

NCDIF y Regresión logística como procedimientos para detectar funcionamiento diferencial del ítem en Saber 5 y 9

Víctor H. Cervantes - Martha Ligia Cuevas
Subdirección de Estadística

Instituto Colombiano para la Evaluación de la Educación - ICFES
vcervantes@icfes.gov.co - mcuevas@icfes.gov.co

II Seminario Internacional de Investigación sobre la Calidad de la Educación

Noviembre 3, 2011

- 1 Antecedentes
Objetivos
- 2 Estudio de simulación
Diseño
- 3 Resultados
- 4 Discusión
Discusión
Recomendaciones
Limitaciones
Futuras Investigaciones
Referencias

- Propiedades de los instrumentos de medición y pruebas
- Validez
- Interpretaciones iguales de las puntuaciones por género, zona, sector y nivel socioeconómico.
- Equidad
- Funcionamiento Diferencial del Ítem (DIF o FDI)
- Sesgo
- Impacto
- Ejemplos
- SABER 5 and 9
- Procedimientos específicos

- El índice no compensatorio de DIF (NCDIF) ha sido propuesto en el marco del DIF y del funcionamiento diferencial de pruebas (DFT o FDT) como una alternativa basada en la teoría de respuesta al ítem (TRI) para la detección de DIF (Raju, van der Linden & Fler, 1995).
- Es definido como:

$$\text{NCDIF}_i = \int_{-\infty}^{\infty} [P_{iF}(\theta) - P_{iR}(\theta)]^2 f_F(\theta) d\theta \quad (1)$$

- El NCDIF puede verse afectado NCDIF por variables tales como razón de tamaños entre los grupos que se comparan, tamaño de muestra, y parece estar relacionado con parámetros específicos de los ítems como ciertos rangos de dificultad y discriminación (Bolt, 2002; Oshima, Raju & Nanda, 2006).

- Se ha encontrado que la prueba estadística propuesta para probar la hipótesis de no DIF no se aproxima a la distribución χ^2 llevando a inflar las tasas de error tipo I para muestras grandes con datos simulados bajo la hipótesis nula (Raju et al, 2006).
- Algunos puntos de corte han sido propuestos (p. ej. 0.006, Raju et al, 2006)
- Aproximaciones de Monte Carlo (Bolt, 2002; Oshima et al, 2006) para decidir si un ítem debe ser considerado con DIF o no.
- Tamaños de muestra grandes llevan a detectar pequeñas desviaciones de la hipótesis nula, por lo que se recomienda evaluar aspectos relacionados con la importancia práctica de los efectos observados (Gómez-Benito, Hidalgo & Padilla, 2009).

- Psicología y educación.
- Coeficientes para el grupo, puntuación total y su interacción.
- Significancia.
- Variable dependiente: respuesta de los examinados.
- Suele verse afectado por razón de tamaños de muestra, tamaño de muestra, dificultad y discriminación (Herrera, 2005; Santana, 2009)
- RL:

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta + \beta_2g + \beta_3\theta g \quad (2)$$

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta + \beta_2g \quad (3)$$

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta \quad (4)$$

- Identificar algunas variables que pueden afectar el NCDIF y la RL
- Encontrar los mejores puntos de corte para el NCDIF y el ΔR^2 en la detección de DIF, tomando en cuenta algunas variables que los afectan

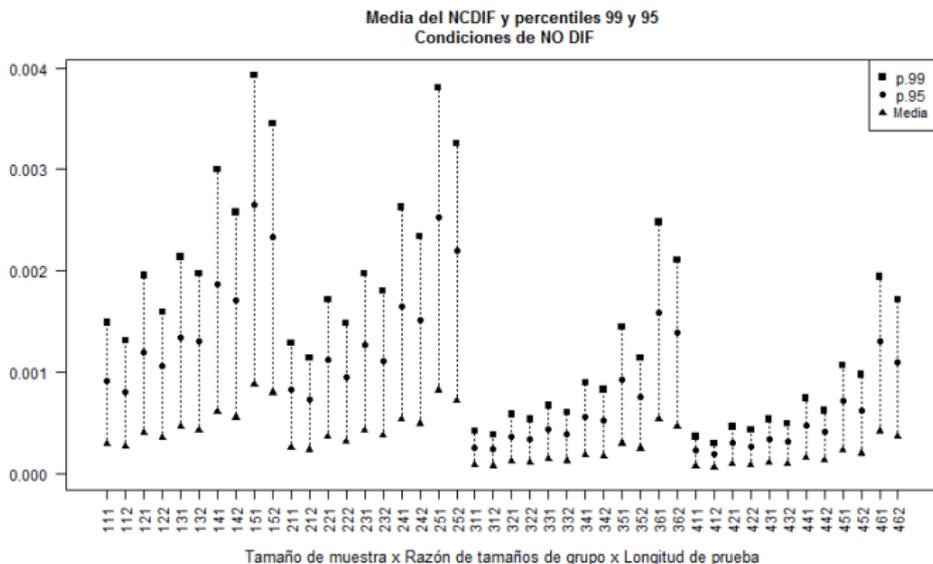
- 1 Antecedentes
Objetivos
- 2 Estudio de simulación
Diseño
- 3 Resultados
- 4 Discusión
Discusión
Recomendaciones
Limitaciones
Futuras Investigaciones
Referencias

