QÜESTIIÓ, vol. 23, 2, p. 239-259, 1999

EFECTO DEL NÚMERO DE INDICADORES POR FACTOR SOBRE LA IDENTIFICACIÓN Y ESTIMACIÓN EN MODELOS ADITIVOS DE ANÁLISIS FACTORIAL CONFIRMATORIO

A. OLIVER
J. M. TOMÁS
P. M. HONTANGAS
Universitat de València*

La matriz multirrasgo-multimétodo (MRMM) es un diseño de investigación de larga tradición en Psicología. Las técnicas de análisis de datos adecuadas para una correcta extracción de conclusiones han estado sujetas a controversia. Parece, no obstante, que diversos modelos de análisis factorial confirmatorio resultan muy adecuados. De entre los diversos modelos, dos de ellos han recibido gran atención, el modelo completo, que apareció primero en la literatura, y el de unicidades correlacionadas, que parece una alternativa razonable a los problemas que aparecen en el primero. Los resultados de ambos modelos en la literatura se refieren a situaciones con un solo indicador por combinación rasgo-método. La presente investigación simula datos de matrices MRMM para múltiples indicadores por combinación rasgo-método y somete a prueba la adecuación de las estimaciones de ambos modelos. Los resultados apuntan a un mejor comportamiento del modelo completo, si bien los sesgos, aunque triviales en cuantía, aumentan conforme aumenta la correlación entre métodos.

Effects of number of indicators per factor on identification and estimation of confirmatory factor analysis models.

Palabras clave: Estimación, sesgo, análisis factorial confirmatorio (AFC), matriz MRMM, número de indicadores, correlación entre métodos.

Clasificación AMS (MSC 2000): 62H25, 62P15

La presente investigación se enmarca en el proyecto PB96-0791 del Programa Sectorial de Promoción General del Conocimiento subvencionado por la DGES, y en sendos proyectos concedidos a los dos primeros autores por el subprograma SEUID (MEC)-Royal Society de Londres. Los autores desean agradecer los comentarios de un revisor, que mejoraron sustancialmente la versión final de este trabajo. Una versión preliminar de este trabajo se presentó en el *V Congreso de Metodología de las Ciencias Humanas y Sociales*, Sevilla, 23-26 de Septiembre, 1997.

* Àrea de Metodologia de les Ciències del Comportament. Facultat de Psicologia. Universitat de València. Av. Blasco Ibáñez, 21. 46010 València. España. Correo electrónico: oliver@uv.es.

– Recibido en noviembre 1998.
– Aceptado en febrero de 1999.

1. INTRODUCCIÓN

El diseño de investigación mediante la matriz multirrasgo-multimétodo fue presentado por Campbell y Fiske (1959). En su aplicación paradigmática se miden distintas variables mediante diversos métodos. Así, puede medirse extraversión, neuroticismo y depresión por medio de tres métodos de medida (autoinforme, registro clínico y observación externa, por ejemplo). Este ejemplo de diseño sería una matriz multirrasgo-múltimétodo 3×3, con tres rasgos y tres métodos. El número de rasgos y de métodos utilizados en la literatura es muy variable, y pueden encontrase diseños de 3×2, 4×4, 9×9, etc. El objetivo de la matriz multirrasgo-multimétodo era ofrecer un diseño de investigación que permitiera analizar la validez convergente-discriminante de las medidas incluidas.

Este diseño de investigación sigue plenamente vigente, aunque el análisis de datos propuesto por estos autores ha sido ampliamente criticado (entre otros, Bagozzi, 1993; Schmitt y Stults, 1986; Widaman, 1985). Como respuesta a estas críticas, se han planteado aplicaciones de diferentes tipos de técnicas para mejorar la extracción de conclusiones. De entre éstas, las soluciones más prometedoras parecen ser las basadas en distintos modelos de análisis factorial confirmatorio (Jöreskog, 1971, 1972; Marsh, 1988, 1989, Rindskopf, 1983).

Dentro del análisis factorial confirmatorio se proponen diversas parametrizaciones o modelos, que varían en términos de supuestos, condiciones de aplicación y adecuación de las estimaciones. El primero en aparecer es el modelo completo, o modelo de rasgos correlacionados y métodos correlacionados (Jöreskog, 1971). Una representación de éste puede verse en la figura 1 (a) para tres rasgos y tres métodos. Como una alternativa a este modelo aparece el modelo de unicidades correlacionadas presentado por Marsh (1988, 1989), basándose en trabajos de Kenny (1976, 1979). Esta parametrización extrae las conclusiones sobre efectos de método a partir de la correlación entre unicidades de las medidas de distintos rasgos que tienen en común un mismo método. Una representación de este modelo, también para tres rasgos y tres métodos, aparece en la figura 1 (b).

El modelo completo ha sufrido fuertes críticas ya que, aunque cuenta con ventajas frente a otras técnicas, presenta muchos problemas, especialmente de identificación y estimación (ver, entre otros, Brannick y Spector, 1990; Kenny y Kashy, 1992; Kumar y Dillon, 1992; Marsh y Bailey, 1991; Wothke, 1996). Por su parte, el modelo de unicidades correlacionadas ha aparecido como una alternativa razonablemente robusta a violaciones de sus supuestos, que ofrece estimaciones adecuadas de la validez convergente-discriminante y de los efectos de método, y que presenta problemas de identificación y/o estimación con escasa probabilidad (Marsh y Bailey, 1991). Los resultados obtenidos se derivan de estudios analíticos y empíricos, con datos reales y de simulación, y resultan altamente consistentes entre sí, lo que ha permitido a un número de autores recomendar la elección del modelo de unicidades correlacionadas en estudios de matrices multirrasgo-multimétodo (Brannick y Spector, 1990; Marsh y Bailey, 1991). Los resultados anteriores se han establecido a partir de datos de matrices multirrasgo-multimétodo en su definición más prototípica. Esto es, cuando se utiliza una sola medida, una variable, como indicador de cada combinación rasgo-método. Si volvemos a la figura 1 (a), la primera variable observable es una medida del primer rasgo tomada mediante el primer método de medida, y es la única medida de ese rasgo tomada con ese método.

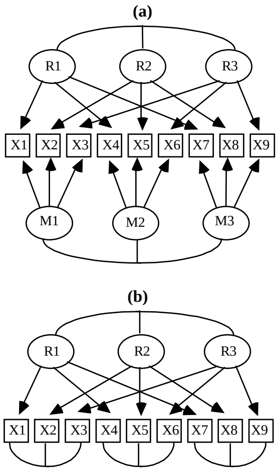

Figura 1. Modelo completo (a) y modelo de unicidades correlacionadas (b).

