

Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto

Paula Elosua
Universidad del País Vasco

Desde la perspectiva lineal del factor común la equivalencia de medida se equipara con la invarianza factorial estricta. En este trabajo mostramos un modo progresivo de evaluar la invarianza factorial que analiza de forma secuencial la invarianza configural, la invarianza métrica y la invarianza factorial estricta; en una segunda etapa del análisis y tras rechazar un modelo de invarianza se lleva a cabo un análisis individualizado ítem a ítem para obtener información sobre las causas de la no-invarianza. El modo de proceder se ejemplifica con el estudio de la invarianza en la adaptación de tests. Se adapta al euskera una escala de autoconcepto (AFA-A) y se analiza la invarianza factorial entre las versiones original y adaptada. Los resultados concluyen la equivalencia configural, métrica y fuerte entre todos los ítems de la escala. Además 29 de los 31 ítems presentan equivalencia factorial estricta.

Progressive way in strict factorial invariance. The common factor model equate the equivalence between tests with the strict factor invariance. This paper illustrates one sequential way to evaluate the factor invariance. First, we analyze the configural invariance, the metric invariance and the strict factor invariance. Second we put our attention on item analysis, to look for information about non-invariance focus. This way to work is showed by an empirical study in the field of test adaptation. We adapted to Basque language one test about self-concept (AFA-A) and then we analyze the strict factor invariance. The results showed the configural, metric and strong invariance was presented in the adaptation of test. Also 29 item (n= 31) have the property of strict factor invariance.

La comparación entre los resultados obtenidos con un test en contextos diferentes es una práctica que desde una perspectiva psicométrica sólo es admisible si se justifica empíricamente su equivalencia formal y sustantiva (Elosua, 2003). No tendría sentido contraponer puntuaciones de personas pertenecientes a dos grupos culturales en un constructo que no tuviera la misma representación (significado) en las dos culturas. Tampoco sería correcto interpretar unívocamente los resultados de un test (o de formas de éste) si su aplicación en dos contextos se asocia con errores de medida diferentes. La pertinencia de ambos «usos» está supeditada a la invarianza de medida de las puntuaciones.

La invarianza de medida se define con relación a un grupo o forma de un test, de tal modo que el significado formal y sustantivo de la medición es independiente respecto a ellos. Formalmente diríamos que la distribución de puntuaciones observadas depende sólo del espacio latente multidimensional (Mellenbergh, 1989; Meredith, 1993) del que aquellas son indicadores. A saber,

$$f(\mathbf{Y}|\eta, g) = f(\mathbf{Y}|\eta) \quad (1)$$

donde:

$f(\bullet)$ es la función de distribución de probabilidad de la variable observada.

η es el vector d -dimensional de puntuaciones factoriales.

g es el valor de una variable de selección de grupo G o forma del test.

Según la ecuación 1 los valores que pudieran adoptar las variables observadas (en nuestro caso respuestas a ítems) dependerían únicamente de los valores que adquirieran las variables latentes (nos referiremos indistintamente a variable latente, factor o dimensión), lo cual implica la independencia entre \mathbf{Y} y G . De no cumplirse 1 sería imposible asignar el mismo significado sustantivo a las puntuaciones, y en consecuencia sería incorrecta su equiparación. En esta última situación, las respuestas dependerían del espacio latente medido (η) y de otra variable asociada al grupo o forma con la que aquella interferiría causando un error sistemático conocido como sesgo de medida. Dado que el sesgo puede aparecer cuando se producen cambios contextuales en el uso de un test respecto a sus condiciones originales (cambio de idioma, cambio de grupo normativo, cambio de condiciones de administración, cambio de objetivos...) en estas circunstancias sería conveniente evaluar la invarianza métrica de las nuevas puntuaciones.

De la definición de invarianza (ecuación 1) se deduce que para su estudio es necesario especificar una relación funcional entre las variables manifiestas y el espacio latente. Ésta básicamente puede ser lineal y no-lineal. El primer tipo se correspondería con los modelos de ecuaciones estructurales, mientras que el segundo se relacionaría con los modelos de respuesta al ítem (Lord, 1980).

Fecha recepción: 16-4-04 • Fecha aceptación: 18-10-04

Correspondencia: Paula Elosua

Facultad de Psicología

Universidad del País Vasco

20009 San Sebastián (Spain)

E-mail: pspelolp@sc.ehu.es

El acercamiento al problema de la invarianza de medida desde el modelo lineal parte de la conocida ecuación que define la puntuación observada (o vector de puntuaciones) (y), a través de un modelo de regresión en el que la variable independiente unidimensional es latente (Jöreskog, 1971).

$$y = v + \Lambda \eta + \varepsilon \quad (2)$$

donde:

- y es el valor de la variable aleatoria n -dimensional \mathbf{Y} .
- v es el vector n -dimensional de interceptos.
- Λ es la matriz $n \times d$ de coeficientes regresores o pesos factoriales de los d factores subyacentes a \mathbf{Y} ($n > d$).
- η es el vector d -dimensional de puntuaciones factoriales.
- ε es el residual (parte unicidad y parte error) n -dimensional de la regresión de \mathbf{Y} sobre η .

