Revista Electrónica de Metodología Aplicada
2010, Vol. 15 nº 1, pp. 12-18

Detección del DIF en ítems politómicos mediante el uso de los métodos Mantel-Haenszel

Angel M. Fidalgo[1], Laura Quintanilla[2], Rubén Fernández[3],
Francisco Pons[4] y María Ester Aguerri[5]

[1]Universidad de Oviedo, [2]UNED, [3]Universidad de Almería,
[4]Universidad de Oslo, [5]Universidad de Buenos Aires

RESUMEN

Los métodos Mantel-Haenszel (MH) constituyen uno de los procedimientos de referencia usados para la detección del funcionamiento diferencial de los ítems (DIF), tanto en ítems dicotómicos como politómicos. Hasta la fecha los estadísticos empleados para evaluar el DIF han sido el estadístico ji-cuadrado MH, el test Mante-Haenszel generalizado y el test de Mantel, pese a que sólo permiten evaluar el DIF en dos grupos simultáneamente. En este artículo se analiza que estadísticos MH permiten analizar el DIF en ítems politómicos cuando deseamos analizar múltiples grupos.

Palabras clave: Estadísticos Mantel-Haenszel generalizados, Funcionamiento diferencial de los ítems, DIF, Ítems politómicos, Métodos Mantel-Haenszel.

ABSTRACT

Mantel-Haenszel (MH) methods constitute one of the most popular non-parametric differential item functioning (DIF) detection procedures, and have been applied for detecting DIF in both dichotomous and polytomous items. To date, the statistics used for assessing DIF has been the MH chi-square statistic, the generalized Mantel-Haenszel test and the Mantel test, although they limit the analysis to two groups. This article analyses which MH statistics permit DIF assessment in multiple groups and polytomous items.

Keywords: Differential item functioning, DIF, Generalized Mantel-Haenszel statistics, Mantel-Haenszel methods, Politomous items.

.

Este artículo ha sido posible gracias a la financiación proporcionada por el Ministerio Español de Ciencia y Educación [Proyectos de investigación SEJ2006-07491 y PCI2006-A7-0553].

Dirección para correspondencia:
Ángel M. Fidalgo.
Departamento de Psicología, Universidad de Oviedo.
Plaza de Feijoo s/n, 33003 Oviedo, España
e-mail: fidalgo@uniovi.es.

1.- Introducción

Como es sabido, el funcionamiento diferencial de los ítems constituye una amenaza sobre la validez de las puntuaciones obtenidas en tests y escalas psicológicas y educativas. Desde un punto de vista técnico se dice que un ítem funciona diferencialmente cuando personas que tienen el mismo nivel en la variable medida por el test, pero que pertenecen a diferentes subgrupos (por ejemplo, hombres/mujeres), tienen diferente probabilidad de contestarlo correctamente. En el ejemplo puesto, el ítem daría diferentes valores en función del género de la persona evaluada. Se comprenderá porque la investigación de técnicas estadísticas que permitan detectar el funcionamiento diferencial de los ítems (differential item functioning, DIF) ha sido un área de creciente interés psicométrico. Se han propuesto multitud de procedimientos para detectar el DIF, desde complejos estadísticos basados en la TRI hasta simples pruebas ji-cuadrado (para revisiones en español sobre esta temática véase: Fidalgo, 1996; Hidalgo y Gómez, 1999). De entre todos ellos, y desde el principio, han destacado los métodos Mantel-Haenszel (MH), hasta el punto de convertirse en los procedimientos de detección de referencia. Este artículo hace un revisión de los diferentes estadísticos MH que podemos emplear para evaluar el DIF en ítems politómicos, de sus equivalencias y ventajas.