Notas: Las unicidades se eliminan por simplicidad; las líneas curvas representan covarianzas

Es el único indicador de la combinación primer rasgo por primer método, y de la misma forma ocurre para cada combinación rasgo-método. Lo mismo sucede cuando el modelo es el de unicidades correlacionadas, que supone una parametrización distinta, pero comparte con el anterior modelo las variables observables o medidas. Aún siendo lo más usual en la literatura, ninguna característica del diseño MRMM hace pensar que escoger una única variable observable por combinación rasgo-método sea la única posibilidad, ni siquiera la más recomendable. El uso de más de un indicador o variable observable por combinación de rasgo-método, bien corresponda a distintas medidas de un mismo constructo o a la misma variable medida repetidamente en el tiempo, puede conllevar importantes ventajas. Estas ventajas son comparables a aquéllas que en general se producen al aumentar el número de indicadores en cualquier modelo de análisis factorial confirmatorio. De hecho, «la aproximación de un solo indicador debe verse como un caso especial de la aproximación más general de múltiples indicadores, en la cual se imponen sobre la estructura de los indicadores múltiples restricciones implícitas, a veces no probadas, y quizá arbitrarias» (Marsh, 1993, p. 78). Saris y Andrews (1991) proponen la formulación como modelo de análisis factorial de segundo orden con un solo indicador por combinación rasgo-método. Posteriormente, Marsh (1993) propuso el uso de múltiples indicadores para representar cada combinación rasgo-método como un factor latente, usando estos factores latentes de primer orden como indicadores de los factores de rasgo y de método, que se modelarían, por tanto, como factores de segundo orden. Una representación de un modelo de este tipo puede verse en la figura 2. Además, esta aproximación puede usarse para aplicar las directrices de Campbell y Fiske (1959), eliminando algunos de sus problemas. Por ejemplo, se utilizan variables latentes libres de error, en lugar de variables observables (Marsh, 1993).

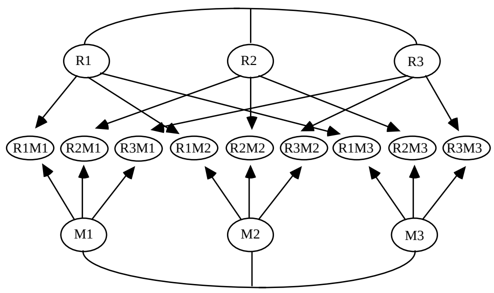

Figura 2. Modelo de matriz multirrasgo-multimétodo con indicadores múltiples analizado mediante análisis factorial confirmatorio de segundo orden.

Notas: Se omiten las variables observables y las unicidades por simplicidad; cada factor de primer orden se infiere a partir de dos o más variables observables; las unicidades, residuales, se eliminan por simplicidad; las líneas curvas representan covarianzas.

Aún sin considerar factores de segundo orden, el análisis factorial confirmatorio de matrices multirrasgo-multimétodo se puede aplicar, y se ha aplicado con éxito, a situaciones en que se tiene más de una variable (indicador) por combinación de rasgo-método. Este es un caso típico cuando se analizan efectos de método en el análisis de ítems de una escala e incluso con combinaciones o paquetes de ítems. En estos casos, los efectos de rasgo se modelan como factores de primer orden, y los efectos de método bien como factores de primer orden, bien mediante unicidades correlacionadas, en función de que se utilice el modelo completo o el de unicidades correlacionadas. Un modelo de estas características puede verse en la figura 3, donde las variables observables pueden ser ítems de una escala, o simplemente distintas medidas observables de un mismo constructo.

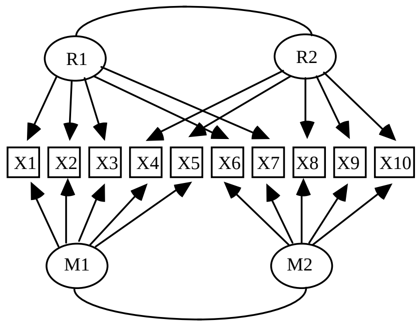

Figura 3. Modelo multirrasgo-multimétodo con más de un indicador por combinación de rasgo-método parametrizado como un modelo completo.

Notas: Se omiten las unicidades por simplicidad; las líneas curvas representan covarianzas.

Lo que se plantea en estos modelos son efectos de método asociados a los ítems y/o variables observables bajo consideración. Se asume que éstos pueden formularse y puntuarse de formas muy diferentes y que estas características formales, al margen del contenido, del rasgo que se intente medir, pueden alterar las relaciones entre variables (Saris y Meurs, 1990). Estos efectos pueden considerarse de método, y por tanto son susceptibles de estudio bajo la lógica del diseño de matrices multirrasgo-multimétodo. De entre las características más estudiadas en la literatura pueden citarse el número de categorías, los ítems comparativos frente a los absolutos, longitud de las baterías, ítems formulados positivamente frente a ítems en negativo, etc. (por ejemplo, Andrews, 1986; Molenaar, 1986). En el estudio de estas características se han utilizado tanto el modelo de rasgos correlacionados y métodos correlacionados como el modelo de unicidades correlacionadas. Sin embargo, existe una laguna en la literatura al respecto del funcionamiento diferencial de ambos modelos en situaciones donde se emplean múltiples indicadores. La justificación del uso de uno u otro modelo se ha basado en la literatura que analiza la situación paradigmática de un indicador por combinación de rasgo-método. Esta situación puede ser muy diferente, desde un punto de vista estadístico, a un diseño que estudie los efectos de método cuando se presentan múltiples indicadores. Una aplicación concreta puede arrojar luz sobre las diferencias. Un diseño de matriz multirrasgo-multimétodo 3×3 presenta nueve variables observables y seis variables latentes, lo que ofrece una razón de variables a factores de 1.5. El uso de múltiples indicadores se puede ilustrar con el trabajo de Marsh (1996), ampliado posteriormente por Tomás y Oliver (1999), donde se aplicó análisis factorial confirmatorio para estudiar efectos de método en los ítems de la escala de autoestima de Rosenberg. Se asumía que la escala era unifactorial (un factor de rasgo), pero el formular los ítems en forma positiva frente a negativa introducía varianza de método (dos factores de método). Dado que la escala de Rosenberg está compuesta de diez ítems, esto ofrecía una ratio de variables a factores de 3.33. Otro ejemplo de aplicación de análisis factorial confirmatorio al estudio de efectos de método con múltiples indicadores se ofrece en Bagozzi y Heatherton (1994). Estos autores estudian los efectos de dos factores de método en una medida multifactorial de autoestima-estado de 20 ítems que presentaba una ratio de variables observables a factores también de 3.33. En trabajos realizados sobre la versión en castellano del STAI-estado (Hontangas, Oliver y Tomás, 1997), se comparan distintas estructuras factoriales con factores de método, que ofrecían ratios de variables a factores de 5, 6.66 y hasta 10. En resumen, aplicar el análisis factorial confirmatorio de matrices multirrasgo-multimétodo generalmente llevará asociada una mayor razón de variables a factores, lo que puede alterar las virtudes y problemas conocidos hasta el momento con estos modelos.

