

Funcionamiento diferencial de los ítems en una situación de contacto de lenguas

Doris Ferreres Traver, Vicente González Romá y Juana Gómez Benito*
Universitat de València y * Universitat de Barcelona

Se aborda el problema del funcionamiento diferencial de los ítems (Differential Item Functioning, DIF) en una situación de contacto de lenguas, como la que se da en la Comunidad Valenciana. Actualmente, la medición y evaluación de las aptitudes de escolares valencianos se realiza con tests psicológicos en castellano y baremados con muestras castellanohablantes. El presente trabajo estudia si existe DIF en dos pruebas de aptitud elaboradas en castellano, una verbal y otra no verbal, y de uso habitual en el ámbito educativo valenciano. La muestra empleada está compuesta por 775 sujetos. Los resultados indican que la presencia de ítems con DIF es escasa (9 ítems, de los cuales 7 favorecen al grupo de referencia) al comparar el rendimiento de un grupo de escolares cuya lengua familiar y escolar es el castellano (grupo de referencia) con el rendimiento de: a) escolares bilingües cuya lengua familiar es el valenciano, y cuya lengua escolar es el castellano, y b) escolares bilingües cuya lengua familiar y escolar es el valenciano.

Differential item functioning and linguistic characteristics of examinees. The aim of this study is to ascertain whether differential item functioning (DIF) in a verbal ability test is related to examinees' linguistic characteristics. The sample was composed of 775 children between 10 and 14 who attended schools located in the Land of Valencia (Spain). This is a bilingual region with two official languages: Spanish and Valencian. In Valencian schools, it is quite usual to apply verbal tests elaborated in Spanish to children whose family and/or school language is Valencian. The results show that 9 items present DIF when the performance of a group of monolingual Spanish-speakers children was compared with the performance of: a) bilingual children whose family language is Valencian and school language is Spanish, and b) bilingual children whose family and school language is Valencian. In 7 out of the 9 DIF items, DIF was against bilingual subjects.

La utilización de los tests psicológicos para la realización de diagnósticos, asesoramiento y orientación vocacional en el ámbito educativo es una práctica bastante extendida. Sin embargo, una de sus mayores críticas es su limitada utilidad cuando los tests son aplicados a minorías lingüísticas y/o culturales. Esto ha estimulado en la comunidad científica una mayor preocupación por el estudio de las propiedades psicométricas de los tests aplicados a poblaciones con características diferentes a las de aquéllas para las que fueron creados. Esta circunstancia junto a la aparición del funcionamiento diferencial de los ítems (*Differential Item Functioning*, DIF en adelante) desató la conocida problemática de la medición psicológica en poblaciones minoritarias y, a consecuencia de ello, los tests psicológicos dejaron de ser vistos como una medida imparcial de las propiedades psicológicas (Fidalgo, 1996).

A partir de ese momento, el estudio del DIF recibe una amplia atención en la aplicación de tests a grupos minoritarios definidos por su etnia, raza y género, e igualmente en escenarios bilingües,

comúnmente conocidos como *situaciones de contacto de lenguas*. La enorme trascendencia de este fenómeno en esta situación se pone de manifiesto cuando se constata la frecuencia con que un instrumento de medida elaborado para usarse en una población monolingüe también se aplica a grupos bilingües, con una lengua materna minoritaria, y que a su vez conviven en una misma zona geográfica. Ejemplos claros de una situación de contacto de lenguas se encuentran en Canadá, Estados Unidos, Bélgica, Luxemburgo, Suiza y España.

Algunos estudios desarrollados en estos países se han centrado en averiguar si la aplicación de un test elaborado para una población monolingüe presenta problemas de DIF cuando es aplicado a una población bilingüe minoritaria. Drasgow y Hulin (1986) estudiaron si el JDI (Job Descriptive Index; Smith, Kendall y Hulin, 1969), un test muy utilizado para medir la satisfacción laboral, presentaba DIF cuando la versión original en inglés era aplicada a una muestra monolingüe (norteamericanos) y a una muestra bilingüe hispana (cubanos y puertorriqueños). Los análisis mostraron que el 4% de los ítems tenían problemas de funcionamiento diferencial (3 ítems de un total de 72). En otro estudio (Candell y Hulin, 1987), la misma versión inglesa del JDI fue administrada a una muestra monolingüe (canadienses anglófonos) y a otra bilingüe (canadienses francófonos). En este caso, la muestra bilingüe identificó el «francés» como su primera lengua, pero prefirió respon-

der la versión inglesa del citado test. Los resultados mostraron que ninguno de los 89 ítems tenían problemas de DIF, aunque dos de ellos estaban muy próximos al nivel de significación estadística considerado ($\alpha=0.001$). En resumen, estos estudios íntegramente desarrollados con adultos y en contextos laborales, mostraron una escasa presencia del DIF al comparar muestras monolingües y bilingües. No obstante, estos resultados plantean también una serie de interrogantes, tales como si se obtendrán los mismos resultados en ámbitos educativos, si existe alguna relación entre DIF y las competencias lingüísticas de monolingües y bilingües, o si la presencia del DIF está determinada por el tipo de prueba analizada. Estas cuestiones son abordadas a continuación.