2.- Estadísticos MH

Los métodos Mantel-Haenszel se han empleado tanto para la evaluación del DIF en ítems puntuados dicotómicamente como en ítems politómicos. En el caso de los ítems dicotómicos fueron Holland y Thayer (1988) quienes propusieron emplear el estadístico ji-cuadrado MH (χ^2_{MH}) desarrollado por Mantel and Haenszel (1959). También en el caso de los ítems politómicos se han propuesto varios estadísticos basados en los trabajos originales de Mantel y Haenszel: el test MH generalizado (Mantel & Haenszel, 1959; Zwick, Donoghue, & Grima, 1993) y el test de Mantel (Mantel, 1963; Zwick et al., 1993). El test MH generalizado considera las categorías de respuesta del ítem politómico como una variable nominal. De tal forma que la hipótesis alternativa (H_1) de DIF especifica que la distribución de las respuestas al ítem difiere entre los grupos. De otra parte, el test de Mantel considera la naturaleza ordenada de las categorías de respuesta del ítem politómico, especificando en este caso la H_1 de DIF que es la media de las puntuaciones correspondientes a las categorias de respuesta lo que difiere entre los grupos. La principal limitación de ambos estadísticos para evaluar el DIF es que sólo es posible comparar dos grupos simultáneamente. Por tanto, en caso de que se desee evaluar el DIF entre varios grupos sería necesario realizar tantos análisis por ítem como pares de comparaciones posibles haya; lo que necesariamente afectará bien a la tasa de error de tipo I (si no se controla aplicando, por ejemplo, la corrección de Bonferroni), bien a la potencia de prueba (si se controla la tasa de error de tipo I aplicando la corrección de Bonferroni); véase a este respecto Fidalgo y Scalon (2010). El lector puede encontrar una buena aplicación práctica tanto del test MH generalizado como del test de Mantel en Elosua y López-Jauregui (2007).

Afortunadamente, tal y como han planteado Fidalgo y Madeira (2008), dentro de los métodos MH existen mejores alternativas para evaluar el DIF en ítems politómicos que el test de Mantel y el test MH generalizado. Como se señala allí, en 1978 Landis, Heyman y Koch formularon un estadístico MH generalizado que subsumía tanto el test MH generalizado y el test de Mantel, como el estadístico χ^2_{MH}. A parte de la simplificación inherente a una única formulación, su principal ventaja para detectar el DIF es que permite, mediante un único test de significación, evaluar el DIF simultáneamente en varios grupos, y tanto en ítems dicotómicos como politómicos (Fidalgo y Scalon, 2010). A continuación se describe el estadístico MH generalizado y su correspondencia con los estadísticos habitualmente empleados para detectar el DIF.

3.- Estadístico MH generalizado

En 1978 Landis et al. propusieron un estadístico MH generalizado para el análisis de tablas de contingencia de dimensiones *Q*: *R* x *C*, siendo *Q* el número de estratos o tablas de contingencia, *R* el número de filas en cada tabla y *C* el número de columnas en cada tabla. En la Tabla 1 se muestra la estructura y notación correspondiente a dicha tabla general.

Niveles	Categorías de la variable de respuesta						
del factor	1	2	·	*j*	·	*C*	Total
1	n_{h11}	n_{h12}	·	n_{h1j}	·	n_{h1C}	$N_{h1\cdot}$
2	n_{h21}	n_{h22}	·	n_{h2j}	·	n_{h2C}	$N_{h2\cdot}$
⋮	⋮	⋮	⋮	⋮	.	⋮	⋮
i	n_{hi1}	n_{hi2}	·	n_{hij}	·	n_{hiC}	$N_{hi\cdot}$
⋮	⋮	⋮	⋮	⋮	.	⋮	⋮
R	n_{hR1}	n_{hR2}	·	n_{hRj}	·	n_{hRC}	$N_{hR\cdot}$
Total	$N_{h\cdot 1}$	$N_{h\cdot 2}$	.	$N_{h\cdot j}$	·	$N_{h\cdot C}$	$N_{h\cdot\cdot}$

Nota: En un análisis de DIF, los niveles del factor serían los diferentes grupos a comparar y la variable de respuesta, las categorías de respuesta del ítem. Cada tabla correspondería a un nivel de la covariable o variable de emparejamiento, que suele ser la puntuación total en el test.