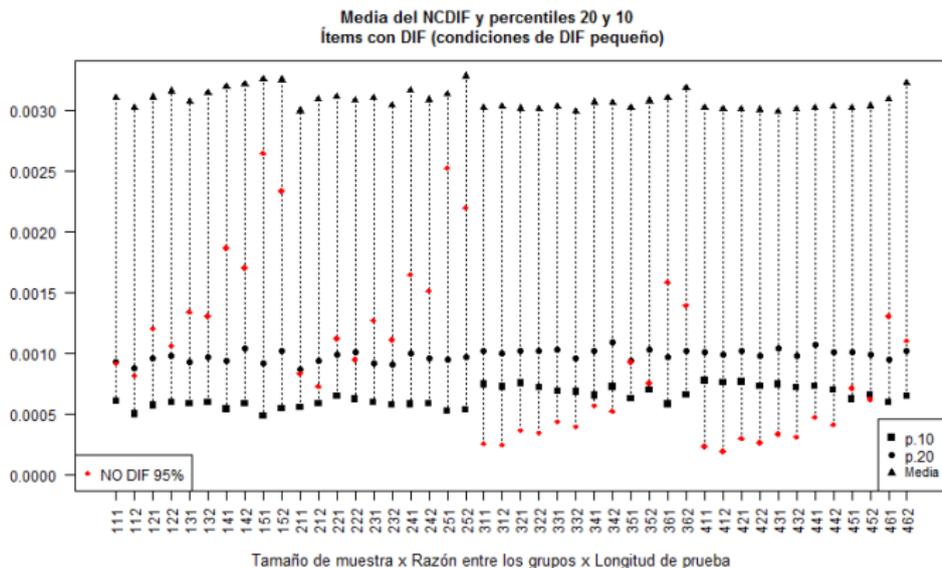
- Tamaño de muestra: $n = 7500, 8500, 26000$ y 33000 .
- Razón de tamaños entre los grupos para NCDIF: 1, 3, 4, 6, 10 y 20.
- Razón de tamaños entre los grupos para RL: 1, 3, 4, 6, 10, 20, 150 y 250.
- Longitud de la prueba: $K = 18$ y 26 .
- Magnitud de DIF: No DIF, dos ítems con DIF (pequeño y grande), y cuatro ítems con DIF (pequeño y grande).
- 220 condiciones experimentales para NCDIF y 270 para RL.
- Réplicas por condición: $R = 200$.
- Dos ítems con DIF uniforme y dos ítems con DIF no uniforme cuando la condición tenía cuatro ítems.
- Variables fijas: dificultad y discriminación.

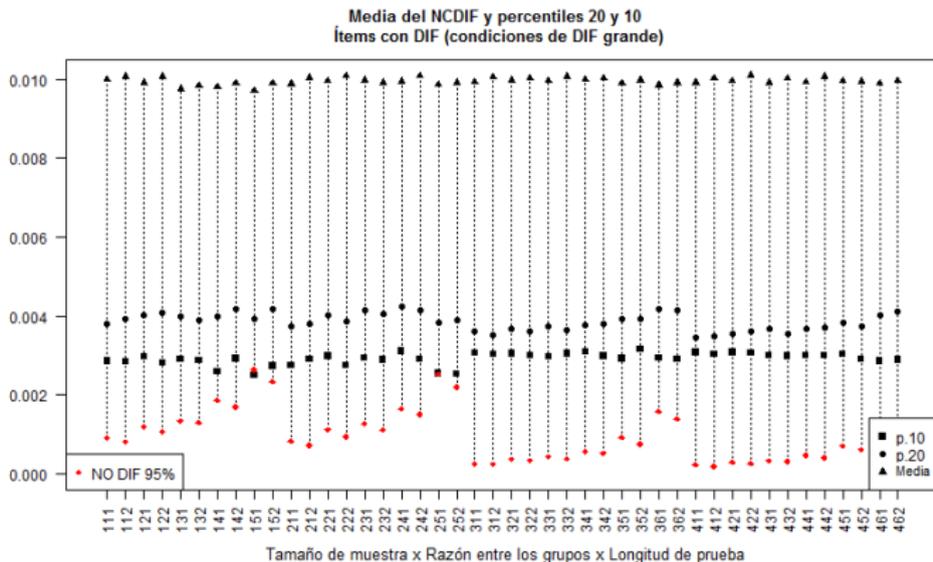
- El NCDIF fue usado como medida de tamaño del efecto y como variable dependiente para el DIF no compensatorio.
- El ΔR^2 de McFadden se usó como medida de tamaño del efecto y variable dependiente para la RL.
- Para detectar ítems con DIF uniforme se usó el ΔR^2 cuando se introducía el efecto de grupo.
- Para detectar ítems con DIF uniforme se usó el ΔR^2 cuando se introducía el efecto de la interacción.
- Modelos lineales para definir cuáles variables afectaban la medida de tamaño del efecto.

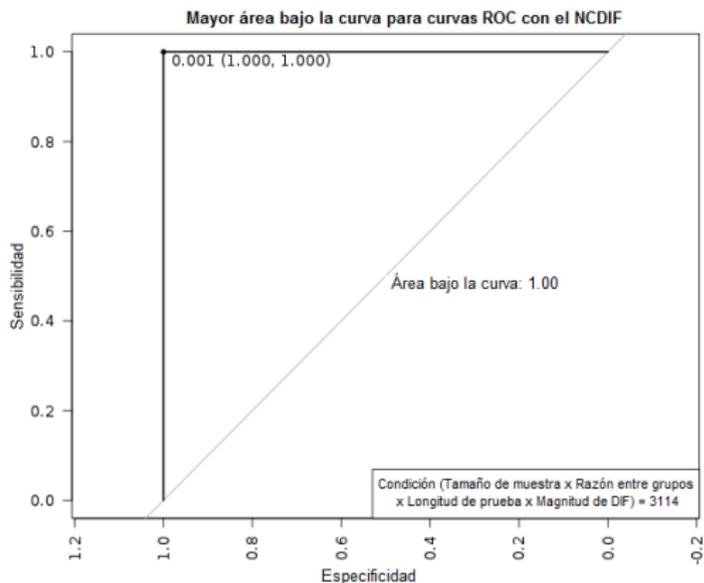
- Curvas operador receptor (COR o ROC).
- Para cada condición.
- Para cada combinación de variables con efecto sobre el NCDIF y el ΔR^2 .
- Selección de los mejores puntos de corte en forma eficiente, es decir, manteniendo buenas tasas de error tipo I y potencia.