El objetivo de este trabajo es comparar la capacidad del modelo de rasgos correlacionados y métodos correlacionados frente al modelo de unicidades correlacionadas, para estimar de forma adecuada parámetros poblacionales conocidos allá donde se evalúen efectos de método cuando exista más de un indicador por combinación de rasgo-método (ver figuras 3 y 4). Se estudian ambos modelos a través de diversos niveles de correlación entre factores de método, y variando también el número de indicadores por combinación de rasgo-método. Las hipótesis principales que se pretenden contrastar son:

- Al respecto del modelo completo:
  a) presentará menores problemas de identificación y estimará mejor a medida que aumente el número de indicadores por combinación rasgo-método, y
  b) estimará de forma adecuada a todos los niveles de correlación entre métodos.

- Al respecto del modelo de unicidades correlacionadas:
  a) presentará problemas de estimación mayores a medida que aumente el número de indicadores, y
  b) tendrá mayores problemas de estimación a medida que aumente la correlación entre métodos.

MÉTODO

Modelos hipotéticos y generación de las muestras

Por tratarse de un experimento Monte Carlo, se extraen muestras de modelos poblacionales conocidos, en este caso de matrices multirrasgo-multimétodo. Los modelos poblacionales se corresponden en genérico con el modelo de la figura 3.

Tanto la generación de datos como la estimación de los modelos de ecuaciones estructurales se realizó mediante el paquete estadístico EQS 5.1 (Bentler, 1995). Las capacidades de simulación del programa permitieron calcular los dieciséis conjuntos de matrices de varianzas-covarianzas necesarias para el diseño, con cien replicaciones por conjunto. Las variables se generaron para que cumplieran el supuesto de normalidad multivariada. El número de iteraciones máximo para alcanzar la convergencia en el proceso de estimación se fijó a 100.

Diseño

El diseño es factorial mixto 4×4×2 con medidas repetidas en el último factor (tipo de modelo de AFC). Se manipularon las siguientes variables:

- *Número de indicadores por combinación de rasgo-método*, con cuatro niveles, 2, 3, 4 y 5 indicadores.
- *Magnitud de la correlación entre factores de método*, con cuatro niveles, desde factores ortogonales, a valores de .2, .4 y .6. La correlación nula entre métodos es asumida por el modelo de unicidades correlacionadas. Por el contrario, el nivel más alto de correlación, .6, puede considerarse una correlación entre métodos elevada. El estudio más completo en este contexto consideró correlaciones entre factores de método desde .25 a .49 (Marsh y Bailey, 1991).
- Modelo estimado, con dos niveles, modelo de rasgos correlacionados y métodos correlacionados (que se corresponde en genérico con el modelo de la figura 3) y modelo de unicidades correlacionadas (que se corresponde en genérico con el modelo de la figura 4).

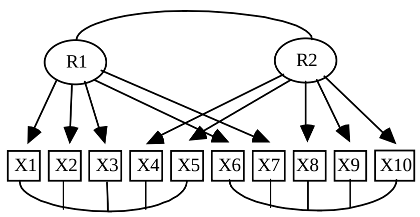

Figura 4. Modelo multirrasgo-multimétodo con más de un indicador por combinación de rasgo-método parametrizado como un modelo de unicidades correlacionadas.

Notas: Se omiten las unicidades por simplicidad; las líneas curvas representan covarianzas.

Las variables dependientes pueden incluirse en tres bloques:

a) Problemas de identificación-estimación, que incluye la variable convergencia (si el modelo converge o no); la variable soluciones impropias (si el modelo ha presentado o no valores fuera de rango, etc.), y soluciones bien definidas (si el proceso de estimación converge y no se presenta ninguna solución impropia, entonces la solución se considera bien definida).
b) Correlación entre rasgos. Para cada modelo ha de estimarse la correlación entre los dos factores de rasgo planteados en la generación de datos. Esta correlación presentaba un valor poblacional de .4. Por lo tanto, la variable dependiente necesaria para evaluar la adecuación de la estimación de este parámetro es la correlación estimada para cada una de las replicaciones.
c) Saturaciones factoriales. A diferencia del caso de las correlaciones entre rasgos, en que el valor estimado es suficiente como variable dependiente, en el caso de las saturaciones se necesitan dos variables dependientes para la evaluación de la adecuación de la estimación, que respectivamente miden el signo y cuantía en valor absoluto de la discrepancia entre las estimaciones y sus valores poblacionales. Estas dos variables dependientes, las mismas utilizadas por Marsh y Bailey (1991) son, respectivamente:
   – Sesgo, medido como el promedio para cada modelo del alejamiento de la saturación estimada frente a su valor poblacional.
   – Precisión, operacionalizada como la raíz cuadrada de la media de la diferencia entre las saturaciones estimadas y su valor poblacional al cuadrado. Esto es, la raíz del error cuadrático medio (RMSE, root mean square error).

El tamaño de la muestra se controló por constancia. A través de todos los modelos se empleó un tamaño muestral de 500. Este tamaño de la muestra se considera muy adecuado en estudios de simulación en modelos de ecuaciones estructurales (Anderson & Gerbing, 1984; Boomsma, 1982). El método de estimación también se controló por constancia, se empleó máxima verosimilitud dado que en la generación de los datos se asumió la normalidad multivariada.

RESULTADOS

Los resultados se ofrecen separados para los distintos grupos de variables dependientes consideradas: a) problemas de estimación, considerando las convergencias, soluciones impropias y soluciones mal definidas; b) la estimación de la correlación entre rasgos (validez discriminante), y c) la estimación de las saturaciones de rasgo (validez convergente).

No convergencias, soluciones impropias y soluciones mal definidas

Las frecuencias de no convergencias, las soluciones impropias y las soluciones mal definidas pueden consultarse en la tabla 1.