Tabla 1. Estructura de la Tabla de contingencia en el *h*-ésimo estrato.

El estadístico MH generalizado que sirve para someter a contrastación la hipótesis nula de no-asociación entre el factor y la variable de respuesta, controlando el efecto de la covariable, es definido en término de matrices por Landis et alius (1978) por:

$$Q_{GMH} = \left\{\sum_{h=1}^{Q}(\mathbf{n}_h - \mathbf{m}_h)'\,\mathbf{A}_h'\right\}\left\{\sum_{h=1}^{Q}\mathbf{A}_h\mathbf{V}_h\mathbf{A}_h'\right\}^{-1}\left\{\sum_{h=1}^{Q}\mathbf{A}_h(\mathbf{n}_h - \mathbf{m}_h)\right\}. \qquad (1)$$

donde $\mathbf{n}_h$, $\mathbf{m}_h$, $\mathbf{V}_h$ and $\mathbf{A}_h$ son, respectivamente, el vector de frecuencias observadas, el vector de frecuencias esperadas, la matriz de covarianzas, y una matriz de funciones lineales definidas de acuerdo con la hipótesis alternativa (H_1) de interés. Dichos vectores y matrices se definen a partir de la Tabla 1 como sigue:

$$\mathbf{n}_h = (n_{h11}, n_{h21}, \ldots, n_{hRC})' \quad (CR \text{ x } 1),$$
$$\mathbf{m}_h = N_{h\cdot\cdot}(\mathbf{p}_{h\cdot *} \otimes \mathbf{p}_{h*\cdot}) \quad (CR \text{ x } 1),$$

$$\mathbf{V}_h = N_{h\cdot\cdot}^2 / (N_{h\cdot\cdot} - 1)\{(\mathbf{D}_{p_{h\cdot *}} - \mathbf{p}_{h\cdot *}\mathbf{p}_{h\cdot *}^{'}) \otimes (\mathbf{D}_{p_{h*\cdot}} - \mathbf{p}_{h*\cdot}\mathbf{p}_{h*\cdot}^{'})\} \quad (CR \text{ x } CR),$$

donde $\mathbf{p}_{h\cdot *}$ and $\mathbf{p}_{h*\cdot}$ son, respectivamente, vectores con las proporciones marginales de las columnas ($p_{h\cdot j} = N_{h\cdot j} / N_{h\cdot\cdot}$) y con las proporciones marginales de las filas ($p_{hi\cdot} = N_{hi\cdot} / N_{h\cdot\cdot}$), $\otimes$ denota el producto multiplicación de Kronecker, $\mathbf{D}_{p_{h\cdot *}}$ es una matriz diagonal con los elementos del vector $\mathbf{p}_{h\cdot *}$ en la diagonal principal, y $\mathbf{D}_{p_{h*\cdot}}$ es una matriz diagonal con los elementos del vector $\mathbf{p}_{h*\cdot}$ en la diagonal principal.

Como se ha señalado anteriormente, la expresión 1 dará lugar a diferentes estadísticos, en función de cómo se defina la matriz de funciones lineales $\mathbf{A}_h$ ($\mathbf{A}_h = \mathbf{C}_h \otimes \mathbf{R}_h$). Así obtendremos los siguientes estadísticos que pueden emplearse para la detección del DIF en ítems politómicos:

