

LA INVARIANZA EN LA ESTRUCTURA FACTORIAL DEL RAVEN EN GRUPOS DE DELINCUENTES Y NO DELINCUENTES

Eliseo Chico Librán
Universidad de Tarragona

Este artículo analiza el rendimiento obtenido en el test de las Matrices Progresivas de Raven por dos grupos, uno de delincuentes y otro de no-delincuentes, ambos con un nivel cultural y educacional similar. La comparación entre ambos grupos muestra unas puntuaciones directas más bajas en el grupo de delincuentes. Realizado un análisis factorial confirmatorio multigrupo para evaluar si la estructura factorial del test es esencialmente la misma en ambos grupos, los resultados sugieren que las diferencias encontradas en las puntuaciones son ampliamente debidas a las diferencias correspondientes en el factor común que el test intenta medir. Sin embargo, la hipótesis de que el test no muestra funcionamiento diferencial en absoluto, no se puede probar.

Invariance in factorial structure of Raven test in delinquent and non delinquent groups. This paper analyzes the performance of a delinquent and a non-delinquent group of young males with a similar educational level in the Raven Standard Progressive Matrices test. The comparison of the direct scores shows a lower performance for the delinquent group. A multigroup confirmatory factor analysis is used to assess if the factor structure of the test is essentially the same in both groups and the results suggest that the differences found in the observed scores are largely due to corresponding differences in the common factor that the test intends to measure. The hypothesis that the test is totally unbiased, however, cannot be proved.

Hay bastante evidencia empírica que demuestra una consistente y negativa relación entre inteligencia general y conducta criminal. Esta relación entre ambas variables se establece entre un rango de $-.20$ a $-.30$ (Eysenck y Gudjonsson, 1989) o una diferencia en CI de 8 puntos entre delincuentes y población general (Hirschi y Hindelang, 1977).

Se han propuesto varias teorías para explicar esta relación. En concreto, se van a revisar dos corrientes teóricas claramente opuestas.

La primera corriente asume que el CI bajo es un factor a tener en cuenta en la conducta antisocial (ver p. e. Gordon, 1986). Desde este punto de vista, por lo tanto, se espera que en el factor "g" los grupos de delincuentes tengan un nivel más bajo que la población no-criminal (Jensen y Faulstich, 1988).

La segunda corriente asume que el peor rendimiento mostrado por los delincuentes

Correspondencia: Eliseo Chico Librán
Universidad «Rovira i Virgili»
Departamento de Psicología
Carretera de Valls, s/n. 43007 Tarragona (Spain)

en los tests de inteligencia es, en gran parte, debido a las diferencias socio-culturales (West y Farrington, 1973) Hay una crítica muy extendida que señala que los tests de inteligencia, especialmente, aquellos en los que predominan ítems verbales, tienen un gran sesgo cultural y perjudican a los grupos socialmente menos favorecidos (p. e. los delincuentes). De hecho, el estudio del sesgo cultural fue lo que impulsó los estudios sobre el posible sesgo en los tests (Camilli, 1994). Y desde el inicio de estos estudios los resultados indican que no existe evidencia firme de la influencia del sesgo cultural en los tests (Jensen, 1980)

Se han propuesto diferentes perspectivas para explicar el sesgo que se produce en los tests mentales dentro de la población delinencial. Sin embargo, ninguno de ellos se puede considerar completamente satisfactorio. Voy a analizar, aunque sea brevemente, dos posturas que se pueden utilizar conjuntamente: a) comparar las puntuaciones de grupos delincuentes y no-delincuentes, pertenecientes ambos grupos a parecidos niveles socio-culturales, y b) comparar las puntuaciones de delincuentes y no-delincuentes en tests de baja influencia cultural.

En cuanto a la primera perspectiva, la evidencia empírica muestra, en general, que las diferencias en CI entre grupos delincuentes y no-delincuentes se mantienen dentro de las diferentes categorías de clase social (Hirschi y Hindelang, 1977; Díaz, Beleña y Báquena, 1994). En este mismo sentido, los estudios de McGarvey, Gabrielli, Bentler y Mednick (1981), estudios que están en la línea de investigación de la influencia del sesgo cultural en el posible sesgo de los tests, encontraron una constante correlación negativa entre CI y delincuencia después de eliminar la influencia del status socio-económico.