El presente estudio

La consideración de la problemática del DIF en un país como España es especialmente importante. La existencia de varias Comunidades Autónomas con lenguas oficiales diferentes al castellano ha hecho que la situación de contacto de lenguas en nuestro país se encuentre bastante extendida. Un ejemplo es la Comunidad Valenciana (CV), donde el uso y aprendizaje de sus dos lenguas oficiales, el castellano y el valenciano, es un hecho frecuente y cada vez más normalizado. La presencia de ambas lenguas oficiales en la CV difiere notablemente según la región geográfica, pudiéndose distinguir una zona castellanohablante y otra valencianohablante. A nivel genérico, los programas educativos vigentes en la CV pueden clasificarse en dos categorías: el programa monolingüe, con una enseñanza íntegramente en castellano, y el programa bilingüe, que se divide en: 1. el modelo de mantenimiento, donde las habilidades lectoescritoras se aprenden en castellano, y ésta es la lengua escolar en toda la enseñanza obligatoria, a excepción de algunas áreas curriculares en 3º de Primaria y 1º de ESO, y 2. el modelo de enriquecimiento, donde las habilidades lectoescritoras se adquieren en valenciano, y el castellano se imparte como asignatura. Estas peculiaridades pueden ser relevantes para la adecuación de cualquier evaluación psicológica desarrollada en el ámbito educativo valenciano. Sin embargo, parecen ignorarse con frecuencia, y la evaluación de las aptitudes intelectuales de los escolares valencianos se realiza generalmente mediante la administración de tests psicológicos elaborados en castellano y baremados con muestras castellanohablantes, con independencia de la lengua familiar y/o escolar del alumno (Ferrerres, 1998; Ferreres, González-Romá y Gómez, 2000). La convivencia de estos programas de educación junto con la diversidad lingüística del territorio valenciano nos permite definir una serie de grupos relevantes para nuestro estudio: 1) alumnos cuya lengua familiar (LF) y lengua escolar (LE) es el castellano (grupo LFC+LEC); 2) alumnos cuya LF es el valenciano y LE es el castellano (grupo LFV+LEC); y 3) alumnos cuya LF y LE es el valenciano (grupo LFV+LEV).

La existencia de estos grupos y las prácticas de medición psicológica en el ámbito educativo valenciano nos llevó a plantear una serie de preguntas sobre la adecuación de la medición y la evaluación de las capacidades intelectuales en los escolares valencianos bilingües:

1) ¿Existirá DIF cuando la evaluación de las capacidades intelectuales de los escolares valencianos bilingües se realice mediante pruebas psicológicas elaboradas en castellano y baremadas con muestras monolingües castellanoparlantes?

2) En caso de existir DIF, ¿podría decirse que la aparición del mismo está relacionado con el tipo de prueba psicológica administrada (verbal vs. no verbal)?

En relación con la primera pregunta, hay que señalar que Cummins (1981, 1984), a través de su teoría de los «umbrales» y el principio de «interdependencia lingüística», ha destacado la importancia de una primera lengua bien establecida como paso previo para la adquisición de una segunda lengua a niveles aceptables. En el caso de los escolares bilingües, esto implica que la lengua familiar «minoritaria» (valenciano) debería estar bien consolidada para que los escolares pudieran lograr unos niveles de competencia aceptables en la lengua mayoritaria (castellano). En Educación Primaria es difícil que los escolares bilingües hayan alcanzado ese nivel de consolidación de la lengua familiar. Los niños del grupo LFV+LEC no utilizan el valenciano como lengua escolar, lo cual dificultará su aprendizaje y consolidación. Por otra parte, aunque los niños del grupo LFV+LEV sí usan el valenciano como lengua escolar, cabe pensar que en las etapas iniciales de su escolarización todavía no habrán alcanzado un nivel de competencia que garantice la consolidación de la lengua familiar. Es por ello que, de acuerdo con Cummins, cabe esperar que el nivel de competencia en el uso del castellano de los niños bilingües valencianos (grupos LFV+LEV y LFV+LEC) sea inferior que el nivel de competencia de los niños del grupo LFC+LEC. Así pues, este menor nivel de competencia en castellano de los niños bilingües será el responsable de que su rendimiento en una prueba de aptitud intelectual sea inferior al rendimiento de niños monolingües con igual nivel intelectual. Considerando esta argumentación, se formulan las siguientes hipótesis:

Hipótesis 1: Se detectarán ítems con DIF que perjudiquen el rendimiento en el test de los sujetos cuya lengua familiar es el valenciano y cuya lengua escolar es el castellano (grupo LFV+LEC), respecto a los sujetos cuya lengua familiar y escolar es el castellano (grupo LFC+LEC), cuando sus aptitudes se evalúan con pruebas verbales elaboradas en castellano.

Hipótesis 2: Se detectarán ítems con DIF que perjudiquen el rendimiento en el test de los sujetos cuya lengua familiar y escolar es el valenciano (grupo LFV+LEV), respecto a los sujetos cuya lengua familiar y escolar es el castellano (grupo LFC+LEC), cuando sus aptitudes se evalúan con pruebas verbales elaboradas en castellano.

Respecto a la segunda cuestión planteada, cabe señalar que si la variable responsable de la aparición del DIF en pruebas verbales es la competencia lingüística en castellano de los grupos comparados, entonces no deberían aparecer ítems con DIF cuando se analice la ejecución de los grupos en pruebas no verbales. La evidencia empírica ofrecida por Resing, Bleichrodt y Drenth (1986) y Van de Rijt (1990) parece apoyar esta idea. Observaron que las puntuaciones obtenidas en un test verbal de inteligencia por niños inmigrantes, eran en torno a 14 puntos inferiores a las puntuaciones obtenidas por niños holandeses. Los investigadores concluyeron que las características lingüísticas de los grupos minoritarios interactuaban con el contenido verbal de las pruebas administradas, dificultando el rendimiento en el test de los grupos minoritarios. Considerando estos resultados, otros investigadores holandeses elaboraron un test de inteligencia no verbal para minorías étnicas («Learning Potential Tests for Ethnic Minorities», Hamers,

Hessels y Van Luit, 1991), que fue sometido a un análisis de DIF con el estadístico Mantel-Haenszel. Los resultados revelaron una escasa presencia de DIF, ya que sólo 5 de los 95 ítems mostraron DIF. Además, el funcionamiento diferencial detectado favoreció tanto al grupo de referencia (nativos holandeses) como al grupo focal (emigrantes turcos y marroquíes) (Hamers, Hessels y Penning, 1996). Considerando estos argumentos y evidencias empíricas, se formula la siguiente hipótesis:

Hipótesis 3: Se espera que no aparezcan ítems con DIF que perjudiquen el rendimiento en el test de los sujetos cuya lengua familiar es el valenciano (grupos LFV+LEV y LFV+LEC), respecto a los sujetos cuya lengua familiar y escolar es el castellano (grupo LFC+LEC), cuando sus aptitudes se evalúan con pruebas no verbales.

En suma, el objetivo del presente estudio es determinar si existe DIF en dos tests de inteligencia (uno verbal y otro no verbal), elaborados y baremados en castellano, y si el DIF se debe a las diferentes características lingüísticas de los grupos considerados.