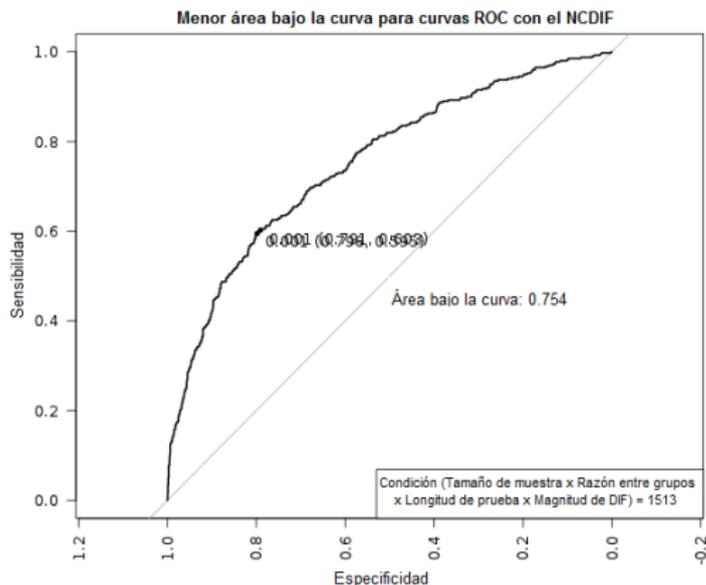
- 1 Antecedentes
Objetivos
- 2 Estudio de simulación
Diseño
- 3 Resultados**
- 4 Discusión
Discusión
Recomendaciones
Limitaciones
Futuras Investigaciones
Referencias



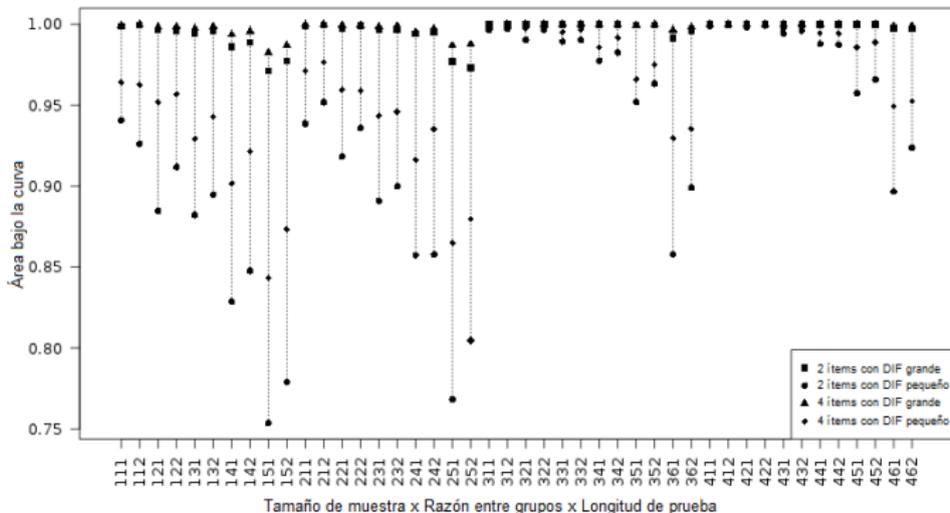




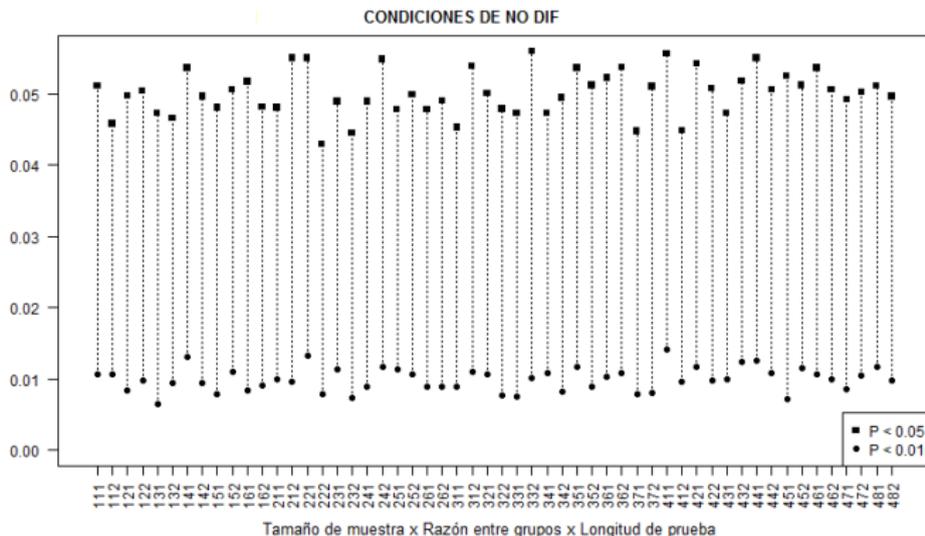




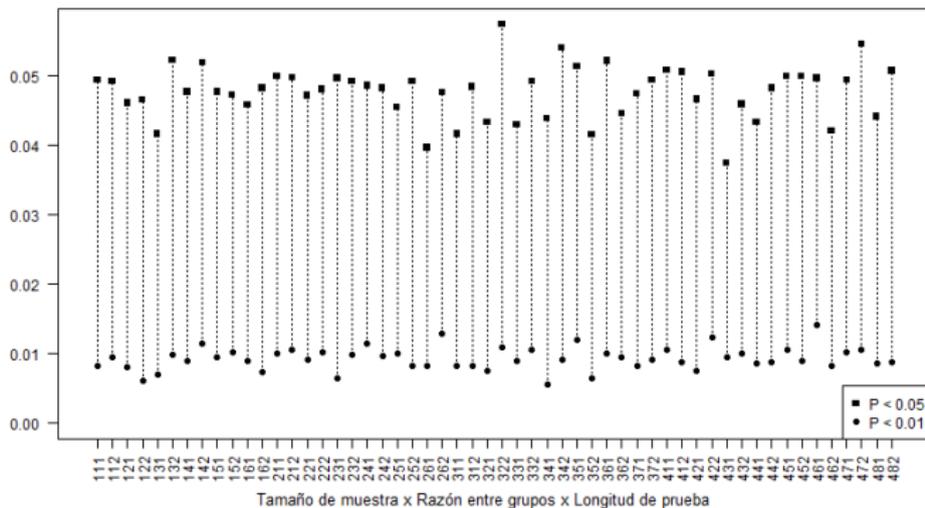
Áreas bajo la curva del NCDIF por cada condición de DIF