Este modelo de medida asume el carácter continuo de la variable observada, la linealidad de la relación entre variables observadas y latentes, la distribución normal de los residuales, la independencia entre ellos, y la independencia entre los errores y las variables latentes. A partir de este modelo es posible representar las medias aritméticas y covarianzas de las variables observadas por medio de parámetros estructurales:

$$\begin{aligned} \mu &= v + \Lambda \alpha & (3) \\ \Sigma &= \Lambda \Psi \Lambda^T + \Theta & (4) \end{aligned}$$

donde:

- μ es el vector de medias de \mathbf{Y} .
- v es el vector n -dimensional de interceptos.
- Λ es la matriz nd de coeficientes regresores.
- α es el vector de medias de η .
- Σ es la matriz de covarianzas de \mathbf{Y} .
- Ψ es la matriz de covarianzas de η .
- Λ^T es la matriz transpuesta de Λ .
- Θ es la matriz de covarianzas de ε .

Según 3 y 4 para garantizar que se compara el mismo constructo sustantivo y que éste además tiene las mismas características métricas entre los grupos, tendremos que determinar la igualdad entre los coeficientes regresores o pesos factoriales (Λ), los valores interceptales (v) y las varianzas/covarianzas residuales (Θ). Desde esta perspectiva, y dado que estamos tratando con un modelo de medida, la invarianza métrica se equipara con la invarianza factorial (Meredith, 1964, 1993; Bloxom, 1972; Ellis, 1993). La equivalencia de los parámetros entre subpoblaciones determinará la invarianza dentro del modelo del factor común conocida como Invarianza Factorial Estricta (*Strict Factor Invariance*).

La invarianza factorial estricta es el mayor nivel de acuerdo alcanzable entre estructuras factoriales. Sin embargo, no es el único. En función del número y condición de los parámetros que permanecen invariantes entre subpoblaciones es posible distinguir tres niveles básicos de invarianza: Invarianza Configural, Invarianza Métrica, Invarianza Factorial Estricta (Tabla 1). La primera (modelo 1) ubica la invarianza en la configuración básica del modelo de medida; la segunda (modelo 2) constriñe la equivalencia a los pesos factoriales y la tercera (modelo 3) añade como restricción la

equivalencia entre valores interceptales. Los dos últimos niveles además pueden extenderse (modelos 2a y 3a) si se considera la equivalencia entre las varianzas/covarianzas de los términos de unicidad/error.

En este marco de trabajo el objetivo del presente estudio es mostrar las ventajas que presenta la evaluación progresiva de la invarianza en la búsqueda de información sobre el sesgo. Dado que las ecuaciones estructurales a comparar entre grupos son modelos anidados que van constriñendo los parámetros en forma secuencial, proponemos un estudio progresivo de la invarianza factorial que vaya aportando información gradual sobre los parámetros estructurales que definen el modelo de medida en cada uno de los grupos. Este análisis sería global en una primera etapa, e individualizado (ítem por ítem) una vez que se rechace un nivel de invarianza.

Esta propuesta teórica es aplicada al problema psicométrico de análisis de la equivalencia en la adaptación de tests. Para ello se estudia progresivamente la invarianza factorial entre las versiones original y adaptada al euskera de una escala sobre autoconcepto.

Método

Participantes

La muestra está compuesta por 1.809 estudiantes de la comunidad autónoma del País Vasco, distribuidos entre el 6.º curso de Educación Primaria, Educación Secundaria y Bachiller. De acuerdo con el diseño utilizado en este trabajo, se han utilizado dos submuestras; la primera se corresponde con la población destinataria del test original (submuestra original= 539), y la segunda (submuestra adaptada= 1.269) se identifica con el grupo normativo de la versión adaptada del test; es un grupo que cursa sus estudios en euskera. En la submuestra adaptada 604 son chicos y 665 son chicas, que se distribuyen por cursos como sigue: $N_{EP6} = 110$, $N_{ESO} = 831$, $N_{BA} = 328$. La submuestra de referencia la forman 262 chicos y 277 chicas, que se reparten en los siguientes cursos: $N_{EP6} = 70$, $N_{ESO1} = 414$ y $N_{BA} = 45$.

La selección de los participantes ha transcurredo en dos fases. En una primera etapa, siguiendo un criterio de doble estratificación en función del tipo centro de enseñanza público/privado/concertado y rural/urbano, se han concretado los centros que podrían participar en el estudio. En una segunda fase, fueron criterios circunstanciales los que determinaron los centros participantes.

