

Identificación de Ítems con Funcionamiento diferencial en SABER 11°

Martha Ligia Cuevas Mendoza* - Aura Nidia Herrera Rojas**
Departamento de Psicología
Universidad Nacional de Colombia

Introducción

SABER 11° es uno de los criterios usados en instituciones de educación superior para seleccionar a sus estudiantes. En consecuencia, si se quiere controlar un posible efecto de factores de inequidad en el ingreso a la universidad debe considerarse la identificación de cualquier tratamiento desigual para grupos que de una u otra forma se han denominado desfavorecidos o minoritarios. Uno de dichos grupos se encuentra compuesto por personas pertenecientes a comunidades indígenas, quienes difieren culturalmente de la mayoría de la población colombiana.

Una de las formas en que se pueden presentar desigualdades en las pruebas se refiere a lo que psicometría se conoce como sesgo en los ítems. El paso inicial para la detección de ítems con sesgo es realizar un análisis del Funcionamiento Diferencial del Ítem (DIF).

La importancia de identificar ítems con DIF en este examen tomando como grupo focal (grupo supuestamente desfavorecido o minoritario) a los evaluados provenientes de resguardos indígenas, surge en dos sentidos: la identificación de posibles fuentes de sesgo cultural que se sumen a las ya encontradas en otros estudios, y en un sentido más práctico, y tal vez el más importante, contribuir para mejorar procedimientos y evitar interferencia de variables no relacionadas con el objeto de medida.

Algunos conceptos importantes

Impacto: diferencias reales en el desempeño en pruebas entre grupos diferentes (Barbero & Prieto, 1997, Gomez & Navas, 1998).

Sesgo: "Error sistemático que distorsiona el significado de las puntuaciones y que está causado por la intervención de habilidades espurias junto a la habilidad principal en un ítem" (Ackerman, 1992; Mellenbergh 1989; Shealy & Stout, 1993, citados por Elosua, Lopez, & Torres, 2000, p. 198).

Grupo focal: grupo que está bajo investigación y que se presume generalmente como desfavorecido.

Grupo de referencia: grupo de la población general, el cual se supone tiene mejor desempeño y no es el grupo de estudio.

DIF: un ítem presenta DIF cuando la probabilidad de responder correctamente no depende únicamente del nivel de habilidad de la persona en el rasgo medido por la prueba sino también de otras variables, dicha probabilidad puede ser diferente para personas pertenecientes a grupos distintos, infringiendo de esta forma el supuesto de invarianza de medida (Elosua, 2006).

Objetivos y preguntas de investigación

- ¿Existen ítems con posible DIF entre indígenas y no indígenas en el examen de Estado SABER 11° aplicado en 2006-2 y 2006-1?
- Aplicar una rutina propuesta al ICFES, resultado de la primera fase del proyecto de investigación: "Identificación de ítems con Sesgo Cultural en las Pruebas de los Exámenes de Estado en Colombia" del grupo Métodos e Instrumentos en ciencias del comportamiento, con el fin de identificar los ítems con DIF en el examen de Estado llevado a cabo en 2006-2 y 2006-1, teniendo como grupo focal a los evaluados indígenas.

* mlcuevasm@unal.edu.co
** anherrerar@unal.edu.co

Metodología

Muestra

- N 2006: 410548. N 2007: 67703 .
- Razón de tamaños entre los grupos 2006: 1/123 y 2007: 1/88.
- 20 muestras aleatorias usando bootstrapping.

Pruebas

Biología, filosofía, física, lenguaje, matemáticas, química y ciencias sociales.

Procedimientos

Mantel-Haenszel:

$$A = |\Delta MH| < 1$$
$$B = 1,5 > |\Delta MH| > 1$$
$$C = |\Delta MH| > 1,5$$

Diferencia de la dificultad y Metrica (Berrío, 2008)

$$D = \left[\ln \left(\frac{b_{r15,2.5}}{b_{f15,2.5}} \right) \right] * -2.35$$

Punto de corte: 0.22.

Análisis de datos

- Estadísticos descriptivos de las tasas de detección general de DIF y por las diferencias en la distribución de la habilidad entre grupos (D de Cohen) y ajuste al modelo.
- Comparación entre procedimientos y Concordancia: Kappa.