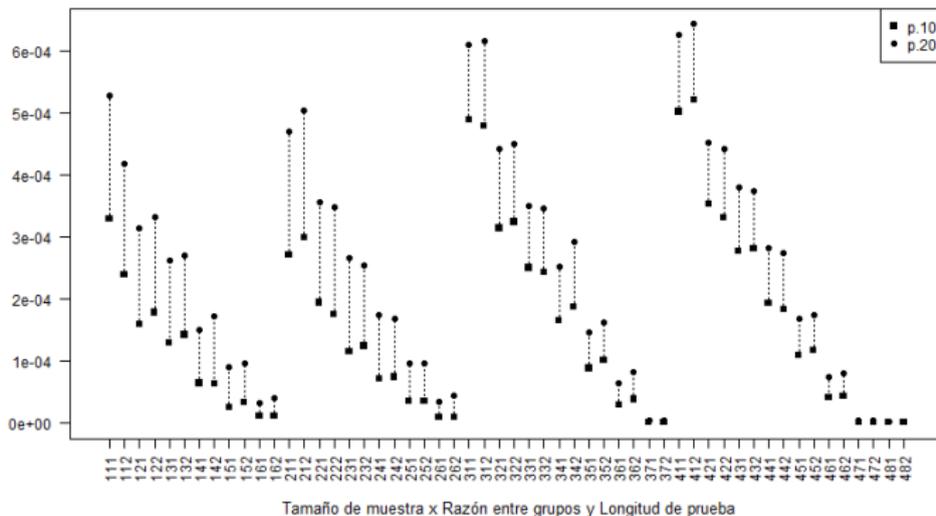
Media de rechazo (0.01, 0.05) de la prueba de DIF uniforme con regresión logística



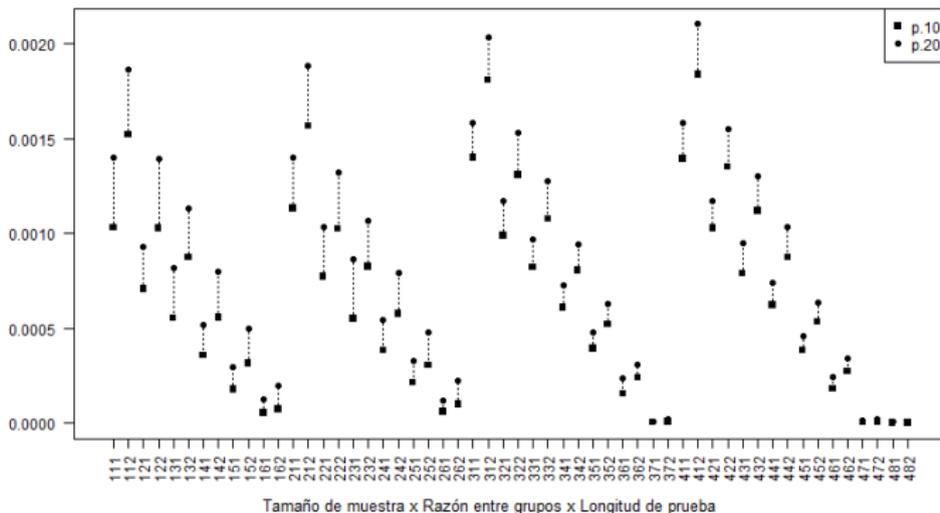
Media de rechazo (0.01, 0.05) de la prueba de DIF no uniforme con regresión logística
CONDICIONES DE NO DIF



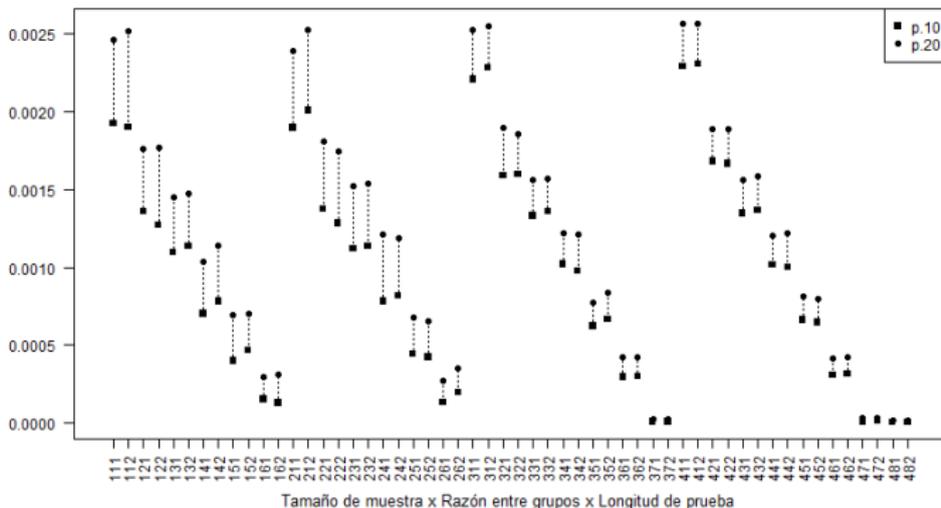
Delta para la detección de DIF uniforme con la regresión logística y percentiles 20 y 10
Items con DIF uniforme (condiciones de DIF pequeño)