Tabla 1. Frecuencias de no convergencias, soluciones impropias y soluciones mal definidas para los dos modelos. Se incluyen las frecuencias por condición y a través de los niveles de las variables independientes (V. I.) en cada modelo.

		Modelo completo			**Modelo de unicidades correlacionadas**		
Tipo de matriz	Correlación métodos	No convergencias	Soluciones impropias	Sol. mal definidas	No convergencias	Soluciones impropias	Sol. mal definidas
	.0	7	18	25	0	100	100
Dos	.2	6	56	61	0	100	100
indicadores	.4	2	52	54	0	100	100
	.6	4	56	58	0	100	100
	.0	2	4	6	0	100	100
Tres	.2	3	24	27	0	100	100
indicadores	.4	5	28	33	0	100	100
	.6	6	40	45	0	100	100
	.0	2	4	6	0	100	100
Cuatro	.2	2	20	22	0	100	100
indicadores	.4	4	24	27	0	100	100
	.6	6	30	34	0	100	100
	.0	0	0	0	0	100	100
Cinco	.2	5	5	10	0	100	100
indicadores	.4	1	5	6	0	100	100
	.6	1	10	11	0	100	100
V. I.							
	2	19	182	198	0	400	400
	3	16	96	111	0	400	400
Indicadores	4	14	78	89	0	400	400
	5	7	20	27	0	400	400
	.0	11	26	37	0	400	400
Corr. entre	.2	16	105	120	0	400	400
métodos	.4	12	109	120	0	400	400
	.6	17	146	148	0	400	400

Con respecto al modelo completo, el número de indicadores presenta un efecto claro sobre la convergencia, las soluciones impropias y, por tanto, las soluciones mal definidas. A medida que aumentan el número de indicadores disminuyen los problemas de identificación y estimación habituales en este modelo. Las soluciones no convergentes aparecen, en cualquier caso con baja frecuencia, mientras que las soluciones impropias son bastante más comunes, especialmente los casos Heywood. Por su parte, la presencia de correlación entre factores de método no afecta de forma apreciable a las no convergencias, que ocurren en una pequeña proporción. Sin embargo sí afecta, y ostensiblemente, a la frecuencia de aparición de soluciones impropias y mal definidas. El número de éstas aumenta drásticamente cuando aparece esta correlación, y tanto más cuanto mayor es ésta.

Por su parte, para el modelo de unicidades correlacionadas no se producen problemas de convergencia. Sin importar el número de indicadores ni la correlación entre métodos, todos los modelos convergen. Sin embargo, las soluciones impropias (y por tanto las mal definidas) ocurren siempre. Estos problemas de estimación suelen ser valores fuera de rango, correlaciones mayores de uno, varianzas de error negativas, etc. Así pues, el modelo de unicidades correlacionadas presentó problemas de estimación graves que afectaban a todos los modelos. Cuando en cualquier modelo de ecuaciones estructurales aparecen soluciones mal definidas, está totalmente contraindicado interpretar los resultados. Esto se debe a la dificultad genérica de establecer la identificación del modelo y, por tanto, a la cautela que debe tenerse cuando cualquier resultado ofrece algún indicio de problemas de identificación. Por lo tanto, y ante la posibilidad de severos problemas de identificación, a partir de aquí se realizarán análisis por separado para el modelo completo y el de unicidades correlacionadas, con especial hincapié en el primero. En los análisis del modelo completo se utilizaron únicamente las soluciones bien definidas. Los análisis del modelo de unicidades correlacionadas se quedarán en el estudio descriptivo de los índices.

Correlación entre factores de rasgo

En la tabla 2 puede consultarse la correlación promedio por condición del estudio, así como las correlaciones promedio para los distintos niveles de las dos variables independientes. Hay que recordar que el valor poblacional se fijó a .4, y es frente a este valor que se deben evaluar las estimaciones. Por tanto, resulta sencillo apreciar las desviaciones frente a este valor. Aunque el modelo de unicidades correlacionadas no ofreció soluciones bien definidas, y por tanto los resultados deben interpretarse con cautela, se ofrecen también en la tabla 2 los promedios de la correlación entre los rasgos a través de niveles y condiciones.

Tabla 2. Número de soluciones bien definidas por celda, correlación media entre rasgos estimada, sesgo de la estimación (Notas: En el caso del modelo de unicidades correlacionadas, no hay ninguna solución bien definida, por lo que se ofrecen los resultados para las soluciones mal definidas, ME = sesgo, RMSR = raíz del error cuadrático medio, D.T. = desviación típica, las medias por nivel son las medias de la variable dependiente para cada nivel de las variables explicativas o independientes).