Tabla 1
Modelos de invarianza factorial

1	Invarianza configural	$\Sigma_1 = \Lambda_1 \Psi_1 \Lambda_1^T + \Theta_1$	$\mu_1 = v_1$
		$\Sigma_2 = \Lambda_2 \Psi_2 \Lambda_2^T + \Theta_2$	$\mu_2 = v_2$
2	Invarianza métrica	$\Sigma_1 = \Lambda \Psi_1 \Lambda^T + \Theta_1$	$\mu_1 = v_1$
		$\Sigma_2 = \Lambda \Psi_2 \Lambda^T + \Theta_2$	$\mu_2 = v_2$
2a	Invarianza métrica con restricciones sobre Θ	$\Sigma_1 = \Lambda \Psi_1 \Lambda^T + \Theta$ $\Sigma_2 = \Lambda \Psi_2 \Lambda^T + \Theta$	$\mu_1 = v_1$ $\mu_2 = v_2$
3	Invarianza fuerte	$\Sigma_1 = \Lambda \Psi_1 \Lambda^T + \Theta_1$	$\mu_1 = v$
		$\Sigma_2 = \Lambda \Psi_2 \Lambda^T + \Theta_2$	$\mu_2 = v + \Lambda \alpha_2$
3a	Invarianza estricta con restricciones sobre Θ	$\Sigma_1 = \Lambda \Psi_1 \Lambda^T + \Theta$	$\mu_1 = v$
		$\Sigma_2 = \Lambda \Psi_2 \Lambda^T + \Theta$	$\mu_2 = v + \Lambda \alpha_2$

Instrumento

El cuestionario para la medida del autoconcepto, AFA-A (Mutsu, García y Gutiérrez, 1997), es un test de rendimiento típico compuesto por 31 ítems de respuesta ordenada (siempre, algunas veces, nunca) derivado de la teoría del autoconcepto propuesta por Shavelson, Hubner y Stanton (1976). Estos autores proponen un modelo multidimensional del autoconcepto que ha sido comprobado empíricamente en numerosos trabajos (Byrne, 1996; Byrne y Shavelson, 1996; González-Pienda y col., 1997, 2002; Marsch, 1990; Vispoel, 1995), y que en el test AFA se operacionaliza a través de cuatro dimensiones: emocional (honestidad y estabilidad), familiar (relaciones con los padres), social (relaciones con iguales) y académica.

Adaptación

La escala original fue adaptada al euskera utilizando el método de la traducción inversa (Brislin, 1970). El proceso tiene lugar en tres etapas; en la primera se traduce la prueba original al euskera; en la segunda se «retrotraduce» al castellano la versión adaptada, y, finalmente, se revisan y analizan las posibles divergencias entre las dos versiones en lengua castellana (original/retrotraducida). Las dos adaptaciones se llevan a cabo de modo independiente por traductores profesionales. En el proceso global han participado tres filólogos.

Procedimiento

La administración de las versiones original y adaptada tuvo lugar dentro del horario lectivo en cada una de las aulas de los centros participantes por personal de apoyo que fue especialmente entrenado para la ocasión.

Análisis progresivo de la invarianza factorial

El primer paso en el análisis progresivo de la invarianza consiste en la exploración de los datos. Esta fase previa es un examen inicial del modelo sustantivo, basado en los resultados de análisis factoriales exploratorios sobre cada una de las submuestras. La información obtenida será la base para la especificación del modelo de invarianza. Todos los modelos evaluados son modelos factoriales confirmatorios multi-grupo estimados sobre las matrices de medias y covarianzas por el procedimiento de máxima verosimilitud, tal y como está implementado en LISREL8.3 (Jöreskog y Sörbom, 1993). Dada la indeterminación de las escalas de las variables latentes y el hecho de que son dos grupos los que tenemos que comparar hemos fijado las unidades de medida para los factores a través de variables de referencia (ver tabla 2).

Es posible considerar los modelos analizados, como modelos anidados en los que se van añadiendo restricciones a medida que aumentan los parámetros a comparar entre grupos. El anidamiento puede adoptar dos vías en función de que se consideren o no se tenga en cuenta las varianzas de error. La primera secuencia estaría constituida por Modelo 1-Modelo 2a-Modelo 3a. La segunda cadena, por su parte, sería la formada por Modelo 1-Modelo 2a-Modelo 2b-Modelo 3a-Modelo 3b.

La evaluación del ajuste de los modelos de ecuaciones estructurales se debe de apoyar conjuntamente con el valor de Chi-cuadrado, con la información aportada por otros índices que juzgan la

bondad del modelo (Bollen y Long, 1993). Entre ellos utilizaremos el índice de ajuste general (GFI; Jöreskog y Sorbom, 1993), la raíz media cuadrática del error de aproximación (RMSEA; Hu y Bentler, 1999), el criterio de información de Akaike (AIC; Akaike, 1974) y el índice de ajuste comparativo (CFI; Bentler, 1990). Además del estudio independiente de cada uno de los modelos, es posible evaluar comparativamente el ajuste de dos modelos anidados. Para ello, siguiendo los criterios propuestos por Cheung y Rensvold (2002), nos fijamos en la diferencia entre los valores en el índice comparativo de Bentler (CFI); si el valor de la diferencia entre dos modelos anidados es superior a 0,01 en favor del modelo menos restricto, debería rechazarse el modelo con más restricciones.