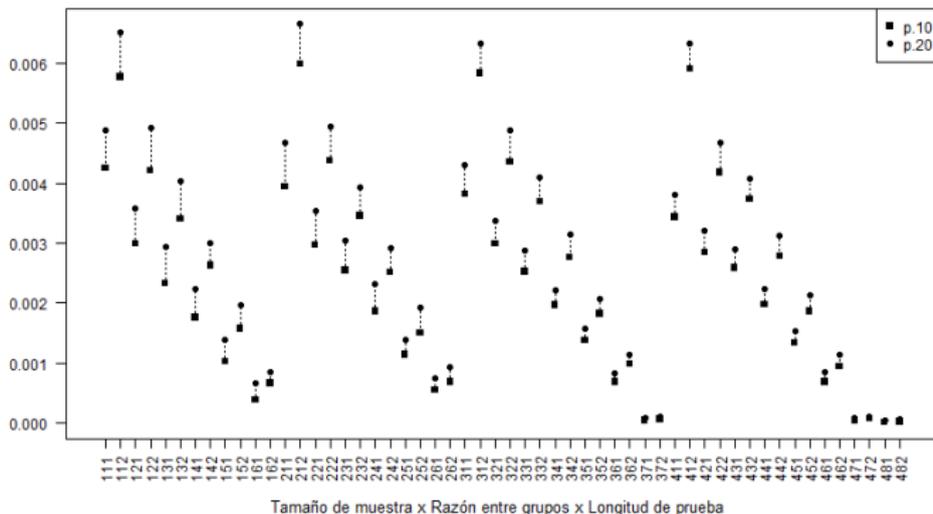
Delta para la detección de DIF no uniforme con regresión logística y percentiles 20 y 10
Items con DIF no uniforme (condiciones con DIF pequeño)



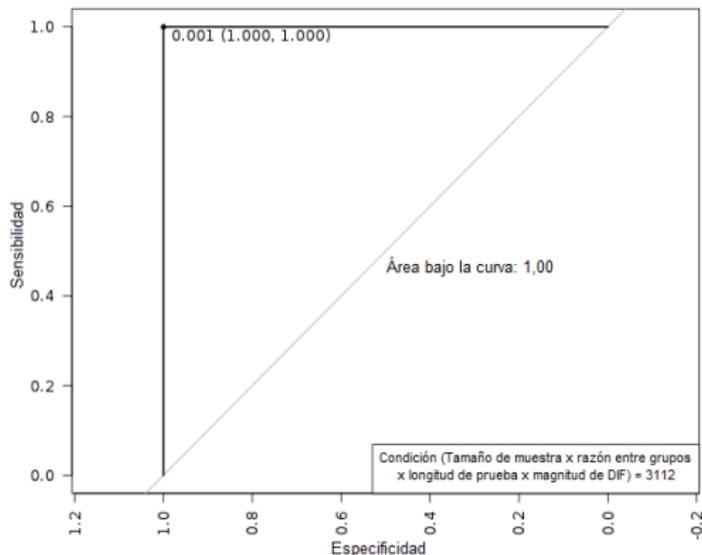
Delta para la detección de DIF uniforme con regresión logística y percentiles 20 y 10
Ítems con DIF uniforme (Condiciones con DIF grande)



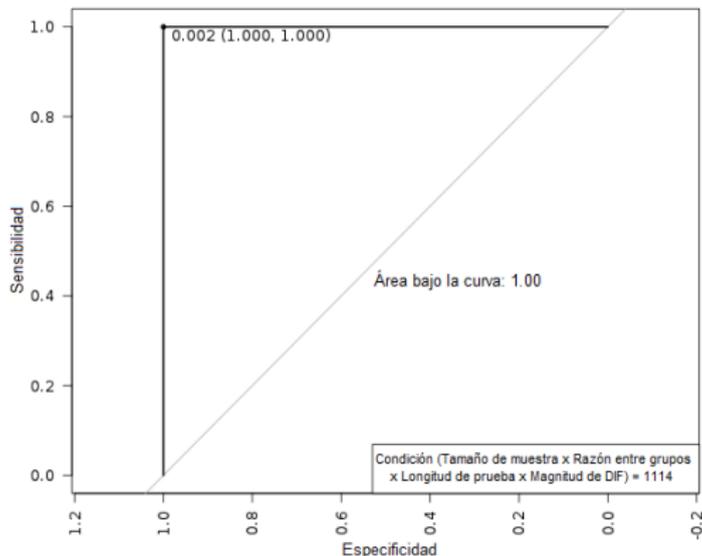
Delta para la detección de DIF no uniforme con regresión logística y percentiles 10 y 20
Ítems con DIF no uniforme (condiciones con DIF grande)



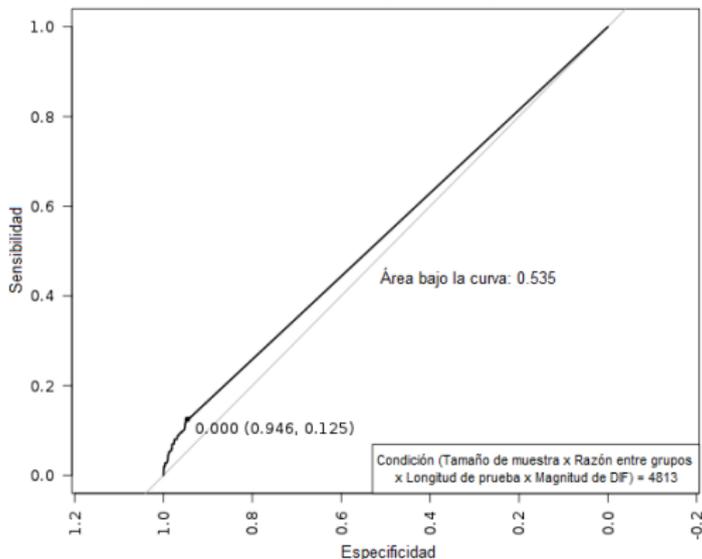
Mayor área bajo la curva de las curvas ROC con el delta para detectar DIF uniforme