En lo referente a tests de baja influencia cultural, algunos estudios que han usado el

test de las Matrices Progresivas de Raven han mostrado que los grupos de delincuentes también tienden a conseguir puntuaciones más bajas que la población general en estos instrumentos no verbales de baja influencia cultural (West, 1969). No obstante, estos resultados se han atribuido también a la falta de motivación en la situación de prueba (West, 1969) o a que el test de Raven presenta procesos de razonamiento mediatizados por el lenguaje (Trasler, 1973), e igualmente, en algún sentido, también se puede considerar como un instrumento sesgado.

En esta investigación se pretende determinar si la interpretación de las puntuaciones en el test de Raven es igualmente válida para el grupo de delincuentes que para el grupo de no-delincuentes. Es decir, si existe invarianza en la estructura factorial del test. La no invarianza sugeriría un funcionamiento diferencial de dicho test a nivel de series. Este objetivo general se intenta desarrollar en dos etapas: a) un estudio sobre si hay diferencias substanciales en la puntuación obtenida por un grupo de delincuentes y otro no-delincuentes. De acuerdo con la evidencia empírica revisada, se espera una puntuación superior en el grupo no-delincente; b) una valoración sobre si la estructura factorial del test, a nivel de conjunto de ítems múltiples, es invariante en ambos grupos. Si la estructura factorial es invariante, entonces las diferencias grupales en las puntuaciones se pueden atribuir a diferencias equivalentes en el constructo que el test intenta medir. Por otra parte, si la invarianza de la estructura factorial no se puede probar, entonces cabe pensar que el test está sesgado respecto a la posible validez diferencial de la interpretación de las puntuaciones aportadas por el test.

El modelo de medida extendido de Lisrel.

La principal finalidad de la investigación es estudiar la invarianza factorial de

la estructura del test en dos grupos. Tal invarianza se puede valorar detalladamente usando el modelo de análisis factorial confirmatorio (AFC) ofrecido por Sörbom (1974). Conviene hacer un breve comentario sobre este modelo de cara a interpretar los resultados de la presente investigación.

Generalmente, se asume que las 5 serie del Raven miden un sólo factor común (g). Así, el modelo factorial considerado aquí es el modelo de Spearman con errores no correlacionados.

El grupo total de sujetos, en el que se espera que se cumpla el modelo de Spearman, se divide en dos grupos (delincuentes y no-delincentes); cada grupo se considera como una muestra procedente de diferente población, y, para cada población (g), se asume el modelo siguiente:

$$(1) \chi_i^{(g)} = \mu_i + \lambda_i f + e_i, \quad i = 1, 2, 3, 4, 5 \text{ (series)}$$

Es decir, la puntuación observada en cada serie para un sujeto perteneciente al grupo "g" se considera que está en parte, explicada por la puntuación factorial común de este sujeto, "f", y en parte, por el error de medida "e". El término " μ_i " es el intercepto (porque el modelo trabaja en puntuaciones directas) y " λ_i " es la carga factorial. El modelo no sesgado asume que tanto " μ_i " como " λ_i " son invariantes en las dos poblaciones comparadas (delincuentes y no delincuentes). Si esto es así, bajo los supuestos habituales acerca del error de medición, el vector de las medias en las 5 series, para una población particular, viene expresado por:

$$(2) \quad \bar{\chi}^{(g)} = \mu + \lambda \bar{f}^{(g)}$$

Donde $\bar{f}^{(g)}$ es la media de las puntuaciones factoriales en la población "g" y μ y λ son el intercepto y vectores de cargas factoriales, respectivamente.