Método

Muestra

La muestra se compone de 2128 alumnos escolarizados en 4º, 5º y 6º de Primaria, y 1º y 2º de Secundaria. En Educación Primaria, se obtuvieron datos de sujetos que cursaban los tres programas educativos mencionados. En Educación Secundaria, se recogieron datos únicamente de sujetos en los modelos monolingüe y de mantenimiento. Esta circunstancia fue la causa principal por la que el estudio se centró únicamente en Educación Primaria. El tamaño muestral en Educación Primaria fue de 775 alumnos. De todo el conjunto, 333 alumnos cursaban el modelo monolingüe en castellano (LEC), 193 alumnos el modelo de mantenimiento (LEC), y 249 el modelo de enriquecimiento (LEV).

Dichos grupos fueron comparados en las variables sexo y edad. En relación con la variable sexo, los resultados obtenidos para la comparación entre los grupos LFC+LEC y LFV+LEC revelaron que no existían diferencias significativas entre los grupos considerados ($z = 2.07$, $p > 0.05$). El mismo resultado se obtuvo para la comparación entre los grupos LFC+LEC y LFV+LEV ($z = 0.23$, $p > 0.05$). En cuanto a la variable edad, se observaron diferencias estadísticamente significativas entre los tres grupos (ANOVA: $F = 7.687$, $p < 0.001$). Las comparaciones post-hoc revelaron diferencias significativas sólo entre los grupos LFC+LEC y LFV+LEV, aunque la diferencia tipificada entre ambas medias no resultó elevada ($d \approx 0.16$) (grupo LFC+LEC: media = 11.06; grupo LFV+LEV: media = 10.79).

Instrumento de medida

El estudio se centra en dos pruebas psicológicas: la escala de Habilidad Mental Verbal (HMV) y la de Habilidad Mental No-Verbal (HMNV) de la «Batería de Aptitudes Diferenciales y Generales» (BADYG, Yuste, 1988, 1995). La escala HMV está compuesta por 40 ítems verbales ordenados por dificultad, y cuya variada tipología es la que sigue: a) constancia de una característica; b) secuencias lógicas con números; c) ordenar palabras sueltas formando una frase correcta; d) hallar el género o característica cla-

sificatoria de una serie de palabras; e) problemas numéricos de comprensión lógico-numérica, y f) problemas de resolución espacio-temporal. La escala HMNV consta de 40 ítems gráficos, ordenados según su dificultad, y clasificados en: a) ítems estáticos, y b) dinámicos. Dicha batería, elaborada en castellano, es una de las pruebas psicológicas más utilizadas en los centros escolares de la CV, es por ello que fue seleccionada para este estudio. De la escala HMV fueron seleccionados aquellos ítems que poseían un alto contenido verbal, debido a la importancia de la naturaleza verbal de los ítems en la formulación de nuestras hipótesis. Se eliminaron aquellos ítems cuyo contenido no estaba directamente relacionado con la aptitud verbal (las tipologías b, e y f), quedando reducida la prueba estudiada a 23 ítems. En la escala HMNV fueron seleccionados de forma aleatoria otros 23 ítems, con el fin de que ambas pruebas tuvieran la misma longitud.

Variables

Lengua familiar (LF). Dicha variable fue medida mediante una encuesta socio-lingüística. La operacionalización de esta variable se realizó como sigue: 1. lengua familiar castellana (LFC): sujetos cuya lengua familiar era siempre el castellano, y 2. lengua familiar valenciana (LFV): sujetos cuya lengua familiar era siempre el valenciano o al menos existía una predominancia clara del valenciano sobre el castellano. Al respecto, véase Ferreres (1998).

Lengua escolar (LE). Esta variable fue operacionalizada según el modelo educativo cursado. Las peculiaridades de cada modelo educativo se indican en la parte introductoria del artículo.

Tipo de escala. Se aplicó la escala HMV, con ítems de contenido verbal; y la escala HMNV, con ítems gráficos.

Análisis

Primero, se realizó un análisis de las propiedades psicométricas de las escalas HMV y HMNV. A continuación, fueron sometidas a un análisis de componentes principales con el fin de evaluar la unidimensionalidad de las mismas. En lo relativo al análisis de DIF, se tomó como grupo de referencia al grupo LFC+LEC, y como grupos focales a los grupos LFV+LEC y LFV+LEV. Atendiendo al tamaño muestral de los grupos comparados, se optó por emplear el estadístico Mantel-Haenszel (MH; Holland y Thayer, 1986) para datos dicotómicos. Entre sus ventajas, destacan su amplio uso, su simplicidad en cálculo e interpretación, y sus buenos resultados con tamaños muestrales pequeños (200 individuos por grupo). La aplicación de esta técnica se realizó mediante el programa MHDIF elaborado por Fidalgo (1994). Dicho programa ofrece los estadísticos MH para el cálculo del DIF uniforme, e implementa la modificación propuesta por Mazor, Clauser y Hambleton (1994) para la detección del DIF no uniforme. Se ha demostrado que esta estrategia en la aplicación del procedimiento MH produce un incremento de la potencia de prueba y una reducción de la tasa de error de Tipo I (Fidalgo, Mellenbergh y Muñoz, 1998, 2000). Como estadísticos para cuantificar la magnitud del DIF ofrece el cociente de razones común MH ($\hat{\delta}_{MH}$) y el estadístico MH delta (MH-D), también ofrece un test de significación estadística (χ^2_{MH}). Además, permite detectar el DIF mediante una depuración bietápica del criterio. En este estudio, se concluyó que los ítems analizados presentaban DIF cuando el estadístico ji-cuadrado MH resultó significativo, a un nivel de significación de $0.05/n^\circ$ de ítems de la escala ($p = 0.05/23 = 0.0021$). Este criterio res-

ponde a un compromiso entre el nivel de significación adoptado por comparación y el elevado número de comparaciones realizadas, a fin de prevenir el error de Tipo I. También, se consideró el estadístico delta (MH-D), un índice de la magnitud del DIF empleado por el Educational Testing Service (ETS). Los ítems con una magnitud igual o mayor que 1.51 son los que presentan problemas de funcionamiento diferencial.