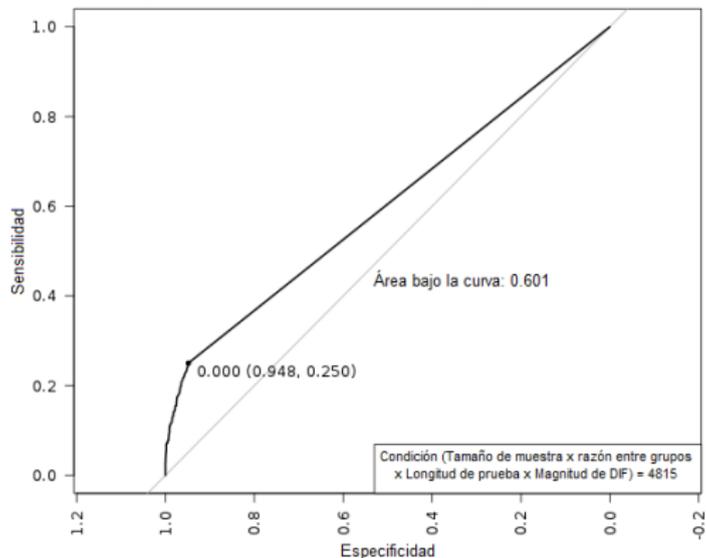
Mayor área bajo la curva de las curvas ROC con el delta para detectar DIF no uniforme



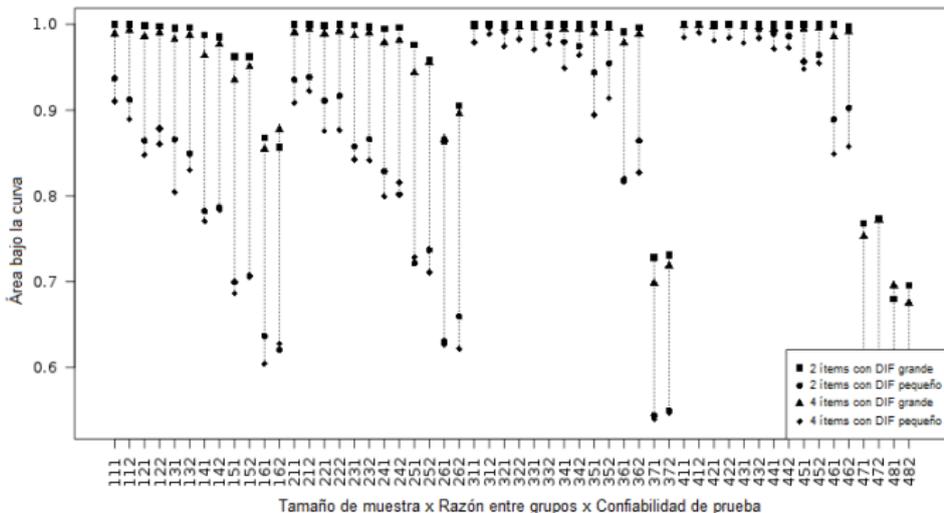
Menor área bajo la curva de las cruvas ROC para el delta para detectar DIF uniforme



Menor área bajo la curva de las curvas ROC con el delta para detectar DIF no uniforme



Áreas bajo la curva del delta para detectar DIF uniforme para cada condición de DIF



Áreas bajo la curva del delta para detectar DIF no uniforme para cada condición de DIF

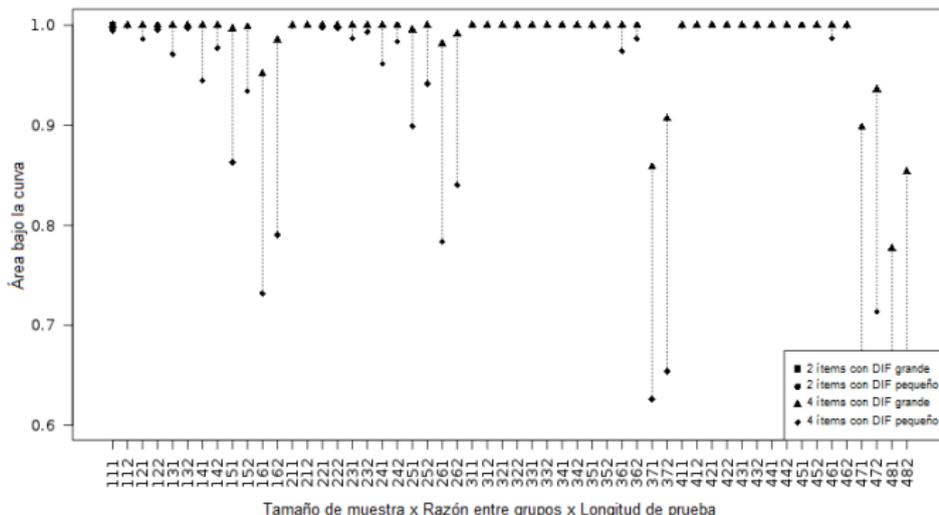


Tabla: Efectos en el NCDIF

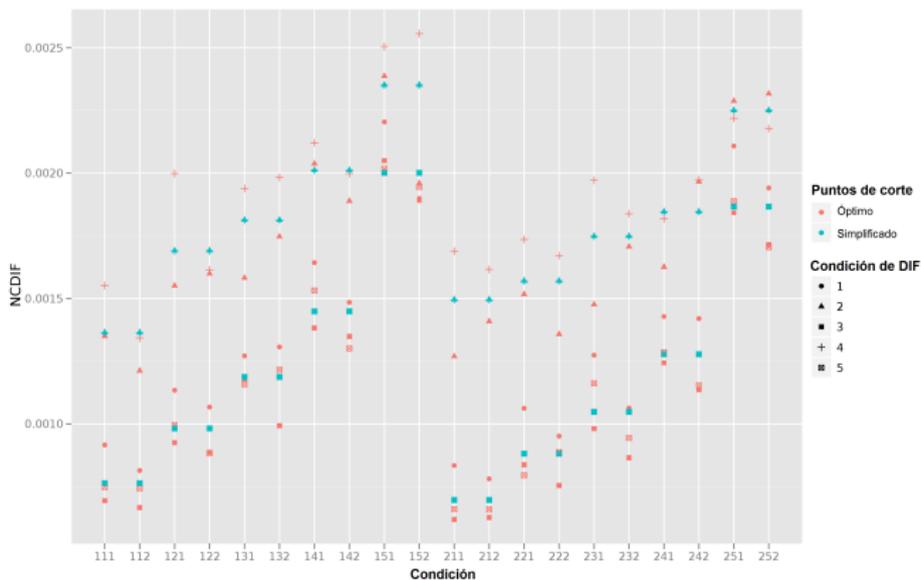
Variable	F	P
Tamaño	364.008	1.52E-195
Razón	185.955	6.48E-165
Magnitud de DIF	2.442	0.0447
Longitud de Prueba	4.733	0.030