El análisis progresivo consiste en un estudio secuencial que comienza con el modelo de invarianza configural, y continúa imponiendo restricciones hasta que se rechaza un modelo. En ese punto del análisis se estudia de modo individual cada uno de los ítems para descubrir cuál(es) de ellos es(son) la causa de la no-invarianza, y con qué parámetros podemos identificar ésta.

Resultados

Análisis preliminares

Las medias aritméticas y las desviaciones estándar obtenidas con la versión original son 70,94 y 6,47, y en la submuestra adaptada 68,97 y 6,52. La diferencia de medias entre ambas submuestras es estadísticamente significativa ($t= 5,867$; $p<0,0001$).

El coeficiente de fiabilidad estimado por el alpha de Cronbach es 0,714 para la submuestra original, y 0,74 para la submuestra adaptada. Su equivalencia se comprueba con el estadístico propuesto por Feldt (1969), que en nuestro caso permite aceptar la hipótesis de igualdad entre los coeficientes de consistencia interna ($w= 1$; $F_{(539,1268)}= 0,092$).

Invarianza factorial

La primera exploración factorial de los datos, además de apoyar la organización tetradimensional del autoconcepto propuesta en el cuestionario analizado, permite una evaluación tosca de la equivalencia entre estructuras factoriales a través del índice de Tucker (1951), conocido también como coeficiente de proporcionalidad. Ante la imposibilidad de la construcción de intervalos de confianza para este índice, dado que no se conoce su distribución muestral, se han propuesto varias reglas de decisión según las cuales puede aceptarse la hipótesis de equivalencia factorial si el valor obtenido es superior a 0,95 (Ten Berge, 1986). En nuestro caso, el alto índice obtenido (Índice de Congruencia= 0,986) nos permite continuar con el análisis de la invarianza.

La estimación progresiva de la invarianza comienza con el modelo de invarianza configural (Tabla 2). Los índices de ajuste obtenidos (Tabla 3) permiten aceptar la equivalencia de los modelos de medida básicos entre las dos versiones del test. Aunque el valor de Chi-cuadrado excede al exigido para aceptar la hipótesis de invarianza, el resto de índices contradicen esta conclusión. El índice de ajuste general (GFI= 0,90) y la raíz media cuadrática (RMSEA= 0,052) nos permiten aceptar el modelo base de la invarianza.

Añadiendo al modelo base restricciones sobre los coeficientes regresores caracterizamos la invarianza métrica (Modelo 2a). Los valores que se recogen en la tabla 3 permiten aceptar este nivel de

invarianza. El índice de ajuste general (GFI= 0,90) y la raíz media cuadrática (RMSEA= 0,052) siguen aportando información convergente en esta dirección. Además, el criterio de información de Akaike (AIC= 3249) y el índice comparativo de Bentler (CFI= 0,69) no sufren incrementos respecto al modelo anterior. Haciendo uso del criterio para la evaluación de los modelos anidados, la diferencia entre CFI's obtenidas entre el Modelo 2a y el Modelo 1 ($CFI_{\text{mod1-mod2}} = 0,00$) permite aceptar el modelo de invarianza métrica. Podemos concluir hasta ahora que los pesos factoriales son equivalentes en las dos submuestras.

El modelo de anidamiento en este punto del proceso puede derivarse en dos vías; por una parte, imponiendo restricciones sobre las varianzas de error/únicas (Modelo 2b), y, por otro lado, añadiendo al modelo como parámetros fijos los valores interceptales (Modelo 3a). Siguiendo la primera vía, el RMSEA adquiere un valor de 0,063, que supera el punto crítico de 0,50; el GFI, por su parte, tampoco alcanza el valor 0,90. Aunque podría aceptarse un ajuste aceptable del modelo, si evaluamos el anidamiento la conclusión debería de ser otra. La diferencia entre los índices comparativos de Bentler es superior al límite fijado ($CFI_{\text{mod2a-mod2b}} = 0,17$). Por su parte, el índice de Akaike utilizado también para la comparación de

modelos (Wicherts y Dolan, 2004) sufre un incremento considerable ($AIC_{\text{mod2b-mod2a}} = 4338-3249 = 1069$). Evaluando toda la información disponible no aceptaríamos el modelo de Invarianza Métrica con Restricciones sobre las varianzas de error/únicas.