Además, bajo los supuestos habituales en AF, la matriz varianza-covarianza en cada población viene dada por:

$$(3) \quad \Sigma^{(g)} = \gamma \wedge \sigma^2^{(g)} \wedge' + \Psi^{2(g)} \quad (\Psi^{2(g)} \text{ diagonal})$$

Si este modelo se cumple, a partir de (2) y (3) se puede asumir que todas las diferencias encontradas entre las medias obtenidas y las covarianzas en los dos grupos son exclusivamente debidas al hecho de que las puntuaciones factoriales tienen diferentes medias ($\bar{f}^{(g)}$) y varianzas ($\sigma^{2(g)}$) en estos grupos. En otras palabras, todas las diferencias encontradas en las puntuaciones obtenidas en el test se podrían explicar por las diferencias en el factor que el test pretende medir.

Sin embargo, si en una serie en concreto, se estiman diferentes interceptos y/o cargas en los dos grupos, esta serie se puede considerar que está sesgada. Un ejemplo, consideremos dos grupos g_1 y g_2 con interceptos estimado μ_1 y μ_2 y un sujeto por cada grupo con la misma puntuación factorial, pongamos cero. De acuerdo con la fórmula (2) la puntuación esperada sería μ_1 para el sujeto del grupo 1 y μ_2 para el sujeto del grupo 2; es decir, se podrían esperar puntuaciones diferentes en el test para sujetos con la misma puntuación factorial (p. e. nivel g), pero que pertenecen a diferentes grupos. Este es un ejemplo del llamado "sesgo uniforme de test" (Mellenberg, 1989). Cuando no solamente el intercepto, sino también la carga factorial es distinta, las diferencias esperadas en las puntuaciones obtenidas por los sujetos con la misma puntuación factorial, pero que pertenecen a diferentes grupos, varían en diferentes niveles del factor. Esto es lo que se llama "sesgo no uniforme del test" (Mellenberg, 1989). Se debería señalar que, en ambos casos, las diferencias en las puntuaciones obtenidas se podrían atribuir a otras causas que no sean las diferencias en el constructo que se está midiendo.

Método

Instrumentos

En este trabajo se utilizó el test de las Matrices Progresivas de Raven, adaptado el Castellano (Murga, 1970). La fiabilidad del test fluctúa, en función de la edad, de .83 a .93. Tiene una correlación con la Escala de Terman-Merrill de .86 y se ha encontrado una saturación en el factor "g" de .82. En las muestras utilizadas en este estudio, se ha encontrado una fiabilidad de .92 en el grupo de delincuentes y .88 en el grupo de soldados. La utilización de este test para esta investigación se debe a tres razones: a) porque es una medida generalmente aceptada del factor "g"; b) porque se basa en analogías perceptuales y razonamiento inductivo y es considerado por muchos expertos como uno de los tests de menor carga cultural que miden "g" (ver Jensen, 1980); y c) porque fue ideado para medir un constructo unidimensional y, factorialmente, parece medir un único constructo y poca cosa más (Burke, 1958). El test se divide en cinco series de 12 ítems cada una, y dentro de cada serie, los ítems van de menor a mayor dificultad.

Los programas estadísticos utilizados en los análisis fueron el SPSS-PC, Lisrel-8 y Prelis-2.

Participantes y procedimiento

El grupo de delincuentes estaba formado por 942 internos del Centro Penitenciario de Tarragona. Todos los internos tenían nivel de estudios primarios y buena lectura comprensiva. El nivel de estudios del grupo se obtuvo por información de los profesores de E.G.B. del Centro.

Los participantes del grupo de referencia fueron 585 soldados que cumplían el servicio militar en Tarragona. De este grupo se seleccionó 400 soldados que tenían

un nivel educativo similar al de los internos del Centro Penitenciario. El nivel de estudios de este grupo se obtuvo mediante petición de datos académicos cuando se administró el test.

El rango de edad de los internos fue de 17-30 años (media = 24.9), mientras que el rango de edad de los soldados fue de 18-23 años (media = 19.1)

Análisis y Resultados

Diferencias en el rendimiento de los grupos

En primer lugar, se obtuvieron los descriptivos del total del test y de las 5 series en ambos grupos. Las puntuaciones medias en el total del test fueron 39.26 para los delincuentes y 46.83 para los soldados. En la Tabla 1 se presentan las medias, varianzas, asimetrías y curtosis obtenidas por los dos grupos, así como los valores "t" y P para cada serie.