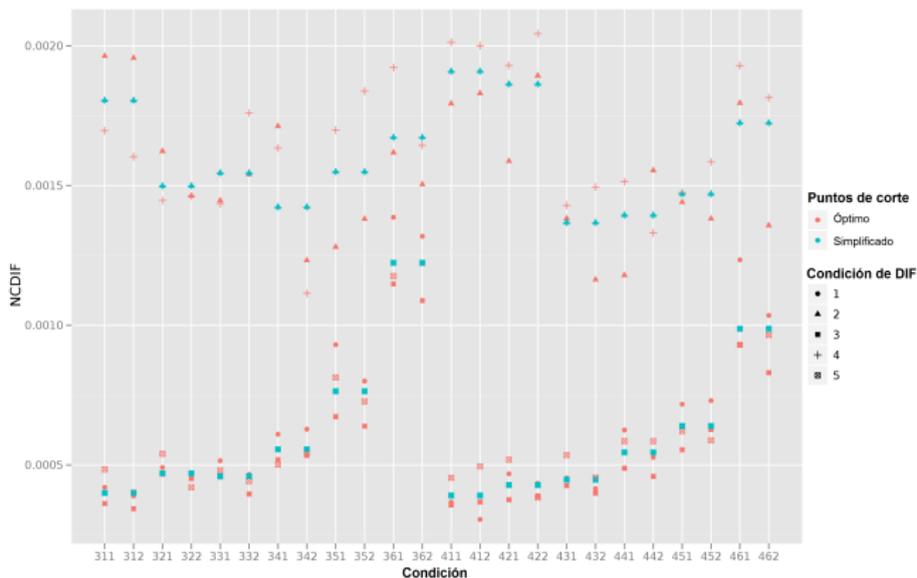
Tabla: Efectos sobre el ΔR^2 para DIF uniforme

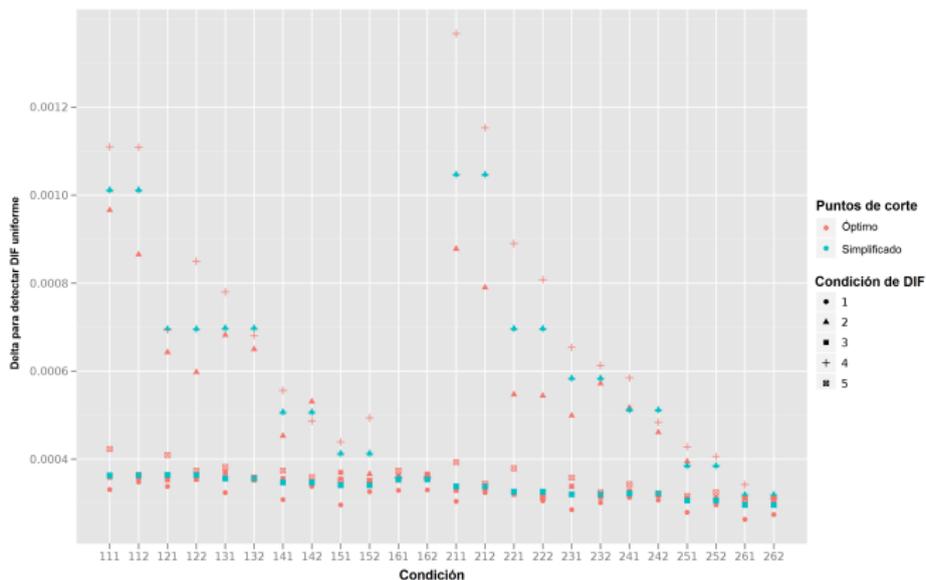
Variable	F	P
Tamaño	70.610	9.69E-44
Razón	473.287	0
Magnitud de DIF	892.215	0
Longitud de prueba	3.983	0.046
Tamaño \times Magnitud de DIF	3.778	9.64E-05
Razón \times Magnitud de DIF	17.012	3.40E-58

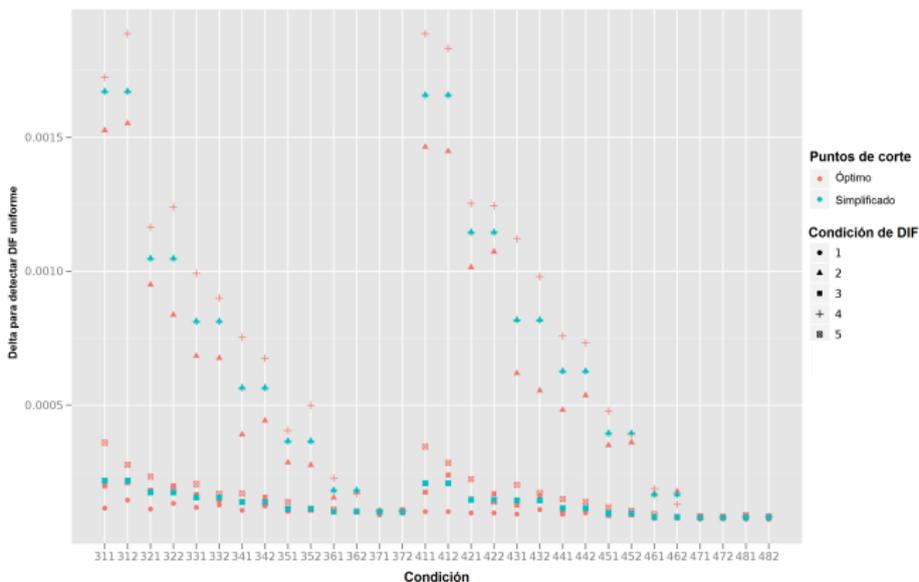
Tabla: Efectos en el ΔR^2 para DIF No Uniforme