Resultados

Análisis descriptivos

Respecto a la escala HMV, los dos grupos focales (LFV+LEC, media= 11.78; LFV+LEV, media= 11.77) ofrecieron un peor rendimiento que su grupo de referencia (LFC+LEC, media= 12.09), aunque las diferencias observadas no fueron estadísticamente significativas (ANOVA: $F= 0.710, p= 0.492$). Por el contrario, en la escala HMNV apareciendo diferencias estadísticamente significativas entre los tres grupos (ANOVA: $F= 5.888, p= 0.003$) (LFC+LEC, media= 12.63; LFV+LEC, media= 13.67; LFV+LEV, media= 13.8). Los análisis *a posteriori* realizados mediante la prueba de Bonferroni revelaron que las diferencias significativas

aparecen entre los grupos LFC+LEC y LFV+LEC (diferencia de medias= -1.04, $p= 0.031$) y los grupos LFC+LEC y LFV+LEV (diferencia de medias= -1.16, $p= 0.006$). A su vez, se procedió a estimar el tamaño del efecto mediante el estadístico *d* (Cohen, 1969). Los valores oscilaron entre 0.08 y 0.24 que, siguiendo los criterios de Cohen, indican un tamaño del efecto pequeño ($d \approx 0.20$). Respecto a la fiabilidad, ambas escalas ofrecieron una consistencia interna aceptable. Los coeficientes KR-20 oscilaron entre 0.70 y 0.74 en la escala HMV, y entre 0.78 y 0.80 en la escala HMNV. Cabe resaltar que el ítem 11 de la escala HMV fue eliminado por su falta de variabilidad en el grupo LFV+LEV, quedando la escala reducida a 22 ítems.

Unidimensionalidad

La unidimensionalidad del constructo fue evaluada mediante un análisis de componentes principales de la matriz de correlaciones tetracóricas, y la comparación de los valores propios de los componentes (Resie, Smith y Furr, 2001). Los resultados obtenidos en cada grupo y prueba mostraron un salto considerable en los valores propios (VP) situado, para todos los casos, entre el primer y segundo componente (C1 y C2) (Prueba HMV: muestra LFC+LEC:

Tabla 1
Índices de funcionamiento diferencial de los ítems para la escala HMV en la comparación LFC+LEC vs. LFV+LEC

Ítems	DIF uniforme											
	01	02	03	04	05	06	07	08	09	10	11	
MH α	0.88	1.05	1.35	1.48	5.43	0.77	0.75	0.55	1.07	0.96	1.53	
MH D	0.31	-0.11	-0.71	-0.92	-3.98 [#]	0.61	0.68	1.40	-0.16	0.09	-0.99	
MH χ^2	0.26	0.00	0.69	2.29	5.54	0.28	1.65	0.11	0.02	0.02	4.02	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH α	1.45	0.69	0.99	0.62	1.38	0.63	0.84	1.20	0.98	1.10	0.84	1.06
MH D	-0.88	0.86	0.03	1.11	-0.76	1.09	0.40	-0.44	0.04	-0.23	0.40	-0.15
MH χ^2	2.39	2.41	0.00	4.64	2.17	3.97	0.46	0.42	0.00	0.03	0.12	0.01
Ítems	01	02	03	04	DIF no uniforme		07	08	09	10	11	
MH1 α	0.75	1.42	1.59	1.19	5.43	0.88	0.68	0.00	1.01	-	1.61	
MH2 α	1.14	0.21	0.74	2.27	-	0.33	0.88	0.81	1.23	0.64	1.45	
MH1D	0.67	-0.82	-1.09	-0.40	-3.98 [#]	0.29	0.90	-	-0.02	-	-1.12	
MH2D	-0.31	3.63 [#]	0.71	-1.92 [#]	-	2.58 [#]	0.29	0.50	-0.49	1.06	-0.87	
MH1 χ^2	0.87	0.77	1.39	0.18	5.54	0.01	1.81	0.07	0.02	1.47	2.31	
MH2 χ^2	0.04	1.89	0.01	3.25	-	0.36	0.03	0.05	0.05	0.47	1.38	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH1 α	1.25	0.69	0.81	0.60	1.35	0.53	1.44	1.30	0.99	2.00	1.03	0.67
MH2 α	2.14	0.70	1.33	0.66	1.43	0.79	0.58	1.15	0.98	0.99	0.80	1.28
MH1D	-0.52	0.88	0.49	1.19	-0.71	1.48	-0.86	-0.61	0.02	-1.62 [#]	-0.07	0.94
MH2D	-1.79 [#]	0.84	-0.68	0.99	-0.84	0.55	1.27	-0.32	0.04	0.03	0.51	-0.58
MH1 χ^2	0.47	1.03	0.43	3.12	0.96	4.23	0.77	0.24	0.05	0.28	0.11	0.36
MH2 χ^2	2.56	1.05	0.52	1.18	0.94	0.25	3.05	0.08	0.01	0.02	0.17	0.43

El símbolo «-» se debe a que no es posible calcular el estadístico requerido debido al número insuficiente de casos en algunos niveles de puntuaciones para calcular el estadístico MH; HMV: Habilidad Mental Verbal; LF: Lengua familiar; LE: Lengua de escolarización; C: Castellano; V: Valenciano; MH α : cociente de razones común; MH-D: estadístico delta; MH χ^2 : test de significación estadística; MH1 α : cociente de razones común en el grupo inferior a la media; MH2 α : cociente de razones común en el grupo superior a la media; MH1D: estadístico delta en el grupo inferior a la media; MH2D: estadístico delta en el grupo superior a la media; MH1 χ^2 : test de significación estadística en el grupo inferior a la media; MH2 χ^2 : test de significación estadística en el grupo superior a la media; * $p \leq 0.005$; [#] $MHD \geq 1.51$.

C1= 3.31 y C2= 1.73; muestra LFV+LEC: C1= 3.86 y C2= 1.96; muestra LFV+LEV: C1= 3.72 y C2= 1.84. Prueba HMNV: muestra LFC+LEC: C1= 4.48 y C2= 1.63; muestra LFV+LEC: C1= 4.83 y C2= 1.54; muestra LFV+LEV: C1= 4.92 y C2= 1.51). Estos resultados revelan la existencia de un primer componente dominante con un VP que, como mínimo, es 1.9 veces el valor del VP del segundo componente. Valores similares han sido utilizados por otros investigadores para justificar una unidimensionalidad suficiente en las pruebas analizadas (e. g., Reise et al., 2001).