			Correlación entre rasgos			**Saturaciones de rasgo**			
Modelo	Matriz	ρ	N	r	D.T.	ME	D.T.	RMSR	D.T.
Completo			1175	.3863	.0970	−.0103	.0484	.0687	.0406
	Dos indicadores	.0	75	.4067	.0565	.0020	.0201	.0589	.0166
		.2	39	.4156	.1086	−.0266	.0688	.0942	.0568
		.4	46	.3363	.1450	−.0413	.0831	.1168	.0770
		.6	42	.3712	.1372	−.0170	.0624	.1032	.0549
	Tres indicadores	.0	94	.4021	.0538	−.0001	.0199	.0515	.0131
		.2	73	.3885	.0766	−.0095	.0425	.0657	.0271
		.4	67	.3627	.1310	−.0207	.0605	.0815	.0458
		.6	55	.3697	.1326	−.0192	.0680	.0924	.0525
	Cuatro indicadores	.0	94	.4055	.0499	−.0005	.0170	.0507	.0102
		.2	78	.3713	.1054	−.0156	.0499	.0693	.0322
		.4	73	.3779	.1293	−.0168	.0603	.0779	.0410
		.6	66	.3664	.1227	−.0243	.0681	.0906	.0598
	Cinco indicadores	.0	100	.3988	.0457	−.0007	.0174	.0448	.0079
		.2	90	.4010	.0568	.0019	.0277	.0534	.0128
		.4	94	.3857	.0875	−.0044	.0384	.0582	.0219
		.6	89	.3857	.1065	−.0119	.0533	.0685	.0370
Medias por nivel	Número de indicadores	2	202	.3850	.1135	−.0173	.0606	.0881	.0567
		3	289	.3834	.0994	−.0109	.0486	.0698	.0385
		4	311	.3821	.1039	−.0131	.0510	.0702	.0406
		5	373	.3929	.0773	−.0037	.0365	.0559	.0240
	Correlación de métodos	.0	363	.4030	.0511	.0000	.0185	.0509	.0129
		.2	280	.3915	.0862	−.0099	.0462	.0667	.0336
		.4	280	.3700	.1209	−.0176	.0596	.0785	.0492
		.6	252	.3747	.1217	−.0176	.0620	.0852	.0516
Unicidades			1600	.4358	.0571	.0226	.0249	.0674	.0220
	Dos indicadores	.0	100	.4015	.0573	.0056	.0214	.0704	.0291
		.2	100	.4334	.0582	.0165	.0222	.0713	.0260
		.4	100	.4512	.0435	.0327	.0230	.0740	.0245
		.6	100	.4834	.0548	.0433	.0229	.0781	.0239
	Tres indicadores	.0	100	.3966	.0558	.0028	.0195	.0691	.0268
		.2	100	.4237	.0532	.0166	.0197	.0611	.0176
		.4	100	.4432	.0550	.0281	.0210	.0653	.0174
		.6	100	.4657	.0526	.0422	.0210	.0716	.0180
	Cuatro indicadores	.0	100	.4003	.0523	.0030	.0177	.0711	.0300
		.2	100	.4257	.0482	.0144	.0184	.0629	.0175
		.4	100	.4495	.0463	.0283	.0184	.0645	.0179
		.6	100	.4719	.0446	.0418	.0184	.0680	.0147
	Cinco indicadores	.0	100	.3959	.0474	.0012	.0179	.0634	.0228
		.2	100	.4190	.0402	.0146	.0201	.0593	.0175
		.4	100	.4449	.0471	.0302	.0193	.0625	.0164
		.6	100	.4666	.0436	.0408	.0201	.0658	.0154
Medias por nivel	Número de indicadores	2	400	.4424	.0613	.0245	.0266	.9735	.0269
		3	400	.4323	.0596	.0224	.0249	.0668	.0268
		4	400	.4369	.0547	.0219	.0233	.0666	.0211
		5	400	.4316	.0519	.0217	.0245	.0628	.0183
	Correlación de métodos	.0	400	.3986	.0532	.0031	.0191	.0685	.0273
		.2	400	.4255	.0504	.0150	.0201	.0636	.0204
		.4	400	.4472	.0480	.0290	.0205	.0665	.0197
		.6	400	.4719	.0494	.0420	.0206	.0708	.0188

Se realizó un ANOVA entre-sujetos 4×4 para contrastar los efectos principales y de interacción sobre la variable dependiente correlación estimada entre factores de rasgo. Para este análisis el tipo de modelo se mantuvo constante; sólo se analizaron las soluciones bien definidas del modelo completo. Únicamente el efecto principal de la variable correlación entre métodos resultó estadísticamente significativo ($F_{3,1159} = 9{,}79, p < ,001$). Dado el elevado tamaño de la muestra, se ofrecen las eta-cuadrado (tabla 3). Cuando los métodos son independientes o su correlación baja, el modelo es capaz de estimar de forma adecuada la correlación entre rasgos (la estimación de la validez discriminante es adecuada). Cuando la correlación entre los factores de método aumenta (.4 y .6) se produce una infraestimación de la correlación entre rasgos. La infraestimación que se produce, incluso a los niveles altos de correlación entre métodos, puede calificarse de ligera.

Tabla 3. Eta-cuadrado para los ANOVAS realizados sobre las diversas variables dependientes consideradas. Se ofrecen tan sólo los resultados para las soluciones bien definidas, y por lo tanto sobre el modelo completo, ya que el modelo de unicidades correlacionadas resultó siempre en soluciones mal definidas.

	Correlaciones entres rasgos	Saturaciones de rasgo	
Factores	r	ME	RMSR
M	.003	.015	.104
C	.025	.031	.142
M × C	.012	.015	.027

Notas: M = número de indicadores; C = correlación entre factores de método; r = correlación estimada entre métodos; ME = sesgo; RMSE = raíz del error cuadrático medio. Se redondeó al tercer decimal. Se analizaron sólo los 1175 modelos con soluciones bien definidas.

Saturaciones factoriales de rasgo

Para estudiar el efecto principal del número de indicadores y de la correlación entre métodos y su interacción sobre la estimación de las saturaciones de rasgo, se consideraron dos variables dependientes sesgo y precisión (RMSE) de la estimación. Se realizó un ANOVA entre-sujetos 4×4 para cada variable dependiente. Como en el caso del ANOVA realizado sobre la variable correlación entre rasgo, se utilizaron sólo las soluciones bien definidas, y por tanto sólo se incluyeron en el análisis las estimaciones mediante el modelo completo. Las medias de los efectos principales y de la interacción, para las dos variables dependientes, pueden consultarse en la tabla 2. Con propósitos descriptivos, también se presentan en esta tabla los promedios de estos dos índices para las soluciones estimadas mediante el modelo de unicidades correlacionadas.

El sesgo promedio para todas las condiciones fue de $-.01$, que puede considerarse poco apreciable. La cuantía del sesgo viene afectada por el número de indicadores ($F_{3,1159} = 6{,}08, p < ,001$), por la correlación entre factores de método ($F_{3,1159} = 12, p < ,001$), y en menor cuantía por la interacción de ambas variables ($F_{9,1159} = 1{,}99, p = ,037$). El porcentaje de varianza explicada en su conjunto fue del 5.2 %. El mayor efecto corresponde, al igual que en el caso de la correlación entre rasgos, al de la variable correlación entre métodos (η^2 = .031). Como puede verse en la tabla 2, el modelo completo estima las saturaciones de rasgo (validez convergente) prácticamente sin sesgos cuando los factores de método son independientes. Cuando los factores de método son oblicuos, y a medida que su correlación aumenta hasta moderada-alta, aumenta el sesgo que, como promedio, se sitúa en una diferencia con respecto al valor poblacional de $-.02$. Este sesgo puede considerarse pequeño, incluso en el nivel de mayor correlación entre rasgos, cuando su cuantía resultó mayor. Por su parte, el efecto de interacción y el efecto principal restantes son muy modestos. A destacar, no obstante, que para la condición de cinco indicadores, y prácticamente para cualquier correlación entre métodos, el sesgo es insignificante.