Tabla 1
Análisis de datos estadísticos

Serios	1	2	3	4	5
Medias:					
Total	11.26	9.63	7.97	8.24	3.82
Soldados	11.40	1.56	9.26	9.25	5.43
Delincuentes	11.21	9.31	7.53	7.90	3.27
Diferencia	0.19	1.25	1.73	1.35	2.16
Valor t	-2.95	-8.49	-10.54	-7.85	-12.54
Valor p	0.00	0.00	0.00	0.00	0.00
Varianza:					
Total	0.99	5.59	7.21	7.58	8.14
Soldados	0.76	3.28	4.01	3.52	7.33
Delincuentes	1.05	5.99	7.54	8.51	7.25
Asimetría:					
Total	-2.03	-1.06	-0.91	-1.33	0.54
Soldados	-2.28	-2.27	-1.65	-2.16	-0.08
Delincuentes	-1.95	-0.81	-0.73	-1.11	0.82
Kurtosis:					
Total	5.98	0.42	0.18	1.26	-0.63
Soldados	7.75	7.31	4.40	6.91	-0.81
Delincuentes	5.51	-0.25	-0.28	0.52	-0.03

Una simple inspección ocular de estos datos sugiere que hay un incremento en dificultad en las sucesivas series, excepto en la 3 y 4 (de hecho, en el grupo delincente la serie 4 parece ligeramente más fácil que la serie 3). A un nivel de diferencia de grupo, parece existir una diferencia general en el rendimiento observado, una diferencia que aumenta a medida que va aumentando la dificultad en las series.

A fin de valorar la tendencia señalada más arriba, se llevó a cabo un diseño factorial mixto común usando las cinco series como niveles intrasujetos. Se analizó el diseño con el procedimiento MANOVA en el paquete SPSS. Los resultados se presentan en la Tabla 2.

Tabla 2 Estadísticos univariados y multivariados	
Fuentes de variación	F Univariada
Intersujeto	134.85 (P= .000)
Intrasujeto	2384.06 (P= .000)
Inter. grupo x series	47.32 (P= .000)
Fuentes de variación	Estadís. Multivariados
Intrasujeto	Pillais: .08486 F aprox. 1814.68 (P= .000)
Inter. grupo x series	Pillais: .1171 F aprox. 42.93 (P= .000)
Nota: Los tests multivariados en este diseño dan todos las mismas F aproximadas, por lo que se da solamente el primero de Pillais.	

Aunque las matrices varianza-covarianza en los dos niveles del factor intersujetos no fueron homogéneas, aún usando las correcciones más conservadoras, los tres efectos: a) intrasujetos; b) intersujetos; y c) interacción grupos x series, fueron claramente significativas, tal como se puede ver en la tabla 2. Se esperaban estos resultados simplemente de la inspección ocular de los datos.

Así, a este nivel, se podría suponer que: a) hay diferencias generales en el rendimiento observado de los grupos delincentes y no delincentes (efecto intersujetos); b) hay una gradación general en la dificultad de las series del Raven para los dos grupos (efecto intrasujeto) y c) las diferencias en comportamiento entre los grupos delincentes y no delincentes aumenta a medida que aumenta la dificultad de las series (efecto interacción).

Comparación de estructuras factoriales

Consideraciones preliminares

El estudio de las distribuciones de las series en cada grupo mostró que en ningún caso se podían considerar estrictamente normales. Los tests univariados y multivariados de asimetría y curtosis mostraron desviaciones significativas respecto al supuesto de normalidad.

De acuerdo con estudios similares (e. j. Muthen y Kaplan, 1985) esta desviación de la normalidad se puede pensar que produce sesgo en el test de bondad de ajuste si se usa el método de estimación de máxima verosimilitud de estimación. Para corregir este problema, se escogió el estimador "Asymptotic Distribution Free" (ADF) para la estimación de este modelo (ver Bollen, 1989).