$Q_{GMH(1)}$ o estadístico MH generalizado nominal. Cuando tanto el factor (los grupos) como la variable de respuesta (las categorías del ítem) son variables nominales, la H_1 especifica que la distribución de la variable de respuesta difiere a lo largo de los niveles del factor. Aquí, $\boldsymbol{R}_h = [\boldsymbol{I}_{R\text{-}1}, \text{-}\boldsymbol{J}_{R\text{-}1}]$ y $\boldsymbol{C}_h = [\boldsymbol{I}_{C\text{-}1}, \text{-}\boldsymbol{J}_{C\text{-}1}]$, donde $\boldsymbol{I}_{R\text{-}1}$ es una matriz de identidad de dimensión (R-1 x R-1), y $\boldsymbol{J}_{R\text{-}1}$ es un vector de unos. Por tanto, la dimensión de $\boldsymbol{R}_h$ será (R-1 x R). De igual modo, $\boldsymbol{I}_{C\text{-}1}$ es una matriz de identidad de dimensión (C-1 x C-1), y $\boldsymbol{J}_{C\text{-}1}$ es un vector de unos. Bajo H_0, $Q_{GMH(1)}$ sigue aproximadamente una distribución ji-cuadrado con grados de libertad (gl) igual a gl = (R-1)(C-1).

$Q_{GMH(2)}$ o estadístico MH generalizado ordinal. Cuando la variable de respuesta (las categorías del ítem) es una variable medida como mínimo en una escala ordinal, la H_1 establece que la media de las respuestas difiere a lo largo de los niveles del factor. Aquí, $\boldsymbol{R}_h$ se define de igual manera que en el caso anterior y $\boldsymbol{C}_h = (c_{h1}, \ldots, c_{hC})$ es un vector de dimensión (1 x C), donde c_{hj} es una puntuación que refleja apropiadamente la naturaleza ordinal de la variable de respuesta en el estrato h-ésimo. En la literatura sobre DIF se suelen asignar enteros sucesivos a las diferentes categorías del ítem, aunque los valores de $\boldsymbol{C}_h$ admiten otras muchas posibilidades (Fidalgo y Madeira, 2008; Fidalgo et al. 2008, Julio). Bajo H_0, $Q_{GMH(2)}$ sigue aproximadamente una distribución ji-cuadrado con gl = (R-1).

El lector interesado puede encontrar información más detallada sobre estos estadísticos en Fidalgo (2005) y Fidalgo y Madeira (2008).

4. – Correspondencias entre los estadísticos MH

Los estadísticos presentados anteriormente subsumen tanto el estadístico χ^2_{MH}, como el resto de los estadísticos MH generalizados formulados por Mantel y Haenszel. Así podemos establecer las siguientes equivalencias entre los estadísticos habitualmente empleados en la literatura sobre DIF y los estadísticos $Q_{GMH(1)}$ y $Q_{GMH(2)}$.

1. *Test Mantel-Haenszel generalizado.* Cuando tenemos en el factor sólo 2 niveles, esto es, cuando en un análisis del DIF comparamos sólo dos grupos, $Q_{GMH(1)}$ es idéntico al test MH generalizado propuesto por Mantel y Haenszel (1959).

2. *Mantel test.* En el caso especial de 2 niveles en el factor (2 grupos), $Q_{GMH(2)}$ es idéntico al test de Mantel propuesto por Mantel (1963).

3. *Ji-cuadrado Mantel-Haenszel.* Cuando tenemos ítems puntuados dicotómicamente y dos grupos, $Q_{GMH(1)} = Q_{GMH(2)} = \chi^2_{MH}$. Hacemos notar que para que se cumpla esta correspondencia χ^2_{MH} debe calcularse sin la corrección de continuidad que habitualmente incluye.