La segunda bifurcación del modelo de invarianza métrica nos llevaría a evaluar la equivalencia entre los valores interceptales, Modelo 3a. Los índices (Tabla 3) muestran un buen ajuste de este modelo, tanto evaluado de modo independiente como analizándolo respecto a su anidamiento con el Modelo 2a de invarianza métrica. La diferencia entre los índices comparativos de Bentler es de 0,01; el índice de ajuste general es 0,90 y la raíz media cuadrática es 0,052. Aceptada la invarianza fuerte, concluimos la equivalencia entre los parámetros λ y ν ; los dos modelos evaluados son equivalentes respecto a los coeficientes factoriales y a los valores interceptales.

Queda, por tanto, evaluar el grado más alto de invarianza, la invarianza estricta (Modelo 3b) (Tablas 2 y 3). Los valores mostrados en la tabla 3 nos inducen a rechazar este modelo. Al igual que en el caso 2b, la diferencia entre los índices AIC y CFI entre los dos modelos nos permiten llegar a esta conclusión ($CFI_{\text{mod3a-mod3b}} = 0,17$; $AIC_{\text{mod3b-mod3a}} = 4528-3447 = 1081$). La información aportada por los índices RMSEA y GFI inducen a la misma valoración.

Tabla 2
Estructura factorial del cuestionario AFA-A en sus versiones original, adaptada e invariante. Modelos de invarianza 1 y 3 b

Ítem	Original	Adaptada	Invarianza estricta	
	λ	λ	λ	ν
ACADÉMICO				
05 Hago bien los trabajos escolares	1,00	1,00	1,00	2,28
07 Puedo dibujar bien	0,76	0,54	0,58	2,01
08 Soy lento en terminar los trabajos escolares	0,88	0,98	0,95	2,28
11 Hago cosas a mano muy bien	0,57	0,56	0,51	2,17
14 Con frecuencia soy voluntario en la escuela	0,68	0,71	0,37	1,90
15 Detesto la escuela	0,25	0,92	0,86	1,95
16 Mis profesores me consideran inteligente y trabajador	1,24	1,49	1,37	2,13
23 Olvido pronto lo que aprendo	0,98	0,83	0,84	2,15
26 Trabajo mucho en clase	1,01	1,11	1,05	2,26
27 Juego con mis compañeros y compañeras	0,67	0,56	0,56	2,68
EMOCIONAL				
02 Me pongo nervioso cuando me llama algún profesor	0,38	0,79	0,74	2,15
03 Cuando me porto mal en clase, me siento disgustado	0,27	0,48	0,39	1,82
04 Me desanimo cuando algo me sale mal	0,51	0,73	0,64	1,84
06 Me avergüenzo de muchas cosas que hago	0,72	0,79	0,72	2,28
09 Soy nervioso	1,00	1,00	0,93	1,92
10 Me pongo nervioso cuando hablo en clase	1,07	1,07	1,00	2,14
12 Me preocupo mucho por todo	0,43	0,42	0,39	1,77
19 Soy torpe en muchas cosas	0,48	0,38	0,38	2,19
28 Tengo miedo de algunas cosas	0,62	0,64	0,59	2,01
31 Soy honrado con los demás y conmigo mismo	0,22	0,32	0,29	2,48
SOCIAL				
01 Es difícil para mí mantener amigos	1,05	0,92	0,93	2,73
13 Me gusta mi forma de ser	1,27	1,21	1,26	2,57
17 Tengo muchos amigos	1,00	1,00	1,00	2,84
18 Soy un chico alegre	0,95	1,14	1,09	2,64
24 Consigo fácilmente amigos	1,04	1,29	1,22	2,43
FAMILIAR				
20 Me gustan las peleas y riñas	0,82	0,69	0,78	2,64
21 Mi familia está decepcionada de mí	1,00	1,00	1,00	2,79
22 Soy criticado en casa	0,84	1,00	0,87	2,61
25 Pierdo mi paciencia fácilmente	0,83	0,60	0,74	2,02
29 Me enfado si los demás no hacen lo que yo digo	0,70	0,55	0,65	2,67
30 Soy violento con mis amigos y familiares	0,65	0,59	0,73	2,70

Del total de modelos evaluados únicamente son los modelos que especifican equivalencia entre las varianzas de error/únicas los que no alcanzan un buen ajuste (Modelo 2b y Modelo 3b).

En el estudio progresivo de la invarianza factorial pasamos a un estudio detallado de los ítems que nos permita extraer información sobre el origen de la no-invarianza. Dado que el problema está concentrado en los errores de medida de las variables observadas fijamos en ellos nuestra atención. En el modelo de invarianza estricta, los errores de medida tienen una media aritmética de 0,32 y una desviación típica de 0,11. Del total de variables analizadas ($n=31$), son dos (ítem 8 e ítem 21) los que tienen valores más extremos, 0,56 y 0,72, respectivamente. Con esta información volvemos a evaluar el modelo de invarianza estricta relajando el criterio de igualdad de varianzas de error para estas dos variables. Este análisis se efectúa de modo individual para cada uno de los ítems con el fin de comprobar su efecto sobre el modelo completo. Después se procederá a estudiar secuencialmente el modelo de equivalencia estricta eliminando cada vez un ítem hasta que el ajuste sea aceptable.

