lo que hay que considerar es si AFC se aplica como herramienta analítica que nos permite profundizar en la evaluación de propiedades psicométricas a partir de la estructura “óptima” identificada previamente

mediante AFE. Esta aplicación del AFC no sería estrictamente confirmatoria, se utilizaría con el fin de imponer nuevas restricciones sobre la estructura identificada en la fase previa, permitiendo poner a prueba más propiedades del instrumento (las que se consideren pertinentes en cada estudio). Este tipo de cuestiones son más técnicas que sustantivas, y así debería quedar reflejado en la justificación que se hace de las distintas herramientas y procedimientos de análisis.

El concepto de unidimensionalidad esencial (o multidimensionalidad esencial, si el interés se centra en las subescalas) es fundamental para entender la utilidad del modelado mediante AFC. Los modelos clásicos como el unifactorial y el de factores correlacionados no permiten evaluar la unidimensionalidad esencial de un conjunto de datos, si bien probablemente son los que mejor reflejan la conceptualización de factores y el proceso de selección de ítems que se produjo en la creación de numerosos instrumentos. Más aún si tenemos en cuenta que muchos de estos instrumentos se llevan empleando desde hace varias décadas.

El modelo bifactor permite descomponer las respuestas a los ítems en dos fuentes de variación o efectos directos y, de esta manera, permite evaluar hasta qué punto la estructura de los datos es más unidimensional que multidimensional o viceversa. No obstante, no hay que confundir el (sobre)ajuste que obtienen los modelos bifactor con la existencia de un factor general a nivel teórico. En este sentido, nuestra propuesta es utilizar bifactor como herramienta de evaluación psicométrica en combinación con los modelos AFC clásicos, no compitiendo con ellos para ver cuál tiene mejor ajuste. Esto es, no es que el ajuste no deba ser valorado, sino que la conclusión sobre cuál es el mejor modelo no debe girar exclusivamente sobre aquel que tenga mejor ajuste (Brown, 2015; Klein, 2015). Como decíamos, probablemente los modelos AF que mejor conectan con la teoría sean los modelos clásicos de un solo factor o de factores correlacionados y, por tanto, con el proceso de selección de ítems y con el tipo de predicciones a nivel sustantivo que se pueden esperar. La excepción puede estar en contextos de evaluación de constructos en los que se conceptualizan

relaciones jerárquicas entre factores, en donde el modelo de segundo orden y el bifactor pueden ser claros competidores con otros modelos clásicos, especialmente en el campo de la inteligencia y cuando se utilizan instrumentos de rendimiento óptimo.

En este trabajo hemos señalado que el modelo bifactor presenta con más frecuencia de la que podría parecer ciertos resultados anómalos (Eid et al., 2017), como estimación de varianzas negativas y aparición de varianzas de factores estadísticamente nulas. Una alternativa de utilidad a nivel de modelado estadístico puede ser el modelo bifactor S-1. Este modelo impone más restricciones que el modelo bifactor, al dejar de especificar uno de los group factor. Como hemos ilustrado mediante un caso aplicado, este tipo de restricciones evitan resultados anómalos, siendo más útil en nuestro caso que el modelo bifactor, además de evitar en cierta medida el sobreajuste. Para este tipo de aplicaciones resulta fundamental valorar la aparición de varianzas negativas, incluso en soluciones convergentes, además de poner a prueba la varianza de los factores mediante tests estadísticos (para lo que es preciso fijar la métrica de los factores fijando a uno la carga factorial de un marcador por factor; ver Brown, 2015). Por otro lado, tanto bifactor como S-1 permiten valorar el grado de determinación factorial del factor general y de las facetas a través de índices como ECV y ω_H . Este tipo de evaluación supone también una importante contribución del modelado mediante AFC, ya que nos va a permitir tomar decisiones sobre la adecuación de las puntuaciones del instrumento (total o subescalas). Además, lo anterior también contribuye a la evaluación del poder predictivo de los modelos ya que las puntuaciones factoriales derivadas (sean la puntuación total o la de las subescalas) son variables más depuradas, más precisas, al controlar la varianza que se debe a unos factores o a otros.

A nivel sustantivo, un aspecto diferencial clave entre bifactor y bifactor S-1 es la forma en la que conectan con el modelo teórico preexistente. Así, el modelo bifactor establece la hipótesis de un factor general (tipo factor g de inteligencia) que subyace a todas las facetas o dominios de contenido, mientras que el modelo S-1 utiliza precisamente una de dichas facetas como

factor general de referencia. Estos dos factores generales no implican concepciones sustantivas equivalentes. Como mostramos en este trabajo, existen instrumentos en donde uno de los group factor se conceptualiza como un factor de carácter general. En estos casos, parece difícil justificar la existencia de un factor general subyacente a todas las facetas cuando una de dichas facetas implica, a su vez, elementos de carácter general, transversal o jerárquico respecto al resto de facetas. A nuestro juicio, no son pocos los instrumentos en los que se puede dar esta situación. Por tanto, el modelo bifactor S-1 puede contribuir enormemente en los próximos años a la evaluación de estructura interna en conexión con la teoría preexistente y con el proceso de redacción y de selección de ítems.