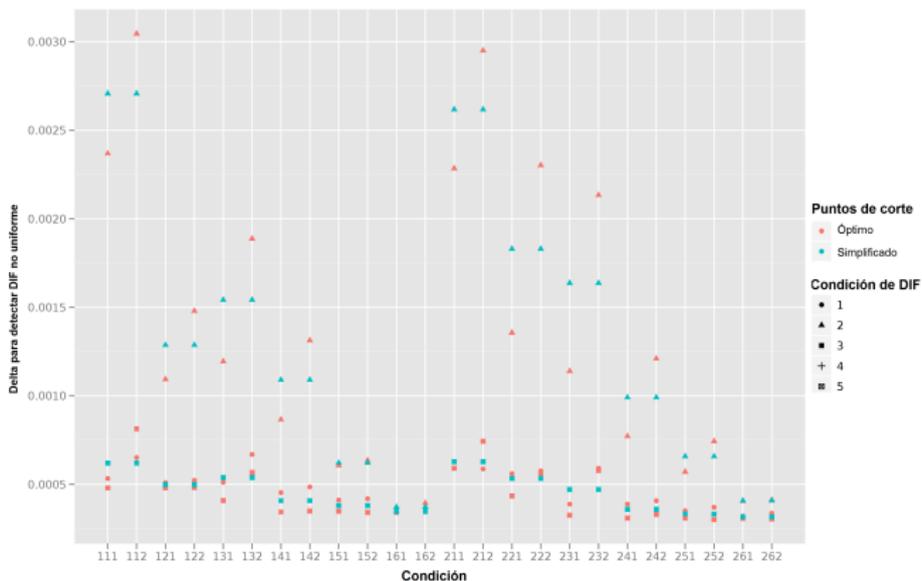
Variable	F	P
Tamaño	27.601	2.95E-17
Razón	133.511	6.49E-145
Magnitud de DIF	638.029	4.98E-113
Tamaño \times Magnitud de DIF	4.081	0.007
Razón \times Magnitud de DIF	30.403	7.42E-39
Magnitud de DIF \times Longitud de prueba	4.122	0.043

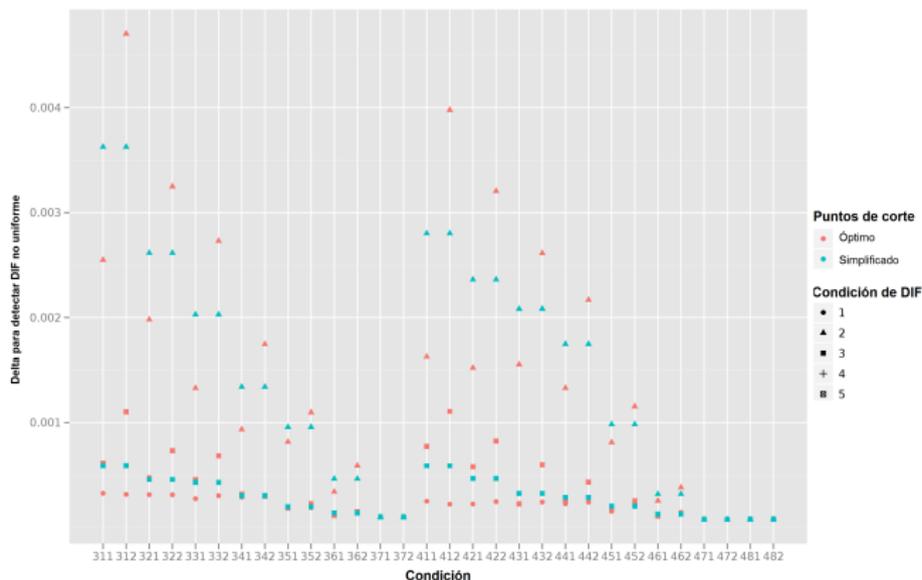












- 1 Antecedentes
Objetivos
- 2 Estudio de simulación
Diseño
- 3 Resultados
- 4 **Discusión**
Discusión
Recomendaciones
Limitaciones
Futuras Investigaciones
Referencias

- Los resultados concuerdan con otros estudios como el de Bolt (2002) y Santana (2009).
- Es necesario contar con puntos de corte específicos de las medidas de tamaño del efecto para condiciones particulares.
- Aunque algunos puntos de corte fueron obtenidos para cada condición y algunas variables tiene efecto sobre las medidas de tamaño del efecto, estos puntos de corte pudieron ser simplificados sin perder potencia ni control sobre las tasas de error Tipo I para algunas condiciones.
- El NCDIF parece poco eficiente para detectar DIF cuando uno de los grupos tiene un tamaño de muestra relativamente pequeño.

- Usar puntos de corte específicos para condiciones particulares con datos reales

- Este estudio no tomó en cuenta variables como impacto que pueden producir diferentes resultados.
- Se requieren grandes tamaños de muestra grandes para un uso apropiado del NCDIF.

- Explorar el efecto del impacto en el NCDIF y en el ΔR^2
- Explorar el tamaño de muestra mínimo requerido a nivel del ítem para el NCDIF

Gracias !!

Víctor Cervantes y Martha Cuevas
vcervantes@icfes.gov.co - mcuevas@icfes.gov.co

- Bolt, D. (2002) A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods, *Applied measurement in education*, 15,113-141.
- Gómez-Benito, J., Hidalgo, M. D., & Padilla, J. L. (2009). Efficacy of effect size measures in logistic regression. An application for detecting DIF, *Methodology*, 5, 18-25.
- Herrera, A. N. (2005) *Efecto del tamaño de muestra y razón de tamaños de muestra en la detección de funcionamiento diferencial de los ítems*. Tesis de Doctorado en Evaluación y Tecnología Informática en Ciencias del Comportamiento. Universidad de Barcelona, España.
- Oshima, T., Raju, N. & Nanda, A. A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework, *Journal of educational measurement*, 43, 1-17.
- Raju, N., van der Linden, W. & Fleer, P. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied psychological measurement*, 19, 353-368.
- Santana, A. C. (2009) *Efecto de la razón de tamaños de muestra en la detección de funcionamiento diferencial de los ítems a través del procedimiento de regresión logística*. Tesis de Maestría en Psicología. Universidad Nacional de Colombia, Colombia.