Evaluamos primero el modelo sin restricción sobre la varianza de error para la variable 21, después se reespecifica el modelo para la variable 8 y finalmente se relajan los supuestos simultáneamente para los dos ítems. Los índices correspondientes a estos modelos aparecen en la tabla 3. En la primera línea leemos la información referida al ítem 21. El modelo sufre una mejora respecto al modelo de invarianza estricta (RMSEA= 0,055; GFI= 0,90), sin embargo, la diferencia entre los índices de Bentler es mayor al punto de corte fijado de 0,01 ($CFI_{\text{mod21-mod3a}} = 0,03$). Si eliminamos del modelo el ítem 8, la mejora obtenida en el ajuste es menor que en caso anterior; este hecho evidencia una influencia mayor del ítem 21 sobre el modelo. El análisis del modelo en el que se liberan los parámetros para los dos ítems ofrece un buen ajuste (Tabla 3). Los valores de RMSEA (0,052), GFI (0,90) o CFI (0,67) permiten aceptar la hipótesis de equivalencia. Nótese además que la diferencia entre CFI de este modelo y el modelo de invarianza fuerte es de 0,01, y entre los índices AIC es de 45.

Como consecuencia puede considerarse que las versiones original y adaptada del cuestionario AFA-A tienen una estructura factorial que se ajusta al modelo de invarianza factorial estricta para todos los ítems excepto para dos. Estos últimos alcanzan la invarianza factorial fuerte.

Discusión

La posibilidad de comparar las medidas obtenidas en situaciones diferentes es un prerrequisito de toda ciencia (Dorans y Holland, 2000) que exige el cumplimiento de la invarianza métrica.

En la medición mediante tests su evaluación se operacionaliza a través del análisis de una relación funcional entre el espacio latente y sus indicadores manifiestos, que puede ser lineal o no-lineal. Desde el primer acercamiento, el estudio de la invarianza se ha limitado en muchas ocasiones al cálculo de un coeficiente de congruencia factorial basado únicamente en la información contenida en las saturaciones factoriales. Este procedimiento de análisis de la invarianza factorial es inadecuado porque no tiene en cuenta la totalidad de los parámetros que la definen, y porque en caso de rechazar la invarianza no aporta información alguna sobre el origen de ésta.

El modo de operar mostrado en este trabajo parte de la relación funcional lineal entre el espacio latente y las variables observadas y analiza progresivamente por medio de modelos factoriales confirmatorios multigrupo todos los parámetros que definen el modelo de medida: coeficiente regresor, valor interceptal y componente de error/unicidad. La evaluación progresiva es global en una primera etapa, e individualizada después. La secuencia del análisis propuesto va añadiendo información sobre el tipo de invarianza alcanzado en función de los parámetros comparados, a la par que permite un acercamiento centrado en el ítem. Evidentemente esta aproximación sólo cobra sentido si existe un grado de invarianza mínimo entre grupos. En estas circunstancias el análisis secuencial y progresivo de los ítems permite al investigador profundizar en las semejanzas/divergencias entre grupos, dado que el foco de la invarianza puede relacionarse sólo con algunos indicativos del factor (ítems) y no con todos ellos.

La invarianza configural es la forma más elemental de invarianza, y el modelo básico de análisis en todo estudio de equivalencia. Únicamente exige que las dimensiones latentes queden especificadas por las mismas variables manifiestas en las dos subpoblaciones. El rechazo de la hipótesis de invarianza configural implica la falta de equivalencia sustantiva de constructos entre subpoblaciones o culturas. El segundo nivel de invarianza examina, además, la igualdad de coeficientes regresores. Dentro del nivel de invarianza métrica es posible todavía aumentar el nivel de equivalencia restringiendo las matrices de varianzas/covarianzas de unicidad/error en las dos subpoblaciones. La invarianza métrica permite la comparación entre varianzas factoriales y pendientes regresoras. Si la equivalencia configural ha sido demostrada, pero no podemos admitir la invarianza métrica, es posible evaluar sucesivamente cada una de las variables observadas (ítems) para estudiar la causa del desajuste. En este nivel de análisis las diferencias entre los coeficientes regresores podría interpretarse como la diferencia en la relevancia de cada uno de los ítems en el factor que definen. Este análisis progresivo aportaría información sobre las diferencias entre los grupos respecto a cada una de los ítems in-