El modelo AFC de Sörbom solamente proporciona diferencias en la media y en la varianza de las puntuaciones factoriales estimadas en los diferentes grupos que son evaluados, pero la métrica para el factor es indeterminada. Para obtener una solución, se tuvo que arbitrar una serie de restricciones para fijar una escala para el factor que es común para los dos grupos. En este artículo se usó la restricción más usual que consiste en fijar el valor de la primera carga a uno de los dos grupos y fijar la media de las puntuaciones factoriales en cero en

el otro grupo. (ver Jöreskog y Sörbom, 1989 para más detalle).

Ajuste del modelo

La evaluación del ajuste para los diferentes modelos examinados se basaron en los siguientes índices: a) valor mínimo de la Función de discrepancia, b) raíz media cuadrática del error de aproximación (RMSEA), c) índice de validación cruzada (ECVI), d) test chi-cuadrado de ajuste exacto y e) test de ajuste aproximado. El procedimiento de investigación basado en estos índices se describe con detalle en Browne y Cudeck (1993).

Los resultados del ajuste para los diferentes modelos examinados se resumen en la Tabla 3.

por las restricciones antes descritas, en este modelo los interceptos y las cargas factoriales eran libres de variar en los grupos.

Una inspección de los índices de ajuste en la table 3 muestra que el modelo de Línea base es admisible así, por el momento, se puede asumir que, a nivel de series, el test se comporta unidimensionalmente en ambos grupos. En segundo lugar, se ajustó totalmente un modelo factorial común invariante. Este modelo está anidado en el modelo de Línea Base porque se podía obtener a partir de aquel restringiendo algunos parámetros del modelo de Línea base, de forma que los valores fuesen los mismos en ambos grupos. La hipótesis en este caso fue que, no solamente el test es unidimensional en ambos grupos, sino que también la estructura factorial es la misma.

Como muestra la tabla 3, el ajuste completo para este modelo no se puede considerar aceptable. Además, usando la propiedad de que la diferencia en ajuste entre dos modelos anidados se distribuye asimismo como un chi-cuadrado con grados de libertad igual a la diferencia en los grados de libertad para los dos modelos, si se compara el modelo Línea base y el modelo de plena invariancia, la diferencia es: Chi-cuadrado (8) = 129.01, p=0.000, que indica que las restricciones impuestas son inaceptables. Finalmente, se examinó un modelo invariante parcial. Se obtuvo del modelo de plena invariancia, relajando las restricciones de igualdad en las cargas factoriales para las series 4 y 5. La tabla 3 muestra, no solamente que el ajuste completo de este modelo es aceptable, sino también que su ajuste relativo es mejor que el modelo de Línea base (como se muestra por el RMSEA, ECVI, y el test de ajuste aproximado). Finalmente, la diferencia del test no es significativa: Chi-cuadrado (6) = 7.21, p=0.30. Basado en estos resultados, se decidió que este modelo proporciona un

Tabla 3
Resumen de análisis

Modelo	d.f.	Fs	RMSA	ECVI
Línea Base	10	.038	.055	.082
Invar. Plena	18	.130	.082	.170
Invar. Parcial	16	.043	.044	.079

Modelo	χ^2	pTEF	pTCF	Decisión
Línea Base	50.52	.000	.27	Aceptado
Invar. Plena	179.53	.000	.00	Rechazado
Invar. Parcial	57.53	.000	.77	Aceptado

Nota: D.F. = grados de libertad; Fs= valor mínimo de la función de discrepancia; RMSEA= raíz media cuadrática del error de aproximación; ECVI= índice de validación cruzada; χ^2 = test de Chi cuadrado de ajuste exacto; PTEF= probabilidad en relación al rest de ajuste exacto; pTCF= probabilidad en relación al test de ajuste aproximado.

El primer modelo en ser examinado fue el modelo de Línea base. Se basó en la hipótesis de que solamente un factor común subyace en el rendimiento en las cinco series del Raven en ambos grupos. Excepto

ajuste razonablemente parsimonioso para estos datos. Los parámetros estimados correspondientes al modelo invariante parcial se muestran en la Tabla 4.