También se realizó un ANOVA entre-sujetos 4$\times$4 sobre la precisión (RMSE), el segundo de los indicadores considerados. El efecto principal del número de indicadores resultó estadísticamente significativo ($F_{3,1159} = 44{,}92, p < ,001$), así como el de la correlación entre factores de método ($F_{3,1159} = 64{,}03, p < ,001$), y la interacción ($F_{9,1159} = 3{,}59, p < ,001$). El porcentaje de varianza explicado por los tres efectos alcanzó un 21.5 %. En la tabla 3 pueden consultarse los valores de eta-cuadrado para cada variable independiente y su interacción. A la vista de las medias ofrecidas en la tabla 2, el efecto principal del número de indicadores resulta muy claro, aumentando la precisión de la estimación a medida que se introduce mayor número de indicadores por combinación rasgo-método. El efecto principal de la variable correlación entre métodos también resulta claro, el modelo completo presenta una mayor precisión conforme disminuye la correlación entre los factores de método. En cuanto al efecto de interacción, que resultó apreciablemente más débil que los efectos principales, la estimación es muy precisa cuando el número de indicadores es cinco, aún cuando los factores de método presenten una correlación elevada. No obstante, el sesgo que se produce en conjunto puede, una vez más, considerarse pequeño, incluso para las condiciones en que la precisión es menor.

CONCLUSIONES Y DISCUSIÓN

A la luz de los resultados, el comportamiento de ambos modelos, el completo y el de unicidades correlacionadas, ha sido muy diferente y contrario a parte de la evidencia analítica y de simulación divulgada hasta la fecha. Una primera conclusión general es que el modelo completo ha mostrado mejor comportamiento que el de Marsh para dos o más indicadores por combinación rasgo-método en diseños multirrasgo-multimétodo, las condiciones bajo estudio. Esta conclusión debe complementarse con otras de carácter específico que se desglosan a continuación para cada uno de los modelos.

Los problemas encontrados en el modelo de unicidades correlacionadas son severos y están asociados con las soluciones impropias que parecen indicar problemas de identificación y/o estimación. En estas condiciones los resultados no deberían interpretarse, recomendación razonable dada la tendencia del diseño multirrasgo-multimétodo a generar este tipo de problemas. Resultados empíricos previos no hacían predecible este comportamiento tan inadecuado del modelo de unicidades correlacionadas, que venía siendo utilizado, precisamente, cuando el modelo completo presentaba este tipo de problemas. Analíticamente, no obstante, este comportamiento es comprensible. Tomemos como ejemplo el caso de dos factores de rasgo y dos de método con cinco indicadores por combinación. En este caso, para cuantificar los efectos de método en el modelo completo es necesario estimar veinte parámetros, saturaciones factoriales en este caso, diez por cada factor de método. Si se quiere valorar estos mismos efectos de método desde el modelo de unicidades correlacionadas se deben estimar noventa parámetros, cuarenta y cinco covarianzas de error por método. En estos casos, puede esperarse que aparezcan problemas de estimación, sobre todo en las varianzas de error, que pueden ofrecer fácilmente soluciones negativas, como así ha ocurrido. Como un revisor acertadamente comentó, pueden adoptarse distintas medidas para disminuir la sobreparametrización del modelo de unicidades correlacionadas. Por ejemplo, las covarianzas entre residuales pueden asumirse iguales dentro de cada combinación rasgo-método, etc. Implicando una disminución en el número de covarianzas entre residuales libres a estimar, el número de restricciones varía en función de los indicadores disponibles por combinación rasgo-método. Estas restricciones (fijar a cero las varianzas de error negativas, por ejemplo) son también una práctica habitual en el caso del modelo completo y conseguirían aumentar el porcentaje de soluciones bien definidas. Como solución a posteriori, es sólo recomendable tras descartarse otras alternativas como el empleo de otros modelos, mayor número de indicadores, mayor tamaño muestral, etc. El objetivo de este trabajo era evaluar estos modelos tal y como se definen a priori, sin restricciones, siendo un problema empírico determinar la adecuación de las restricciones en cada aplicación. Para los datos simulados en el presente trabajo no existía tal problema, puesto que los modelos se estimaron con distintos valores para las covarianzas entre unicidades de una misma combinación rasgo-método. Aunque no es aplicable al presente estudio de simulación, el uso de esta constricción en modelos como el de unicidades correlacionadas puede ser una buena alternativa en situaciones aplicadas cuando ambos modelos, completo y de unicidades, ofrezcan soluciones mal definidas. En cualquier caso, una minuciosa evaluación del ajuste ofrecerá datos sobre la plausibilidad de las restricciones empleadas.

Aunque al producirse soluciones mal definidas no deberían interpretarse los resultados, puede ser de interés evaluar el sesgo y precisión de la estimación también para el modelo de unicidades correlacionadas. Un primer resultado llamativo es que con el modelo de unicidades correlacionadas se produce una sobreestimación, tanto en las correlaciones entre rasgos como en las saturaciones factoriales. Por tanto, la evaluación de la validez convergente de las medidas se ve incrementada, apareciendo los rasgos más solapados de lo que en realidad están. Si se analiza en mayor profundidad lo que ocurre en las estimaciones al manipular las dos variables independientes, se halla un patrón de comportamiento claro del modelo. A la vista de los promedios de la tabla 2, puede decirse que el modelo de unicidades correlacionadas sobreestima por igual sin importar el número de indicadores por combinación rasgo-método, o en cualquier caso las diferencias en sesgo son mínimas. Este no es el caso para los diversos niveles de correlación entre métodos. Los sesgos en la estimación aumentan a medida que aumenta la correlación entre métodos. Este resultado era esperado ya que el modelo de unicidades correlacionadas se basa en el supuesto de métodos no correlacionados.

En cuanto al modelo completo cabe destacar varios resultados. Los problemas de identificación y/o estimación han ocurrido, pero con menor frecuencia de la esperada en función de los resultados de la literatura. Se encuentran en ésta porcentajes de soluciones mal definidas que superan el 77 %, mientras en el presente caso de dos o más indicadores por combinación rasgo-método, éstas, en su conjunto, disminuyen hasta un 26.56 %. Este resultado es tanto más interesante cuanto que baja el porcentaje de soluciones mal definidas conforme aumentan el número de indicadores, situándose en tan sólo un 6.75 % cuando éstos son cinco. Un segundo aspecto a resaltar es la infraestimación que en general produce el modelo a través de las distintas condiciones exploradas. Esta infraestimación es, no obstante, pequeña, incluso en las condiciones que han resultado más desfavorables para el modelo. El sesgo y precisión de la estimación ha disminuido, en general, conforme aumentaba el número de indicadores por combinación. Este resultado es acorde a la hipótesis planteada y esperado, al menos, por dos motivos: a) en cualquier tipo de análisis factorial confirmatorio aumentar el número de indicadores por factor facilita la identificación, y b) al introducirse mayor número de indicadores se consigue elevar la razón de variables observables a factores. El efecto de la variable correlación entre factores de método ha resultado contrario al esperado por hipótesis. A mayor correlación, la estimación resultaba más sesgada y menos precisa. No obstante, cabría esperar ausencia de efecto de la variable correlación entre factores de método, ya que el modelo permite la correlación entre rasgos y entre métodos. Aparece, pues, la necesidad de explicar este resultado en términos analíticos. Como resultado más aplicable de este estudio, incidir en que el modelo completo ha mostrado un buen comportamiento bajo las condiciones de este estudio, y se recomienda su uso frente al modelo de unicidades correlacionadas.