Tabla 3
Índices de bondad de ajuste de los modelos de invarianza

Modelo	g.l.	χ^2	RMSEA	GFI	AIC	CFI
1 Invarianza configural	867	2837	0,052	0,90	3274	0,69
2.a. Invarianza métrica	888	2859	0,052	0,90	3249	0,69
2.b. Invarianza métrica con restricción sobre Θ	918	3985	0,063	0,89	4338	0,52
3.a. Invarianza fuerte	915	2996	0,052	0,90	3447	0,68
3.b. Invarianza estricta	945	4117	0,063	0,89	4528	0,51
Sin restricciones sobre Θ en ítem 21	944	3284	0,055	0,90	3703	0,64
Sin restricciones sobre Θ en ítem 8	944	3930	0,061	0,89	4317	0,54
Sin restricciones sobre Θ en ítems 8 y 21	943	3097	0,052	0,90	3492	0,67

dicadores del constructo, que una prueba de ajuste global no proporcionaría. En el nivel más alto de invarianza se examina la igualdad en el origen de las variables latentes, es decir, la equivalencia entre los valores interceptales de los factores. A diferencia de los modelos anteriores, este nivel de invarianza requiere la especificación del vector de medias. Al igual que en el caso de la invarianza métrica, es posible definir dos modelos anidados de invarianza en función de que se impongan restricciones sobre la matriz de varianzas/covarianzas de error. La equivalencia interceptal se traduce en la paridad entre los orígenes de la escala de medida de las variables latentes; condición necesaria para la comparación de las medias latentes.

El estudio de la invarianza desde la perspectiva lineal se aproxima así a la evaluación del funcionamiento diferencial del ítem propuesta desde los modelos de respuesta al ítem, que centran su atención en las funciones no-lineales ítem/espacio latente. Estos últimos han demostrado su efectividad en distintos ámbitos psicométricos entre los que podemos citar la adaptación de pruebas (Elosua y López, 1999; Elosua, López y Torres, 1999; Elosua, López y Egaña, 2000a), uso de tests en contextos bilingües (Elosua y col., 2000; Ferrers, González y Gómez, 2002) o análisis de sesgos relacionados con la tarea (Elosua, López y Egaña, 2000b). Son procedimientos diseñados originalmente para pruebas unidimensionales, que todavía no han sido generalizados para su uso con escalas multidimensionales. De ahí la relevancia de contar con pro-

cedimientos como el propuesto en este trabajo para este tipo de tests que permiten un análisis individualizado, detallado y progresivo de todos y cada uno de los ítems.

Asumiendo ciertas similitudes entre ambos modelos (Elosua y López, 2002; Lord y Novick, 1968; McDonald, 1999) sería posible contraponer los interceptos factoriales con los parámetros de dificultad del ítem y los coeficientes regresores con los parámetros de discriminación. La equivalencia entre parámetros de discriminación daría lugar a la invarianza métrica, y la igualdad entre parámetros de dificultad sería la condición de la invarianza factorial.

Si bien quedan todavía por llevar a cabo estudios de simulación que pongan a prueba la efectividad del procedimiento propuesto bajo diferentes condiciones, creemos que la evaluación progresiva de la invarianza factorial es fundamental en el estudio de la equivalencia entre puntuaciones porque permite extraer información sobre cada uno de los parámetros del modelo de medida y posibilita un estudio en profundidad de cada uno de los ítems que componen el test. Sólo los resultados de un análisis de invarianza sentarán la base psicométrica para la justificación de las comparaciones entre grupos.

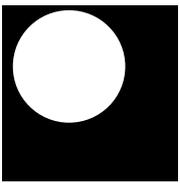
Agradecimientos

Trabajo financiado por la Dirección General de Investigación del Ministerio de Ciencia y Tecnología. BSO2002-00490.