Detección del DIF

En las tablas 1, 2, 3 y 4 se ofrece una cuantificación del DIF uniforme presente en los ítems mediante el cociente de razón común (MH α), además de su transformación a la escala de dificultad empleada por el ETS (delta MH, MHD), y el test de significación estadística (MH χ^2), así como del DIF no uniforme (MH1 α : el cociente de razón común para el grupo inferior a la media muestral; MH2 α : para el grupo superior a la media muestral; MH1D y MH2D: el estadístico delta para el grupo inferior y superior a la media muestral; MH1 χ^2 y MH2 χ^2 : test de significación estadística para el grupo inferior y superior, respectivamente).

En relación con la hipótesis 1, cuando se comparan los grupos LFC+LEC vs. LFV+LEC (ver tabla 1) los resultados referidos al test de significación estadística indican que no existe ningún ítem de la escala HMV que presente DIF. En cambio, los índices de magnitud del DIF, en concreto el MH-D ofrece una magnitud igual o mayor que 1.51 en 6 de los ítems analizados. Se trata del ítem 5 con DIF uniforme y los ítems 2, 4, 6, 12 y 21 con DIF no uniforme. En cuatro de ellos (ítems 4, 5, 12 y 21) el DIF aparece, tal como se esperaba, en contra del grupo focal (grupo LFV+LEC). Estos resultados indican una vez más la conveniencia de utilizar medidas del tamaño del efecto en combinación con los tests de significación estadística (Fidalgo y Ferreres, en prensa). Así pues, en términos generales, los resultados obtenidos muestran el sentido esperado según la primera hipótesis formulada.

Respecto a la hipótesis 2, los resultados del test de significación estadística correspondiente a la comparación LFC+LEC vs. LFV+LEV (ver tabla 3) únicamente revela un ítem con funcionamiento diferencial uniforme (el ítem 14) y, como era esperable, en contra del grupo focal (grupo LFV+LEV). Por su parte, el delta MH ofrece dos ítems (ítems 10 y 14) con problemas de DIF no uniforme, y en contra del grupo focal. Estos resultados son congruentes con lo esperado según la hipótesis 2. Por último, tal co-

Tabla 2
Índices de funcionamiento diferencial de los ítems para la escala HMNV en la comparación LFC+LEC vs. LFV+LEC

Ítems	DIF uniforme											
	01	02	03	04	05	06	07	08	09	10	11	
MH α	0.88	0.75	1.23	0.80	0.68	0.83	1.04	1.02	0.97	0.83	1.13	
MH D	0.29	0.68	-0.48	0.53	0.91	0.44	-0.10	-0.05	0.07	0.45	-0.29	
MH χ^2	0.16	1.04	0.71	1.02	3.10	0.53	0.02	0.00	0.00	0.57	0.08	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH α	0.89	1.76	1.29	1.35	0.96	0.91	1.51	1.15	1.07	1.18	0.82	0.78
MH D	0.26	-1.33	-0.61	-0.71	0.09	0.23	-0.97	-0.32	-0.16	-0.40	0.45	0.59
MH χ^2	0.10	5.72	1.17	1.47	0.00	0.10	2.37	0.19	0.03	0.21	0.57	0.92
Ítems	01	02	03	04	DIF no uniforme		07	08	09	10	11	
MH1 α	0.61	0.85	1.18	0.71	0.64	0.83	0.95	1.22	0.81	0.96	1.36	
MH2 α	1.76	0.55	1.31	0.90	0.73	0.83	1.15	0.57	1.29	0.68	0.77	
MH1D	1.17	0.38	-0.38	0.79	1.05	0.44	0.13	-0.46	0.50	0.11	-0.71	
MH2D	-1.33	1.42	-0.63	0.25	0.75	0.43	-0.34	1.31	-0.59	0.92	0.62	
MH1 χ^2	2.25	0.16	0.18	1.11	2.07	0.27	0.00	0.20	0.33	0.00	0.50	
MH2 χ^2	1.46	1.03	0.41	0.05	0.76	0.11	0.14	0.37	0.26	0.92	0.06	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH1 α	1.08	1.91	1.15	1.76	0.88	1.03	1.53	1.44	0.76	0.82	1.09	0.71
MH2 α	0.62	1.58	1.45	0.97	1.08	0.79	1.50	0.69	1.23	1.53	0.68	0.83
MH1D	-0.19	-1.53 [#]	-0.34	-1.33	0.31	-0.07	-1.00	-0.86	0.66	0.48	-0.21	0.81
MH2D	1.11	-1.08	-0.87	0.07	-0.18	0.54	-0.95	0.89	-0.49	-0.99	0.90	0.44
MH1 χ^2	0.01	3.97	0.09	2.91	0.10	0.00	0.74	1.16	0.18	0.05	0.01	0.59
MH2 χ^2	0.76	1.38	1.14	0.01	0.01	0.29	1.27	0.38	0.38	0.90	1.44	0.21

El símbolo «-» se debe a que no es posible calcular el estadístico requerido debido al número insuficiente de casos en algunos niveles de puntuaciones para calcular el estadístico MH; HMNV: Habilidad Mental no Verbal; LF: Lengua familiar; LE: Lengua de escolarización; C: Castellano; V: Valenciano; MH α : cociente de razones común; MH-D: estadístico delta; MH χ^2 : test de significación estadística; MH1 α : cociente de razones común en el grupo inferior a la media; MH2 α : cociente de razones común en el grupo superior a la media; MH1D: estadístico delta en el grupo inferior a la media; MH2D: estadístico delta en el grupo superior a la media; MH1 χ^2 : test de significación estadística en el grupo inferior a la media; MH2 χ^2 : test de significación estadística en el grupo superior a la media; * $p \leq 0.005$; [#] $MHD \geq 1.51$.

mo se esperaba por la hipótesis 3 (ver tabla 2 y 4), ningún ítem de la escala HMNV en las dos comparaciones realizadas (LFC+LEC vs. LFV+LEC y LFC+LEC vs. LFV+LEV) mostró funcionamiento diferencial, a excepción del ítem 13 en la comparación LFC+LEC vs. LFV+LEC, con una magnitud de DIF ($MHD = -1.53$) muy próxima al criterio adoptado.