Los resultados de este experimento han resultado claros y, en algunos aspectos, relativamente sorprendentes. A nuestro entender abren una nueva vía de investigación poco representada en la literatura, a la vez que introducen un interrogante sobre las principales conclusiones establecidas sobre análisis factorial confirmatorio de matrices multirrasgo-multimétodo. El corpus de conocimiento acumulado sobre los diversos modelos de análisis factorial confirmatorio de matrices multirrasgo-multimétodo ofrecía un veredicto claro: de entre los modelos aditivos disponibles, el completo con sus variantes y el de unicidades correlacionadas, era este último el que había que escoger. A la multitud de problemas de identificación y estimación del modelo completo se había unido el buen comportamiento del modelo de Marsh, junto con su robustez frente a incumplimientos de sus supuestos, lo que lo hacían muy versátil para enfrentar una amplia gama de situaciones. Diversos autores concluyeron, por tanto, que el modelo de Marsh resolvía de forma más adecuada los diseños multirrasgo-multimétodo y era el modelo a elegir. Esta evidencia sólo se mantiene, claro está, mientras se mantengan aproximadamente las mismas condiciones en que se realizaron los estudios. Un aspecto común a todos ellos era la presencia de un único indicador por combinación rasgo-método. La evidencia del comportamiento cuando se presentan dos o más indicadores provenía de dos fuentes. Por un lado la indirecta, ya que en la técnica de análisis factorial confirmatorio, en su sentido más amplio, era conocido que a mayor número de indicadores era tanto menos probable encontrar problemas de identificación y estimación. Y en segundo lugar, los resultados de estudios aplicados en que no se apreciaban mayores problemas de identificación y/o estimación al aplicar el modelo completo. Toda esta evidencia resultaba, no obstante, inconexa y parcial, sugiriendo la necesidad de estudios de simulación, con una selección precisa de variables y un contexto controlado. Aunque deba complementarse con nuevos estudios de simulación que exploren otras condiciones, este trabajo ha intentado contestar de forma precisa a las hipótesis planteadas, abriendo nuevos interrogantes en un campo en que aparentemente las respuestas estaban bien establecidas.

REFERENCIAS

Anderson, J.C. & Gerbing, D.W. (1984). «The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis». *Psychometrika*, 49, 155-172.

Andrews, F.M. (1984). «Construct validity and error components of survey measures: A structural modeling approach». *Public Opinion Quarterly*, 48, 409-442.

Bagozzi, R.P. (1993). «Assessing construct validity in personality research: Applications to measures of self-esteem». *Journal of Research in Personality*, 27, 49-87.

Bagozzi, R.P. & Heatherton, T.F. (1994). «A general approach to representing multifaceted personality constructs: application to state self-esteem». *Structural Equation Modeling*, 1, 1, 35-67.

Bentler, P.M. (1995). *EQS structural equations program manual*. Encino, C.A.: Multivariate Software, Inc.

Boomsma, A. (1982). «The robustness of LISREL against small sample sizes in factor analysis models». In K.G. Jöreskog and H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction (Part I)*. Amsterdam, The Netherlands: North-Holland.

Brannick, M.T. & Spector, P.E. (1990). «Estimation problems in the block-diagonal model of the multitrait-multimethod matrix». *Applied Psychological Measurement*, 14, 325-339.

Campbell, D.T. & Fiske, D.W. (1959). «Convergent and discriminant validation by multitrait-multimethod matrix». *Psychological Bulletin*, 56, 81-105.

Hontangas, P., Oliver, A. y Tomás, J.M. (1997). «Estudio de efectos de método en la validez factorial del STAI-estado». *VII Conferencia Española de Biometría*. Córdoba, 21-24 de septiembre.

Jöreskog, K.G. (1971). «Statistical analysis of sets of congeneric tests». *Psychometrika*, 52, 99-111.

Jöreskog, K.G. (1974). «Analyzing psychological data by structural analysis of covariance matrices». In R.C. Atkinson, D.V. Krantz, R.D. Luce & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1-56). San Francisco CA: W.U. Freeman.

Kenny, D.A. (1976). «An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix». *Journal of Experimental Social Psychology*, 12, 247-252.

Kenny, D.A. (1979). *Correlation and causality*. New York, Wiley.

Kenny, D.A. & Kashy, D.A. (1992). «Analysis of the multitrait-multimethod matrix by confirmatory factor analysis». *Psychological Bulletin*, 112, 165-172.

Kumar, A. & Dillon, W.R. (1992). «An integrative look at the use of additive and multiplicative covariance structure models in the analysis of multitrait-multimethod data». *Journal of Marketing Research*, 29, 51-64.

Marsh, H.W. (1988). «Multitrait-multimethod analyses». In J.P. Keeves (Ed.), *Educational research methodology, measurement and evaluation: An international handbook*. Oxford, Pergamon Press.

Marsh, H.W. (1989). «Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions». *Applied Psychological Measurement*, 13, 335-361.

Marsh, H.W. (1993). «Multitrait-multimethod analyses: inferring each trait-method combination with multiple indicators». *Applied Measurement in Education*, 6, 49-81.

Marsh, H.W. (1996). «Positive and negative global self-esteem: A sustantively meaningful distinction or artifactor?». *Journal of Personality and Social Psychology*, 70, 810-819.

Marsh, H.W. & Bailey, M. (1991). «Confirmatory factor analysis of multitrait-multimethod data: A comparison of alternative models». *Applied Psychological Measurement*, 15, 47-70.

Molenaar, N.J. (1986). *Formuleringseffecten in survey-interviews*. Amsterdam, VU-Uitgeverij.