Resultados

Variable y prueba	Tasa promedio de ítems detectados a través de las muestras					
	Diferencia de la Dificultad		Mantel-Haenszel (B y C)		Ambos procedimientos	
	2006-2	2007-1	2006-2	2007-1	2006-2	2007-1
General						
Biología	0,6%	9,3%	0,0%	5,4%	0,0%	3,2%
Filosofía	0,2%	0,0%	0,0%	2,2%	0,0%	0,0%
Física	0,0%	0,6%	0,2%	1,2%	0,0%	0,4%
Lenguaje	7,1%	9,9%	0,2%	9,7%	0,2%	6,7%
Matemáticas	9,1%	14,5%	0,4%	2,6%	0,4%	2,2%
Química	2,4%	2,6%	0,0%	3,4%	0,0%	2,0%
Ciencias sociales	6,0%	1,7%	3,5%	2,7%	3,5%	0,8%
Impacto						
No Impacto						
Biología	0,0%	1,6%	0,0%	1,4%	0,0%	1,2%
Filosofía	0,0%	0,0%	0,0%	2,2%	0,0%	0,0%
Física	0,0%	0,6%	0,2%	1,2%	0,0%	0,4%
Lenguaje	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Matemáticas	0,0%	0,8%	0,0%	0,0%	0,0%	0,0%
Química	0,6%	2,6%	0,0%	3,4%	0,0%	2,0%
Ciencias sociales	0,0%	0,0%	0,0%	0,6%	0,0%	0,0%
Impacto pequeño						
Biología	0,6%	7,7%	0,0%	4,0%	0,0%	2,0%
Filosofía	0,2%	0,0%	0,0%	0,0%	0,0%	0,0%
Física	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Lenguaje	7,1%	9,9%	0,2%	9,7%	0,2%	6,7%
Matemáticas	9,1%	13,7%	0,4%	2,6%	0,4%	2,2%
Química	1,8%	0,0	0,0%	0,0%	0,0%	0,0%
Ciencias sociales	6,0%	1,7%	3,5%	2,1%	3,5%	0,8%