Discusión

El objetivo de este estudio era determinar si existe DIF en los ítems de dos tests de inteligencia (uno verbal y otro no verbal) elaborados en castellano y aplicados a grupos con características lingüísticas diferentes de las del grupo para el que fueron desarrollados. De acuerdo con las dos primeras hipótesis, esperábamos que aparecerían ítems con DIF en la prueba verbal que perjudicaran el rendimiento de los dos grupos focales (sujetos cuya lengua familiar (LF) es el valenciano y cuya lengua de escolarización (LE) es el castellano, y sujetos cuya LF y LE es el valenciano) cuando son comparados con el grupo de referencia (sujetos cuya LF y LE es el castellano). Los resultados obtenidos indican la existencia de ítems con DIF generalmente en el sentido avanzado por las hipótesis formuladas.

En la comparación LFC+LEC vs. LFV+LEC se detectaron 6 ítems con DIF, de los cuales 4 perjudicaban al grupo focal, mientras que en la comparación LFC+LEC vs. LFV+LEV se identificaron 2 ítems con funcionamiento diferencial, ambos en contra del grupo focal. Estos resultados son congruentes con la teoría de los «umbrales» y el principio de «interdependencia lingüística» de Cummins (1981, 1984), según los cuales es muy importante que la primera lengua o lengua materna esté bien establecida como paso previo para la adquisición de una segunda lengua a niveles de competencia aceptables. Es poco probable que los escolares bilingües valencianos que todavía se encuentran en las etapas iniciales de su escolarización, hayan adquirido unos niveles de competencia elevados en su lengua materna. De manera que, de acuerdo con Cummins, cabe esperar que sus niveles de competencia en la lengua castellana no sean todavía aceptables, e inferiores a los niveles del grupo de referencia (LFC+LEC). Esta competencia diferente en la lengua castellana sería la responsable de la aparición de ítems con DIF que perjudican el rendimiento de los grupos bilingües.

La teoría de Cummins también permite explicar por qué en la comparación LFC+LEC vs. LFV+LEC aparecen más ítems con DIF que perjudican al grupo focal que en la comparación LFC+LEC vs. LFV+LEV. Cabe esperar que el grupo LFV+LEV

Tabla 3
Índices de funcionamiento diferencial de los ítems para la escala HMV en la comparación LFC+LEC vs. LFV+LEV

Ítems	DIF uniforme											
	01	02	03	04	05	06	07	08	09	10	11	
MH α	1.02	1.34	1.14	1.04	1.53	1.17	0.88	-	1.37	0.71	1.12	
MH D	-0.04	-0.69	-0.30	-0.09	-0.99	-0.38	0.31	-	-0.74	0.79	-0.27	
MH χ^2	0.00	0.93	0.09	0.00	0.11	0.15	0.37	1.09	1.52	0.58	0.26	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH α	1.00	1.43	1.76	0.77	1.32	0.66	0.85	0.85	0.77	1.13	0.83	0.91
MH D	0.00	-0.84	-1.32	0.60	-0.64	0.97	0.39	0.39	0.63	-0.30	0.45	0.21
MH χ^2	0.01	2.95	8.47*	1.56	1.70	3.51	0.52	0.40	0.92	0.10	0.22	0.08
Ítems	01	02	03	04	DIF no uniforme		07	08	09	10	11	
MH1 α	0.95	1.32	0.93	0.85	0.60	0.92	0.81	-	1.10	1.02	1.20	
MH2 α	1.09	1.24	1.83	1.47	-	1.74	1.02	-	1.74	0.45	1.08	
MH1D	0.13	-0.65	0.17	0.37	1.19	0.21	0.48	-	-0.22	-0.04	-0.42	
MH2D	-0.19	-0.50	-1.42	-0.91	-	-1.31	-0.04	-	-1.31	-1.87#	-0.17	
MH1 χ^2	0.01	0.52	0.00	0.16	0.08	0.00	0.54	0.26	0.05	0.06	0.31	
MH2 χ^2	0.01	0.03	0.84	0.50	0.87	0.49	0.01	0.02	0.72	2.15	0.02	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH1 α	1.14	1.22	1.94	0.82	1.14	0.55	1.00	0.93	1.21	0.55	1.22	0.49
MH2 α	0.73	1.55	1.54	0.61	1.44	0.78	0.65	0.69	0.63	1.20	0.74	1.19
MH1D	-0.30	-0.47	-1.56#	0.48	-0.31	1.39	-0.00	0.16	-0.45	1.42	-0.46	1.48
MH2D	0.74	-1.03	-1.02	1.16	-0.86	0.57	1.02	0.88	1.08	-0.42	0.71	-0.40
MH1 χ^2	0.13	0.31	6.31	0.47	0.16	4.18	0.02	0.00	0.03	0.30	0.01	1.45
MH2 χ^2	0.25	2.19	1.92	2.33	1.03	0.38	2.25	1.23	2.20	0.18	0.49	0.24

El símbolo «-» se debe a que no es posible calcular el estadístico requerido debido al número insuficiente de casos en algunos niveles de puntuaciones para calcular el estadístico MH; HMV: Habilidad Mental Verbal; LF: Lengua familiar; LE: Lengua de escolarización; C: Castellano; V: Valenciano; MH α : cociente de razones común; MH-D: estadístico delta; MH χ^2 : test de significación estadística; MH1 α : cociente de razones común en el grupo inferior a la media; MH2 α : cociente de razones común en el grupo superior a la media; MH1D: estadístico delta en el grupo inferior a la media; MH2D: estadístico delta en el grupo superior a la media; MH1 χ^2 : test de significación estadística en el grupo inferior a la media; MH2 χ^2 : test de significación estadística en el grupo superior a la media; * $p \leq 0.005$; # $MHD \geq 1.5$.

tenga más consolidada la lengua materna que el grupo LFV+LEC, ya que en el primer grupo la lengua materna es la lengua utilizada en la escuela de manera generalizada. Esta mejor consolidación del valenciano en los sujetos del grupo LFV+LEV, les permitiría alcanzar unos mejores niveles de competencia en la segunda lengua en comparación con el grupo LFV+LEC.