<i>Tabla 4</i> Parámetros estimados para el modelo aceptado		
1a. Interceptos estimados y cargas factoriales		
Interceptos y errores tip. Delinquent. No delinquent.		
Series Raven		
1	11.17 (0.03)	11.17 (0.03)
2	9.32 (0.08)	9.32 (0.08)
3	7.55 (0.02)	7.55 (0.02)
4	7.91 (0.02)	7.91 (0.02)
5	3.29 (0.02)	3.29 (0.02)
Cargas factor y errores tip. Delinquent. No delinquent.		
Series Raven		
1	1.00 (----)	1.00 (----)
2	3.76 (0.25)	3.76 (0.25)
3	5.16 (0.36)	5.16 (0.36)
4	5.04 (0.37)	3.96 (0.29)
5	4.03 (0.33)	6.83 (0.47)
1b. Medias y varianzas estimadas de las puntuaciones factoriales		
	Delincente	No delincente
Media estimada e intervalo confidencial de 95%	0.00 (--)	0.33 (0.33, 0.41)
Varianza estimada	0.22	0.10

La primera parte de la tabla hace referencia a los parámetros estimados correspondientes a las series del Raven, es decir, interceptos y cargas factoriales. Los interceptos dan información sobre el nivel medio del rendimiento examinado. La forma más fácil de interpretar un intercepto es considerarlo como la puntuación espe-

rada en la serie correspondiente para un sujeto con nivel 0 en la variable latente (o sea, la media fijada para el grupo delincuente). Como muestra la tabla 4, los interceptos son valores decrecientes excepto para las series 3 y 4, que señalan un incremento en la dificultad de las series. Este incremento es estadísticamente significativo, como se puede ver por los intervalos confidenciales establecidos usando los errores típicos. Finalmente, la invarianza en los interceptos para ambos grupos se puede interpretar como un aumento igual en la dificultad de las series a un nivel de puntuación factorial de cero. Como en el caso de las puntuaciones observadas, tal vez el resultado más remarcable aquí es la similar dificultad de las series 3 y 4.

Las cargas factoriales se pueden interpretar como el incremento medio en la puntuación conjunta observada por unidad de incremento de la puntuación factorial. El resultado más importante aquí es la gran diferencia en las cargas factoriales estimadas para la serie quinta de los grupos delinquentes y no delinquentes. La interpretación conceptual de esta diferencia es que, al incrementarse los niveles factoriales, la puntuación observada media en la serie quinta es más alta para los sujetos no delinquentes que para los sujetos delinquentes. A un nivel más teórico (y especulativo), se puede pensar que a niveles más altos de g los sujetos no delinquentes tienden a rendir mejor que los delinquentes en la serie de más dificultad.

El resultado descrito más arriba podría ser atribuido al aburrimiento o la falta de motivación en la situación de prueba en el caso de los delinquentes; si es así, se podrían esperar respuestas azarosas o irreflexivas en esta serie, que es precisamente la de mayor dificultad y la última en ser administrada. Sin embargo, la correlación múltiple entre la serie quinta y las otras cuatro series fue de 0.638 en el grupo delincuente y de

0.628 en el grupo de soldados, es decir, las puntuaciones en la serie quinta son bastante predecibles a partir del comportamiento en las series anteriores, y esta predicción es aproximadamente la misma en ambos grupos. La asunción de las respuestas azaras y negligentes en el caso de los delincentes parece improbable.

La segunda parte de la tabla 4 está relacionada con la estimación de las medias y varianzas de las puntuaciones factoriales en ambos grupos. En primer lugar, cabría señalar que es posible comparar grupos en cuanto a su nivel factorial aún cuando el modelo de plena invarianza no sea aceptado. La única condición es que algunos parámetros no fijados sean invariantes en ambos grupos.

