Investigación

Educación | Icfes

Análisis de DIF

en ítems politómicos en las pruebas Saber



Datos

Icfes:
Pruebas Saber
3°,5°,7° y 9°
2022-2023

ODS:

10: Reducción de las desigualdades

Objetivo

Desarrollar y aplicar una metodología para identificar y clasificar el Funcionamiento Diferencial del Ítem (DIF) en ítems politómicos de las pruebas Saber, utilizando el índice NCDIF y el Modelo de Respuesta Graduada (GRM). Asimismo, proponer ajustes al tamaño del efecto que mejoren la detección del DIF en evaluaciones con muestras grandes.

Justificación

Garantizar la comparabilidad y equidad de las evaluaciones requiere identificar ítems que funcionen de manera diferente entre grupos o aplicaciones. Este desafío es especialmente relevante en ítems politómicos y en evaluaciones a gran escala, donde los grandes tamaños de muestra pueden sobredimensionar diferencias pequeñas. Por ello, es necesario fortalecer las metodologías para detectar y clasificar adecuadamente el Funcionamiento Diferencial del Ítem (DIF).

Enfoques

Diferencial:
identifica posibles diferencias en el funcionamiento de los ítems entre grupos de estudiantes

Temporal:
compara el comportamiento de los ítems entre las aplicaciones de 2022 y 2023.

Territorial y poblacional:
la metodología permite analizar diferencias según características como zona, región o género

Resultados

En 5.°

se analizaron 44 ítems: **30 presentaron DIF despreciable,** 11 moderado y 3 grande.

En 7.°

de 51 ítems analizados, **30 tuvieron DIF despreciable,** 17 moderado y 4 grande.

En 9.°

de 55 ítems analizados, **34 presentaron DIF despreciable,** 13 moderado y 8 grande, la mayor cantidad entre los grados estudiados.



La proporción de ítems con DIF grande fue reducida: aproximadamente 6,8 % en 5.°, 7,8 % en 7.° y 14,5 % en 9.°.

El procedimiento de purificación convergió rápidamente, en 1 o 2 iteraciones, mostrando estabilidad en el proceso de equiparación.

Las constantes de equiparación mostraron valores de kA cercanos a 1 y valores de kB de −0,03, −0,16 y −0,23 para 5.°, 7.° y 9.°, respectivamente, indicando mayores diferencias de dificultad en 9.°.

Conclusiones

La metodología basada en NCDIF es adecuada para detectar y clasificar el funcionamiento diferencial en ítems politómicos y puede incorporarse a los procesos psicométricos de las pruebas Saber.

Los resultados respaldan, en términos generales, **la comparabilidad de la prueba de Acciones y Actitudes entre 2022 y 2023,** dado que se identificó un número reducido de ítems con DIF de gran magnitud.

Los ítems identificados con DIF deben complementarse con una revisión cualitativa de su contenido, para determinar si las diferencias provienen del ítem o de cambios en la población evaluada.

Recomendaciones

Los puntos de corte propuestos para el tamaño del efecto constituyen una herramienta útil para evitar que, en muestras grandes, se clasifiquen como relevantes diferencias de poca magnitud.

Se recomienda continuar validando los puntos de corte mediante estudios de simulación con distintos tamaños de muestra, distribuciones de habilidad y configuraciones de los ítems, antes de ampliar su aplicación a otros contextos.

Investigadores:

Alexander Calderón (Icfes)
Andrés Rodríguez (Icfes)
Sebastián Argüello (Icfes)

www.icfes.gov.co