Respecto a la hipótesis 3, se esperaba que no aparecieran ítems con DIF en la prueba no verbal que perjudicaran el rendimiento en el test de los sujetos bilingües. Sólo se detectó un ítem con DIF en la comparación LFC+LEC vs. LFV+LEC, aunque con un valor del estadístico MHD (-1.53) muy cercano al valor crítico elegido (1.50). Por ello puede afirmarse que la magnitud del DIF detectado no fue elevada. En la comparación LFC+LEC vs. LFV+LEV, tal como se esperaba no se detectó ningún ítem con DIF. En general, estos resultados ofrecen apoyo a la tercera hipótesis formulada.

Este estudio tiene implicaciones prácticas importantes. Los resultados obtenidos indican que los ítems con un claro contenido verbal de la escala HMV de la batería BADYG presentan DIF a través de los grupos lingüísticos analizados, que, generalmente, perjudican el rendimiento de los sujetos bilingües que tienen el valenciano como lengua materna. De nuevo estos resultados nos muestran que existen razones suficientes para pensar que la medi-

ción de un constructo psicológico en una población con características lingüísticas y culturales específicas, mediante la aplicación de instrumentos de medida elaborados para una población diferente, frecuentemente resultará en una medición inadecuada, donde los sujetos del grupo minoritario serán infravalorados en la mayor parte de las ocasiones (Hambleton, 1994; van de Vijver y Poortinga, 1997). Así pues, los estudios sobre DIF son necesarios en todo proceso de medición psicológica en el que estén implicados sujetos pertenecientes a poblaciones con características lingüísticas y/o culturales diferentes a las de la población a la cual va destinado el test (Elosúa, López y Egaña, 2000). La contribución de estos estudios garantizará que los ítems originales de la prueba funcionen de forma equivalente en las dos poblaciones objeto de análisis y que, por tanto, las comparaciones entre ambas poblaciones tengan sentido. Cabe esperar que estos estudios sean aún más necesarios en un futuro, ya que nuestro país está recibiendo un continuo flujo de personas procedentes de otros países, algunos con lenguas y culturas muy diferentes a la nuestra.

Por otra parte, los resultados obtenidos indican que la prueba no verbal analizada, compuesta por ítems gráficos, apenas presenta problemas de DIF, por lo que los ítems analizados pueden ser utilizados para evaluar la inteligencia de los escolares bilingües de

Tabla 4
Índices de funcionamiento diferencial de los ítems para la escala HMNV en la comparación LFC+LEC vs. LFV+LEV

Ítems	DIF uniforme											
	01	02	03	04	05	06	07	08	09	10	11	
MH α	1.09	0.76	1.46	1.17	0.83	0.81	1.41	1.03	1.01	0.78	1.03	
MH D	-0.20	0.66	-0.88	-0.36	0.45	0.50	-0.80	-0.06	-0.03	0.58	-0.07	
MH χ^2	0.07	1.11	3.16	0.56	0.71	0.94	2.93	0.00	0.00	1.01	0.00	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH α	1.24	0.66	0.97	0.97	1.47	0.97	0.93	0.91	1.10	0.99	0.94	0.75
MH D	-0.50	0.97	0.07	0.07	-0.91	0.08	0.17	0.22	-0.22	0.02	0.15	0.67
MH χ^2	0.53	2.83	0.00	0.00	3.30	0.01	0.04	0.14	0.11	0.01	0.01	1.39
Ítems	DIF no uniforme											
Ítems	01	02	03	04	05	06	07	08	09	10	11	
MH1 α	1.21	0.78	1.75	1.07	0.83	0.69	1.40	1.12	1.10	0.74	0.83	
MH2 α	0.86	0.71	1.10	1.27	0.82	1.04	1.41	0.76	0.87	0.89	1.35	
MH1D	-0.44	0.60	-1.32	-0.17	0.44	0.86	-0.79	-0.27	-0.22	0.72	0.44	
MH2D	0.35	0.80	-0.22	-0.56	0.46	-0.09	-0.81	0.64	0.32	0.26	-0.70	
MH1 χ^2	0.33	0.56	4.17	0.02	0.33	1.75	1.24	0.06	0.05	1.05	0.33	
MH2 χ^2	0.03	0.32	0.02	0.60	0.22	0.00	1.38	0.05	0.04	0.01	0.76	
Ítems	12	13	14	15	16	17	18	19	20	21	22	23
MH1 α	1.27	0.62	0.88	0.79	1.61	0.86	0.95	1.04	1.21	1.24	1.62	0.77
MH2 α	1.18	0.77	1.12	1.30	1.33	1.14	0.92	0.79	0.89	0.91	0.68	0.74
MH1D	-0.56	1.13	0.31	0.57	-1.12	0.37	0.13	-0.09	-0.45	-0.51	-1.13	0.62
MH2D	-0.39	0.63	-0.27	-0.61	-0.66	-0.30	0.20	0.55	0.28	0.22	0.90	0.70
MH1 χ^2	0.40	2.51	0.10	0.58	2.58	0.21	0.00	0.00	0.39	0.10	0.68	0.39
MH2 χ^2	0.02	0.22	0.03	0.44	0.62	0.07	0.01	0.48	0.02	0.06	0.86	0.76

El símbolo «-» se debe a que no es posible calcular el estadístico requerido debido al número insuficiente de casos en algunos niveles de puntuaciones para calcular el estadístico MH; HMNV: Habilidad Mental no Verbal; LF: Lengua familiar; LE: Lengua de escolarización; C: Castellano; V: Valenciano; MH α : cociente de razones común; MH-D: estadístico delta; MH χ^2 : test de significación estadística; MH1 α : cociente de razones común en el grupo inferior a la media; MH2 α : cociente de razones común en el grupo superior a la media; MH1D: estadístico delta en el grupo inferior a la media; MH2D: estadístico delta en el grupo superior a la media; MH1 χ^2 : test de significación estadística en el grupo inferior a la media; MH2 χ^2 : test de significación estadística en el grupo superior a la media; * p \leq 0.005; [†] MHD \geq 1.5l.

la Comunidad Valenciana. Este resultado y los obtenidos por otros investigadores (Hamers, Hessels y Van Luit, 1991; Hamers, Hessels y Pennings, 1996) sugieren que los tests compuestos por ítems gráficos pueden ser una alternativa para medir la capacidad intelectual de sujetos con lenguas diferentes a la del grupo para el que fue desarrollado el test.