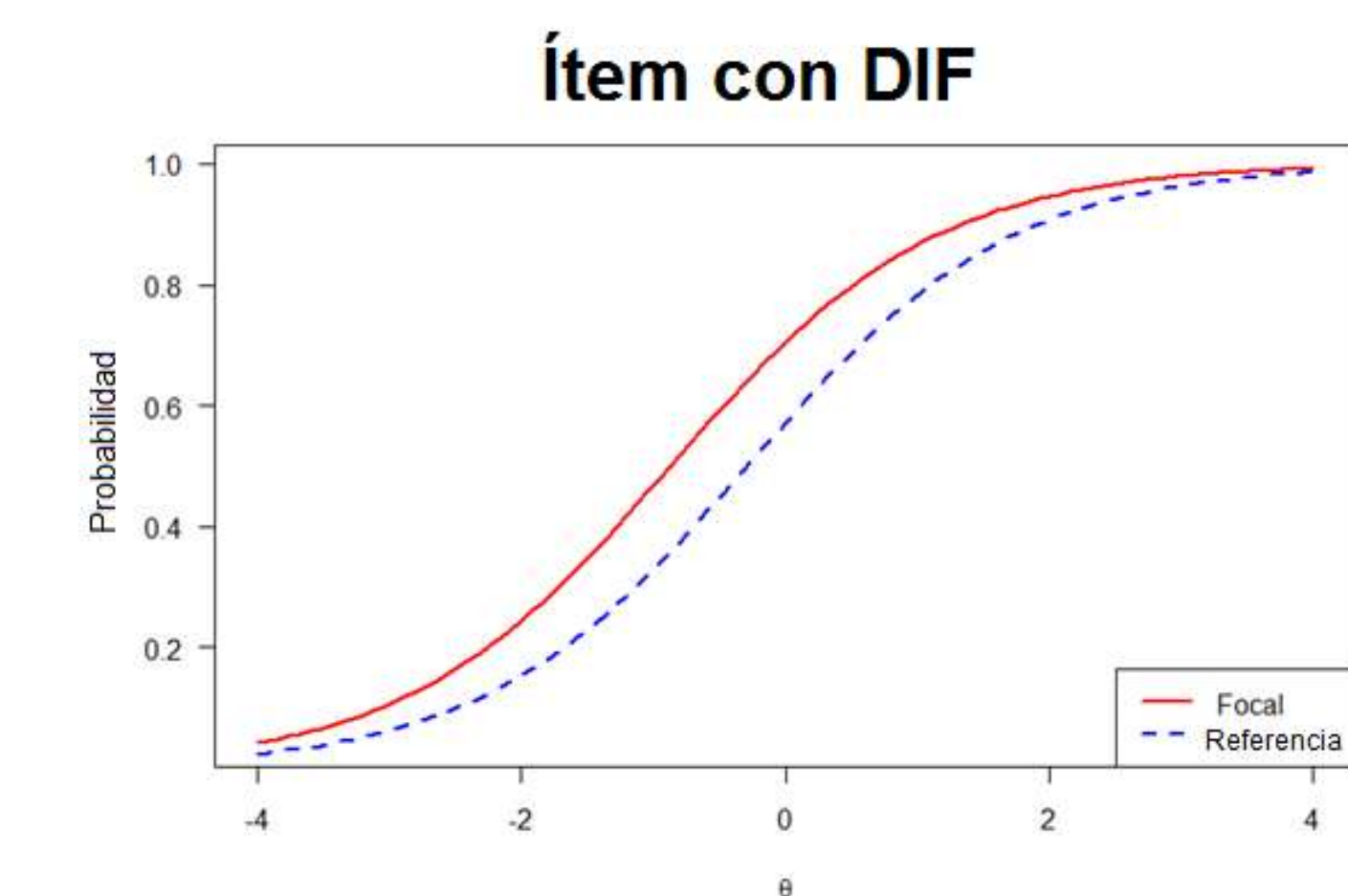
Referencias

Arias, E. (2008). *Efecto de la Razón de Tamaño y el Ajuste del Modelo sobre el Estadístico Mantel-Haenszel y su Métrica Delta en la Detección de DIF*. Tesis de Maestría. Departamento de Psicología, Universidad Nacional de Colombia, Bogotá Colombia.

Berrío, A. I. (2008). *La Razón de Tamaños de muestra y Desajustes del Modelo en la Detección de Ítems con Funcionamiento Diferencial mediante el Procedimiento de Diferencia de Dificultad*. Tesis de Maestría. Departamento de Psicología, Universidad Nacional de Colombia, Bogotá, Colombia.

Elosua, P. (2006). Funcionamiento diferencial del ítem en la evaluación internacional PISA. Detección y comprensión. *RELIEVE*. 12(2), 247-259.

Elosua, P., López, A. & Torres, E. (2000). Desarrollos didácticos y funcionamiento diferencial de los ítems. Problemas inherentes a toda investigación empírica sobre sesgo. *Psicothema*. 12(2), 198-202.



Índice Kappa para ítems detectados a través de las muestras entre Mantel-Haenszel (B y C) y diferencia de la dificultad

	Entre pruebas estadísticas	Entre métricas ΔMH y D
General	0,733** (0,008)	0,394** (0,028)
2006-2	0,658** (0,015)	0,3** (0,047)
2007-1	0,786** (0,010)	0,438** (0,035)

Errores estándar en paréntesis

**p<0,01

Con ambos procedimientos se encuentra que ciencias sociales muestra la mayor tasa de detección en 2006-2 y lenguaje en 2007-1. Teniendo en cuenta el impacto, se encontró una tasa más alta cuando hubo poco impacto (d de Cohen entre 0,2 y 0,5) que cuando no existía. En cuanto al ajuste del modelo, debido a que la mayoría de ítems se ajustaron a través de las muestras, no fue posible establecer una relación entre el modelo de ajuste y la tasa de detecciones.

En cuanto al acuerdo entre los dos procedimientos, se observó la prueba de hipótesis, presenta un buen acuerdo (K = 0,733, p <0,01); sin embargo, el acuerdo entre las métricas es menor y para 2006-2. Se sugiere la revisión de 20 ítems que fueron identificados al menos una vez por los dos procedimientos a través de muestras.

Discusión

La presencia de las tasas de detección más altas en 2007-1 puede estar relacionada con una razón de tamaños menor entre los grupos focal y de referencia que la que se observó en 2006-2, algo ya señalado por Berrío (2008) y Arias (2008) en relación con la diferencia de la dificultad y el MH, respectivamente. Cuando hubo impacto pequeño la tasa de detección creció, lo que corresponde a las conclusiones de Berrío (2008) y Arias (2008). Aunque había ítems con DIF, esto no quiere decir que sean ítems sesgados, para hablar sobre sesgo es necesaria una revisión de expertos en las áreas evaluadas y en las diferencias culturales entre indígenas y el resto de la población con el fin de identificar posibles fuentes de DIF. Si estas fuentes de DIF son relevantes para los atributos a medir, los ítems con DIF son imparciales y no afectan a la validez de la prueba.