Atendiendo a las medias, que constituyen el principal interés de este estudio, la tabla 4 muestra una clara diferencia entre grupos que es estadísticamente significativa, como se puede ver al estudiar los intervalos confidenciales para el grupo no delincuente (la media del otro grupo es un parámetro fijado). Este resultado sugiere que la diferencia de medias intergrupo encontrada en las puntuaciones observadas tiene una correspondencia con la diferencia de medias equivalente en las puntuaciones factoriales.

Discusión

Es importante considerar que en cierto modo se puede hacer una interpretación y una generalización a partir de los resultados de este estudio. Estrictamente hablan-

do, estos resultados muestran que: a) hay diferencias en las puntuaciones observadas en el Raven entre un grupo delincuente y otro no delincuente, ambos con un nivel de educación similar, y b) un modelo de estructura de covarianza con parámetros parcialmente invariantes y una clara diferencia en la media de las puntuaciones factoriales estimadas entre los dos grupos de comparación se ajusta razonablemente bien a los datos observados.

Una interpretación más realista de los resultados hace referencia a la cuestión: cuando se compara el resultado de los delincentes y los no delincentes en el Raven, ¿Hasta qué punto las diferencias observadas reflejan diferencias reales en el constructo que se pretende medir?. En lo que concierne a los dos grupos aquí comparados, la evidencia sugiere que las diferencias observadas están principalmente determinadas por diferencias correspondientes a un sólo factor común que subyace en el test (supuestamente g).

Finalmente, aún cuando los resultados sugieren que el grupo delincuente tiene un nivel más bajo de g, la hipótesis de que el test no muestra funcionamiento diferencial en absoluto debe ser rechazada. Es evidente que se necesita más investigación para poder interpretar adecuadamente estos resultados. En primer lugar, sería necesario llevar a cabo estudios de validación cruzada a fin de investigar si los resultados aquí obtenidos se repiten en diferentes muestras. Si es así, se necesitaría un estudio para poder examinar con detalle los contenidos de las series que se encontraron que no eran inva-

Referencias

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Browne, M.W. y Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bo-

- llen & J.S. Long (Eds.) *Testing structural equation models*. Newbury Park: Sage.
- Burke, H.R. (1958). Raven's Progressive Matrices: A review and critical evaluation. *Journal of Genetic Psychology*, 93, 199-228.
- Camilli, G. (1994). The case against item bias detection techniques based on internal criteria: Do item bias procedures obscure test fairness issues?. En W.P. Holland y H. Wainer (Eds). *Differential item functioning*. Hillsdale, NJ: LEA
- Díaz, A. Beleña, A. y Báguena, M.J. (1994). The role of gender in juvenile delinquency: personality and intelligence. *Personality and Individual Differences*, 16, 309-314.
- Eysenck, H.J. y Gudjonsson (1989). *The courses and cures of criminality*. New York: Plenum Press.
- Gordon, R.A. (1986). Scientific justification and the race-IQ-delinquency model. In T.F. Hartnagel & R.A. Silverman (Eds.) *Critique and explanation: Essays in honor of Gwynne Nettler*. New Brunswick: Transaction Books.
- Hirschi, T. y Hindelang, M.J. (1977). Intelligence and delinquency: A revisionist review. *American Sociological Review*, 42, 571-587.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free press.
- Jensen, A.R. y Faulstich, M.E. (1988). Difference between prisoners and the general population in psychometric g. *Personality and Individual Differences*, 9, 928-928.
- Jöreskog, K.G y Sörbom, D. (1989). *LISREL 7 User's reference guide*. Mooresville: Scientific Software.
- Mellenbergh, G.J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13, 127-143.
- Murga, A. (1970). *P.M. 56 - Matrices progresivas*. Madrid: Mepsa
- Muthen, B. y Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Trasler, G. (1973). Criminal behaviour. In H.J. Eysenck, (Ed.) *Handbook of abnormal psychology*. London: Pitman Medical.
- West, D.J. (1969). *Present conduct and future delinquency*. London: Heinemann.
- West, D.J. y Farrington, D.P. (1973). *Who becomes delinquent?*. London: Heinemann.

Aceptado el 20 de abril de 1996