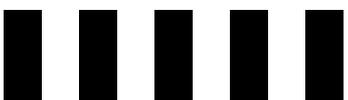
Referencias

- Akaike, H. (1974). A new look at statistical model identification. *Transactions on Automatic Control*, *AC-19*, 716-723.
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.
- Bollen, K.A. y Long, J.S. (1993). Introduction. En K.A. Bollen y J.S. Long (Eds.): *Testing structural equation models* (pp. 1-9) Newbury Park, CA: Sage.
- Bloxom, B. (1972). Alternative approaches to factorial invariance. *Psychometrika*, *37*, 425-440.
- Brislin, R.W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185-216.
- Byrne, B.M. (1996). *Measuring self-concept across the lifespan: issues and instrumentation*. Washington, DC: APA.
- Cheung, G.W. y Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.
- Dorans, N.J. y Holland, P.W. (2000). Population invariance and equitability of tests. *Journal of Educational Measurement*, *37*, 281-306.
- Ellis, J.L. (1993). Subpopulation invariance of patterns in covariate matrices. *British Journal of Mathematical and Statistical Psychology*, *46*, 231-254.
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, *15*(2), 315-321.
- Elosua, P. y López, A. (1999). Funcionamiento diferencial de los ítems y sesgo en la adaptación de dos pruebas verbales. *Psicológica*, *20*, 23-40.
- Elosua, P. y López, A. (2002). Indicadores de dimensionalidad para ítems binarios. *Metodología de las Ciencias del Comportamiento*, *4*(1), 121-137.
- Elosua, P., López, A. y Egaña, J. (2000a). Idioma de aplicación y rendimiento en una prueba de comprensión verbal. *Psicothema*, *12*(2), 201-206.
- Elosua, P., López, A. y Egaña, J. (2000b). Fuentes potenciales de sesgo en una prueba de aptitud numérica. *Psicothema*, *12*(3), 376-382.
- Elosua, P., López, A., Egaña, J., Artamendi, J.A. y Yenes, F. (2000). Funcionamiento diferencial de los ítems en la aplicación de pruebas psicológicas en entornos bilingües. *Metodología de las Ciencias del Comportamiento*, *2*(1), 17-33.
- Elosua, P., López, A. y Torres, E. (1999). Adaptación al euskera de una prueba de inteligencia verbal. *Psicothema*, *11*(1), 151-161.
- Feldt, L.S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two test. *Psychometrika*, *34*, 363-373.
- Ferrres, D., González, V. y Gómez, J. (2002). Funcionamiento diferencial de los ítems en una situación de contacto de lenguas. *Psicothema*, *14*(2), 483-490.
- Hu, L. y Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.
- González-Pienda, J.A., Núñez, J.C., Álvarez, L., González-Pumariega, S., Roces, C., González, P., Muñoz, R. y Bernardo, A. (2002). Inducción parental a la autorregulación, autoconcepto y rendimiento académico. *Psicothema*, *14*(4), 853-860.
- González-Pienda, J.A., Núñez, J.C., González-Pumariega, S. y García, M. (1997). Autoconcepto, autoestima y aprendizaje escolar. *Psicothema*, *9*(2), 271-289.
- Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.
- Jöreskog, K.G. y Sörbom, D. (1993). *LISREL 8: user's guide*. Chicago: Scientific Software International.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: LEA.
- Lord, F.M. y Novick, M.R. (1968). *Statistical Theories of Mental Tests Scores*. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Lubke, G.H., Dolan, C.V., Kelderman, H. y Mellenbergh, G.J. (2003). Weak measurement invariance with respect to unmeasured variables: an implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology*, *56*, 231-248.
- Marsch, H.W. (1990). Causal ordering of academic self-concept and academic achievement: a multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*, 646-656.
- McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Mellenbergh, G. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13, 127-143.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29, 177-185.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Musitu, G., García, F. y Gutiérrez, M. (1997). *AFA. Autoconcepto Forma-A*. Madrid: TEA.
- Shavelson, J., Hubner, J.J. y Stanton, G.C. (1976). Self-concept: validation of construct interpretations. *Review of Educational Research*, 46, 407-442.
- Ten Berge, J.M.F. (1986). Some relations between descriptive comparisons of components from different studies. *Multivariate Behavioral Research*, 21, 29-40.
- Tucker, L.R. (1951). *A method for synthesis of factor analysis studies*. Personnel Research Section Report N.º 984. Washington, DC: Department of the Army.
- Vispoel, W.P. (1995). Self-concept in artistic domains: an extension of the Shavelson, Hubner and Stanton (1976) model. *Journal of Educational Psychology*, 87(1), 134-153.
- Wicherts, J.M. y Dolan, C.V. (2004). A cautionary note on the use of information fit indexes in covariance structure modelling with means. *Structural Equation Modeling*, 11(1), 45-50.



Psicothema



Enviar en sobre a:
Colegio Oficial de Psicólogos del Principado de Asturias
Ildefonso Sánchez del Río, 4 - 1º B
33001 OVIEDO (Spain)



HAY NUMEROSAS RAZONES PARA NO PERDERSE EL 9º CONGRESO EUROPEO DE PSICOLOGÍA

- Los mejores investigadores y profesionales europeos y de otras partes del mundo presentarán los avances más recientes en sus campos de especialización.
- Un Congreso en el que la ciencia psicológica básica y las aplicaciones profesionales van de la mano, enriqueciéndose mutuamente.
- Una oportunidad única para contactar personalmente con los mejores expertos internacionales en tu campo de trabajo.
- Todos los campos científicos y profesionales están contemplados en el Congreso, ningún tema ni enfoque se excluye si viene avalado por la calidad.
- Te invitamos a presentar tus trabajos: Póster, Comunicaciones o Simposios. El Congreso Europeo en Granada, organizado por el Colegio Oficial de Psicólogos, bajo los auspicios de la EFPA, constituye un excelente foro para mostrar el gran nivel alcanzado por la Psicología española, tanto desde el punto de vista científico como profesional. Puedes hacerlo en inglés o en español, tú eliges el idioma.
- Y si aún dudases, no lo creemos, recuerda Granada, esa ciudad en la que, como bien dijo Lorca, ... «las horas son más largas y sabrosas que en ninguna otra ciudad de España...».