5.- Uso de los estadísticos MH en la detección del DIF

La elección entre uno u otro de los estadísticos vistos en un análisis del DIF debe estar guiada por cuatro consideraciones: (a) el número de grupos sobre los que deseamos realizar el análisis del DIF, (b) el tipo de ítem, dicotómico o politómico, (c) si podemos ordenar o no las categorías de respuesta del ítem, y (d) el tipo de DIF que estemos interesados en detectar en función de la hipótesis alternativa que testa cada estadístico. Así por ejemplo, si quisiésemos analizar el DIF en más de 2 grupos, deberíamos emplear necesariamente el estadístico MH generalizado propuesto por Landis et al. (1978). En caso de que las categorías de respuesta del ítem no se puedan ordenar, es decir, estén en una escala de medida nominal, sólo podremos aplicar el estadístico $Q_{GMH(1)}$. Sin embargo, cuando tenemos ítems politómicos ordinales, podremos elegir entre aplicar el estadístico $Q_{GMH(1)}$ o $Q_{GMH(2)}$ en función de patrón de DIF que deseemos detectar. Como se describe en Fidalgo and Madeira (2008), $Q_{GMH(2)}$ incrementa la potencia de prueba respecto de $Q_{GMH(1)}$ para detectar su particular patrón de asociación: que la media de las respuestas difieren a través de los grupos comparados. Por eso, el estadístico $Q_{GMH(2)}$ es más efectivo para detectar lo que se denomina DIF constante, que es cuando el DIF tiene la misma dirección y, más o menos, la misma magnitud a lo largo de todas las categorías de respuesta del ítem. Por otro lado, $Q_{GMH(1)}$, al permitir detectar patrones de asociación más complejos tiene mucha más potencia que $Q_{GMH(2)}$ para detectar el DIF balanceado, que se produce cuando la magnitud del DIF está balanceada a lo largo de las categorías del ítem de forma que pueden cancelarse unas con otras dentro del mismo ítem.

Para concluir, en un intento de presentar de forma clara y resumida la información expuesta, la Figura 1 muestra en forma de un árbol de decisiones qué estadísticos MH debemos emplear en función de las características de los ítems y de la naturaleza y objetivos de nuestro estudio del DIF.