Para finalizar, queremos incidir sobre algunas cuestiones relacionadas con este trabajo. En primer lugar, los resultados de la revisión de la RIDEP que hemos obtenido han sido de utilidad para ilustrar varias de las cuestiones tratadas en este trabajo, si bien no deben considerarse como generalizables. Pensamos que esta práctica de revisión de publicaciones es importante, y debe extenderse a otras revistas y ámbitos de publicación científica, sumándose a los trabajos de revisión desarrollados por parte de otros autores/as. En segundo lugar, respecto al ejemplo aplicado que hemos utilizado para ilustrar el modelo S-1, cabe señalar que se trata de una aplicación concreta, en un contexto de evaluación concreto, por lo que la forma de proceder que hemos mantenido debe considerarse solo de manera orientativa. Es necesario evaluar, en cada caso, si alguno de los group factor tiene entidad suficiente como para poder ser modelizado como factor general de referencia y aplicar, en consecuencia, un modelo S-1. En tercer lugar, aunque hemos visto que S-1 puede prevenir algunos resultados anómalos del bifactor, no los previene todos. En nuestro análisis observamos que algunas cargas factoriales a nivel de faceta eran estadísticamente nulas, lo que puede considerarse como un resultado anómalo (para consultar algunas limitaciones del S-1 se puede consultar el trabajo de Maydeu-Olivares & Coffman, 2006). El modelo S-1 no se ha aplicado con excesiva frecuencia, por lo que todavía debe

estar bajo observación atenta de la comunidad científica.

Referencias

- American Educational Research Association, American Psychological Association & National Council for Measurement in Education [AERA, APA & NCME] (2014). *The Standards for Educational and Psychological Testing*. AERA.
- 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.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225. https://doi.org/10.1207/s15327906mbr4102_5
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*, 241-261.
- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541-562. <https://doi.org/10.1037/met0000083>
- Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence, 6*, 42. <https://doi.org/10.3390/jintelligence6030042>
- Hendy, N., Krammer, G., Schermer, J. A., & Biderman, M. D. (2020). Using bifactor models to identify faking on Big Five questionnaires. *International Journal of Selection and Assessment, 29*(1), 81-99. <https://doi.org/10.1111/ijsa.12316>
- Holzinger, K. J., & Swineford, F. (1937). The bifactor method. *Psychometrika, 2*, 41-54.
- Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema, 26*(3), 395-400. <https://doi.org/10.7334/psicothema2013.349>

- Jackson, D. L., Gillaspay, J. A. Jr., & 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>
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*(2), 183-202.
- Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity, 24*, 387-404.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Ledesma, R. D., Ferrando, P. J., & Tosi, J. D. (2019). Uso del Análisis Factorial Exploratorio en RIDEP. Recomendaciones para autores y revisores. *Revista Iberoamericana de Diagnóstico y Evaluación – Avaliação Psicológica, 3*(52), 173-180.
<https://doi.org/10.21865/RIDEP52.3.13>
- Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología, 30*(3), 1151-1169.
<https://doi.org/10.6018/analesps.30.3.199361>
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(3), 635-694.
<https://doi.org/10.2466/pr0.1957.3.3.635>
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11*(4), 344.
<https://doi.org/10.1037/1082-989X.11.4.344>
- McDonald R. P. (1999). *Test theory: A unified approach*. Lawrence Erlbaum Associates, Inc.
- 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>
- Mokhtari, K., Dimitrov, D. M., & Reichard, C. A. (2018). Revising the Metacognitive Awareness of Reading Strategies Inventory (MARSII) and testing for factorial invariance. *Education Faculty Publications and Presentations, 2*, 219-246.
<https://doi.org/10.14746/ssl.2018.8.2.3>
- Mokhtari, K., & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249-259.
<https://doi.org/10.1037//0022-0663.94.2.249>
- Mulaik, S. A. (2009). *Foundations of factor analysis (2nd Ed.)*. Chapman and Hall/CRC.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.
- Ondé, D., Alvarado, J. M., Sastre, S., & Azañedo, C. M. (2021). Application of S-1 bifactor model to evaluate the structural validity of TMMS-24. *International Journal of Environmental Research and Public Health, 18*(14), 7427.
<https://doi.org/10.3390/ijerph18147427>
- Ondé, D., Jiménez, V., Alvarado, J. M., & Gràcia, M. (2022). Analysis of the structural validity of the Reduced Version of Metacognitive Awareness of Reading Strategies Inventory. *Frontiers in Psychology, 13*.
<https://doi.org/10.3389/fpsyg.2022.894327>
- Pérez, D. O. (2020). Revisión del concepto de causalidad en el marco del Análisis Factorial Confirmatorio. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica, 1*(54), 103-117.
<https://doi.org/10.21865/RIDEP54.1.09>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667-696.
<https://doi.org/10.1080/00273171.2012.715555>
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5-26.
<https://doi.org/10.1177/0013164412449831>
- Revelle, W., & Wilt, J. (2013). The general factor of personality: A general critique. *Journal of Research in Personality, 47*(5), 493-504.
<https://doi.org/10.1016/j.jrp.2013.04.012>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1-36.
<https://doi.org/10.18637/jss.v048.i02>

- Sterner, P., & Goretzko, D. (2022). *Exploratory factor analysis trees: Evaluating measurement invariance between multiple covariates*. OSF. <https://osf.io/7pgrb/>
- 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>
- Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology, 7*, 1832. <https://doi.org/10.3389/fpsyg.2016.01832>