Por último, queremos señalar que este trabajo de investigación presenta algunas limitaciones. Al tratarse de un estudio empírico, no podemos conocer la cantidad de falsos positivos (FP) y falsos negativos (FN) incluidos en los resultados. En este caso, hemos optado por controlar la cantidad de FP estableciendo un nivel de significación estadística más restrictivo (0.05/n° de ítems de la escala: $0.05/23=0.0021$) con el fin de asegurar que los ítems detectados realmente presentaban DIF. La segunda limitación se refiere

al tamaño de los grupos estudiados. Una característica común de los estudios empíricos es la escasa disponibilidad de sujetos en el grupo focal y, en ocasiones, en el grupo de referencia. En este estudio, el tamaño muestral utilizado es suficiente para aplicar el estadístico MH. No obstante, en futuros estudios sería conveniente ampliar los tamaños muestrales de los distintos grupos, y de este modo, aplicar otros métodos de detección de DIF, en especial los basados en la Teoría de la Respuesta al Ítem.

Agradecimientos

Este trabajo ha sido realizado gracias a un proyecto de investigación financiado por la Universitat de València (Proyecto UV00-4132).

Referencias

- Candell, G. y Hulin, Ch. (1987). Cross-Language and Cross-Cultural comparisons in scale translations: Independent sources of information about item functioning. *Journal of Cross-Cultural Psychology*, 17 (4), 417-440.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. NY: Academic Press.
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students in California State Department of Education. En *Schooling and language minority students: A theoretical framework*. LA. Evaluation, Dissemination and Assessment Center.
- Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. En C. Rivera: *Language proficiency and academic achievement*, Clevedon, Multilingual Matters.
- Drasgow, F. y Hulin, C.L. (1986). *Assessing the equivalence of measurement of attitudes and aptitudes across heterogeneous subpopulations*. Unpublished manuscript. University of Illinois at Champaign-Urbana.
- Elosúa, P., López, A. y Egaña, J. (2000). Idioma de aplicación y rendimiento en una prueba de comprensión verbal. *Psicothema*, 12, (2), 201-206.
- Ferreres, D. (1998). *Funcionamiento diferencial de los ítems de una prueba de aptitud intelectual en función de la lengua familiar y la lengua de escolarización*. Tesis doctoral no publicada. Universitat de València.
- Ferreres, D., González-Romá, V. y Gómez-Benito, J. (2000). Comparación del estadístico Mantel-Haenszel y la regresión logística en el funcionamiento diferencial de los ítems en dos pruebas de aptitud intelectual en un contexto bilingüe. *Psicothema*, 12, (2), 214-219.
- Fidalgo, A. (1994). MHDIF: A computer program for detecting uniform and nonuniform differential item functioning with the Mantel-Haenszel procedure. *Applied Psychological Measurement*, 18 (3), 300.
- Fidalgo, A.M. (1996). Funcionamiento diferencial de los ítems. En J. Muñiz (Ed.), *Psicometría* (pp. 371-455). Madrid: Universitat.
- Fidalgo, A.M. y Ferreres, D. (en prensa). Supuestos y consideraciones en los estudios empíricos sobre el funcionamiento diferencial de los ítems. *Psicothema*.
- Fidalgo, A.M., Mellenbergh, G.J. y Muñiz, J. (1998). Comparación del procedimiento Mantel-Haenszel frente a los modelos loglineales en la detección del funcionamiento diferencial de los ítems. *Psicothema*, 10, 219-228.
- Fidalgo, A.M., Mellenbergh, G.J. y Muñiz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *Methods of Psychological Research Online*, 5 (3), 43-53. Internet: <http://www.mpr-online.de>.
- Hambleton, R.K. (1994). Guidelines for adapting educational and psychological tests. A progress report. *European Journal of Psychological Assessment*, 10 (3), 39-52.
- Hamers, J.H.M., Hessels, M.G.P. y Pennings, A.H. (1996). Learning potential in ethnic minority children. *European Journal Psychological Assessment*, 12 (3), 183-192.
- Hamers, J.H.M., Hessels, M.G.P. y Van Luit, J.E.H. (1991). Leertest voor Etnische Minderheden: Test en Handleiding [*Learning potential test for ethnic minorities: Test and Manual*]. Lisse: Swets & Zeitlinger.
- Holland, P.W. y Thayer, D.T. (1986). *Differential item functioning and the Mantel-Haenszel procedure*. Princeton, NJ: Educational Testing Service.
- Hulin, Ch.L. y Mayer, L.J. (1986). Psychometric Equivalence of a Translation of the Job Descriptive Index into Hebrew. *Journal of Applied Psychology*, 71 (1), 83-94.
- Mazor, K., Clauser, B. y Hambleton, R.K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel Procedure. *Educational and Psychological Measurement*, 54 (2), 284-291.
- Reise, S.P., Smith, L. y Furr, R.M. (2001). Invariance on the NEO PI-R Scale. *Multivariate Behavioral Research*, 36 (1), 83-110.
- Resing, W.M., Bleichrodt, N. y Drenth, P.D. (1986). Het gebruik van de RAKIT bij allochtoon etnische groepen. *Nederlands Tijdschrift voor de Psychologie*, 41, 179-188.
- Smith, P.C., Kendall, L.M. y Hulin, C.L. (1969). *Measurement of satisfaction in work and retirement*. Chicago: Rand-McNally.
- van de Rijt, B. (1990). *Reactiesnelheidstest. Een aanvulling voor allochtonen op de bestaande intelligentietests*. Unpublished master's thesis. Tilburg University.
- van de Vijver, F.J.R. y Poortinga, Y.H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13 (1), 29-37.
- Yuste, C. (1988). *Batería de aptitudes diferenciales y generales. BADIY-Elemental*. Madrid: CEPE.
- Yuste, C. (1995). *BADIY-Elemental (2ª edición)*. Madrid: CEPE.