Saris, W.E. & Andrews, F.M. (1991). «Evaluation of measurement instruments using a structural equation modeling approach». In P.P. Biener, R.M. Groves, L.E. Lyberg, N. Mathiowtez & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 575-599). New York: John Wiley & Sons.

Saris, W.E. & van Meurs, A. (1990). *Evaluation of measurement instruments by meta-analysis of multitrait-multimethod studies*. Amsterdam: North Holland.

Schmitt, N. & Stults, D.N. (1986). «Methodology review: Analysis of multitrait-multimethod matrices». *Applied Psychological Measurement*, 10, 1-22.

Tomás, J.M. & Oliver, A. (1999). «Rosenberg´s self-esteem scale: two factors or method effects». *Structural Equation Modeling*, 6, (1), 84-98.

Wothke, W. (1996). «Models for multitrait-multimethod matrix analysis». In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*. New Jersey, LEA.

Widaman, K.F. (1985). «Hierarchically nested covariance structure models for multitrait-multimethod data». *Applied Psychological Measurement*, 9, 1-26.

ENGLISH SUMMARY

EFFECTS OF NUMBER OF INDICATORS PER FACTOR ON IDENTIFICATION AND ESTIMATION OF CONFIRMATORY FACTOR ANALYSIS MODELS

A. OLIVER
J. M. TOMÁS
P. M. HONTANGAS
Universitat de València*

The multitrait-multimethod (MTMM) matrix is a research design with a long tradition in Psychology. However, data analyses of such design have been under controversy. It seems that confirmatory factor analysis (CFA) provides a solid base for the analysis of the MTMM matrix. Among the different CFA approaches to this design, the correlated traits correlated methods (CTCM) model, that was first presented, and the correlated traits correlated uniqueness (CTCU) model, which provides an alternative to the main problems of the first model, have received attention in the literature. Results for these two models have been based on a classic MTMM design, with a single indicator per trait-method combination. The present investigation simulates MTMM data with multiple indicators per trait-method combination, and tests for bias in the estimates of the CTCM and CTCU models. The CTCM model performed better than the CTCU model. The amount of underestimation was trivial. Bias slightly increased when method intercorrelations increased.

Keywords: Estimation, bias, confirmatory factor analysis (CFA), multitrait-multimethod (MTMM) matrix, number of indicators, correlation between methods.

AMS Classification (MSC 2000): 62H25, 62P15

This research was partially funded by project PB96-0791 from Programa Sectorial de Promoción General del Conocimiento., DGES (MEC) and by SEUID-98 (MEC)-Royal Society of London. We thank a reviewer for helpful suggestions. A preliminary version of this paper was presented at *V Congreso de Metodología de las Ciencias Humanas y Sociales* (5th Congress of Methodology for the Behavioural and Social Sciences) held in Sevilla, Spain, 23-26 September 1997.

* Àrea de Metodologia de les Ciències del Comportament. Facultat de Psicologia. Universitat de València. Av. Blasco Ibáñez, 21. 46010 València. España. E-mail: oliver@uv.es.

– Received November 1998.
– Accepted February 1999.

Campbell & Fiske (1959) proposed the multitrait-multimethod (MTMM) design to analyze the convergent and discriminant validity of psychological measures. Altogether with the design, they also proposed data analysis tools. Although the design, as formulated by the authors, is still in use, the data analysis has been critisised (for example, Bagozzi, 1993; Schmidt & Stults, 1986; Widaman, 1985). Confirmatory factor analysis has been defended as the most valuable statistical technique for this design (Jöreskog, 1971, 72; Marsh, 1988, 89; Rindskopf, 1983). However, several confirmatory models exist, based on different assumptions, conditions, etc. Among these, the two most popular are the complete model, or correlated traits-correlated methods (CTCM) model, in which method effects are posited as latent variables, and the correlated traits-correlated uniqueness (CTCU) model, in which method effects are posited as covariances among the uniquenesses. Most analytical, empirical and simulation evidence on the adequacy of these models to extract conclusions of the MTMM design is based on matrices with a single indicator (observed variable) per trait-method combination. The use of multiple indicators has been proposed, however, to overcome major problems in the use of confirmatory factor analysis for the MTMM matrices (Marsh, 1993). In general, the use of multiple indicators per trait-method combination is associated to a higher ratio of observed variables to factors, compared to the typical MTMM design when one indicator per combination exist. The main aim of this paper is to compare the ability of the CTCM and CTCU models to adequately estimate key parameters in the MTMM design. Monte Carlo simulation is used to compare several levels of correlation among methods, varying the number of indicators per trait-method combination.

METHOD

EQS 5.1 (Bentler, 1995) has been used to simulate the data. The simulation design was a $4 \times 4 \times (2)$. The first independent variable was the number of indicators per trait-method combination: 2, 3, 4 and 5 indicators. The second independent variable was the correlation among methods: 0, .2, .4 and .6. Finally, the CTCM and CTCU models were used to model the data. The effect of these three independent variables on identification, correlation among traits and trait factor loadings has been studied.

RESULTS

Regarding the nonconvergence and improper solutions, on one hand the CTCM model had less non-convergent and ill-defined solutions as the correlation among methods decreases. The identification problems also decreased as the number of indicators increased. On the other hand, the CTCU model had no convergence problems, although all the solutions found by this model were ill-defined. In respect to correlation among traits the CTCM models had no bias when methods were independent. There was slight negative bias when methods were correlated. The amount of bias increased as the correlation increased. The CTCU model solutions were not included in the analysis, because there were 100 % of ill-defined solutions.

The CTCM correctly estimated the trait factor loadings (convergent validity) for the condition of independent methods. In the presence of oblique methods, there were slight biases. These biases increased as the correlation among methods increased. The amount of bias (although slight at the different levels) decreased as the number of indicators increased. The accuracy of estimates also increased as the number of indicators increased. The accuracy of estimates also increased as the correlation among methods decreased.

CONCLUSIONS

Both models' behaviour (CTCM and CTCU) has been different to that found in the previous literature. A general conclusion is that the complete model shown a better behaviour when multiple indicators per trait-method combination exist. The CTCU model presented severe problems in the estimation process. 100 % of the solutions were ill-defined.

These results were counterintuitive, given that the CTCU model appeared in the literature to overcome major estimation problems of the CTCM model, when a single indicator per trait-method combination exists. The CTCM model dramatically decreased the number of estimation problems when compared to previous studies when there were a single indicator per combination.

In summary, for the simulated conditions in this work, the CTCM model seems an adequate analytical tool for the analysis of MTMM matrices if there are several indicators per trait-method combination.