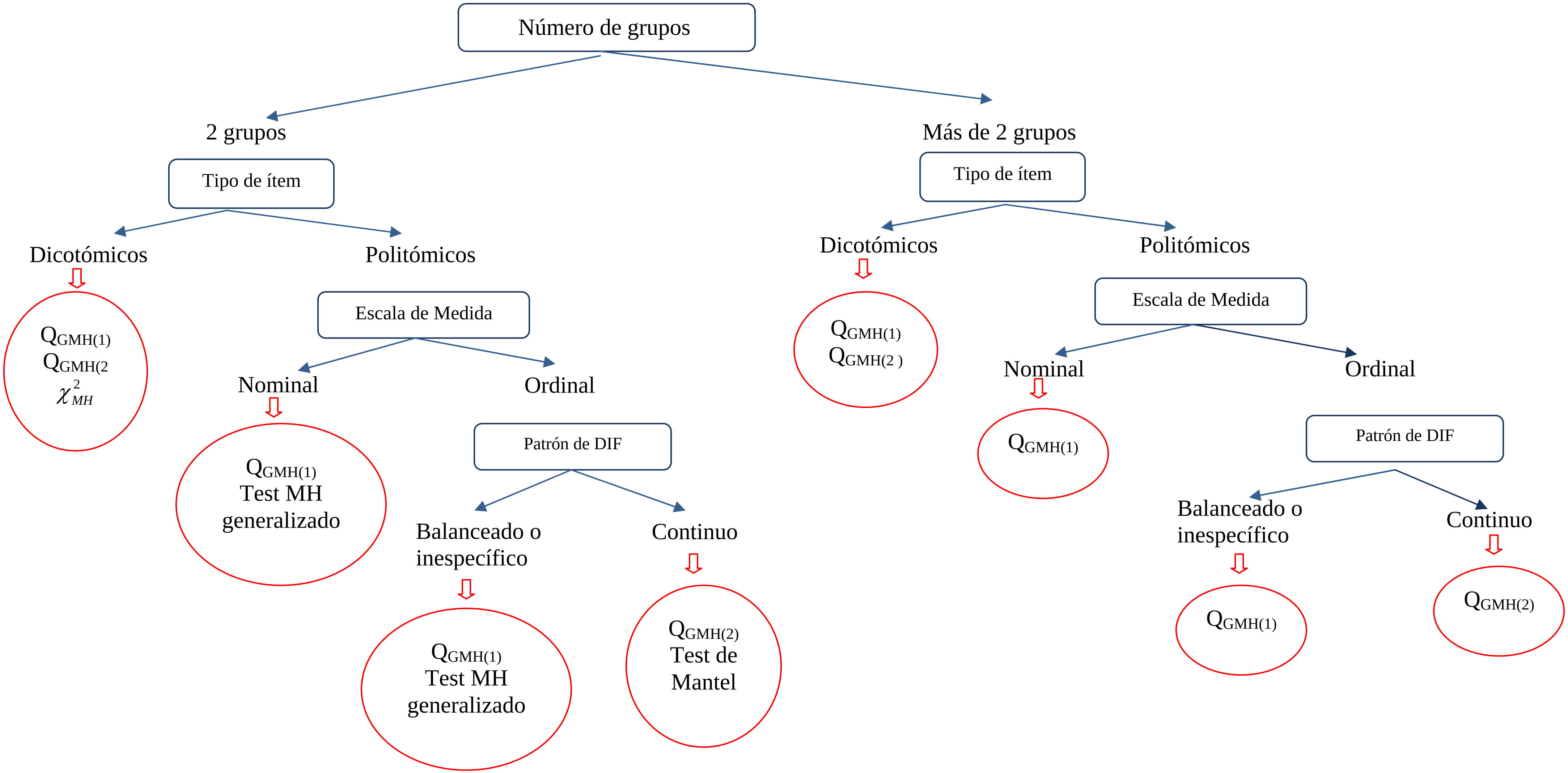

Figura 1. Esquema con los tipos de estadísticos MH que se pueden emplear para detectar el DIF en función de las características del ítem y de los objetivos del análisis del DIF. Cuando en los círculos aparece más de un estadístico el resultado de aplicar uno u otro es equivalente.

6.- Referencias

Elosua, P. y López-Jáuregui, A. (2007). Aplicación de cuatro procedimientos de detección del funcionamiento diferencial sobre ítems politómicos. *Psicothema*, *19*, 329-336.

Fidalgo, A.M. (1996). Funcionamiento diferencial de los ítems. En J. Muñiz (Ed.), *Psicometría* (págs. 371-455). Madrid : Universitas.

Fidalgo, A. M. (2005). Mantel-Haenszel Methods. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol.3, pp. 1120-1126). Chichester: John Wiley & Sons.

Fidalgo, A.M., y Madeira, J.M. (2008). Generalized Mantel-Haenszel methods for DIF detection. *Educational and Psychological Measurement, 68,* 940-958.

Fidalgo, A.M., y Scalon, J.D. (2010). Using Generalized Mantel-Haenszel Statistics to Assess DIF among Multiple Groups. *Journal of Psychoeducational Assessment, 28,* 60-69.

Fidalgo, A.M., Bartram, D., Quintanilla, L., Fernández, R., y Pons, F. (2008, July). Effect of the choice of scores assigned to the response variable on the Mantel test. In A. M. Fidalgo & P. Elosua (Chairs), *Mantel-Haenszel methods for DIF detection*. Symposium conducted at the III European Congress of Methodology, Oviedo, Spain.

Hidalgo, M.D., y Gómez, J. (1999). Técnicas de detección de funcionamiento diferencial en ítems politómicos. *Metodología de las Ciencias del Comportamiento, 1,* 39-60.

Holland, W. P., y Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. En H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: LEA.

Landis, J. R., Heyman, E. R., y Koch, G. G. (1978). Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistical Review, 46*, 237-254.

Mantel, N. (1963). Chi-square tests with one degree of freedom; extension of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58,* 690-700.

Mantel, N., y Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Zwick, R., Donoghue, J. R., y Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*, 233-251.