

Funcionamiento Diferencial del ítem (DIF) en Módulo Latinoamericano de ICCS 2009: Evaluando la validez del instrumento en la muestra de Chile

Loreto Muñoz, Catalina Miranda & Ignacio Rossi

Pontificia Universidad Católica de Chile

Antecedentes, problematización y metodología

El proceso de validación de una prueba de conocimiento debería recolectar evidencia que sustente las interpretaciones y usos de los resultados obtenidos.

En este sentido, un componente clave de la evaluación de la validez e imparcialidad de un instrumento es el funcionamiento diferencial del ítem (DIF).

El DIF se presenta cuando dos grupos con el mismo nivel de habilidad no poseen la misma probabilidad de responder de manera correcta un ítem.

La alerta de DIF en algún ítem del instrumento podría indicar la posibilidad de sesgo y representar un problema de equidad, en la medida que los puntajes obtenidos por los examinados solo deben depender de su habilidad y no de otras características como por ejemplo raza, género o ubicación geográfica.

A partir de esto se considera relevante aplicar y comparar distintos métodos de análisis, los cuales pueden entregar información acerca del comportamiento de los ítems que componen un instrumento.

Objetivos

1. Analizar el funcionamiento diferencial de los ítems (DIF) del instrumento.
2. Identificar en qué medida el instrumento es independiente del género de los examinados.

Datos, variables y método

Se utilizan datos para Chile del módulo latinoamericano del Estudio Internacional de Educación Cívica y Formación Ciudadana (ICCS) 2009, aplicado a una muestra representativa de estudiantes de 8° básico (N=5.192).

El instrumento se compone de 16 ítems de selección múltiple, los cuales se enfocaron en aspectos del conocimiento relevantes para la región.

Los ítems se escalaron utilizando el modelo de Rasch (Rasch, 1960) y el análisis DIF se estimó a partir de dos modelos: Logístico y Análisis de facetas.

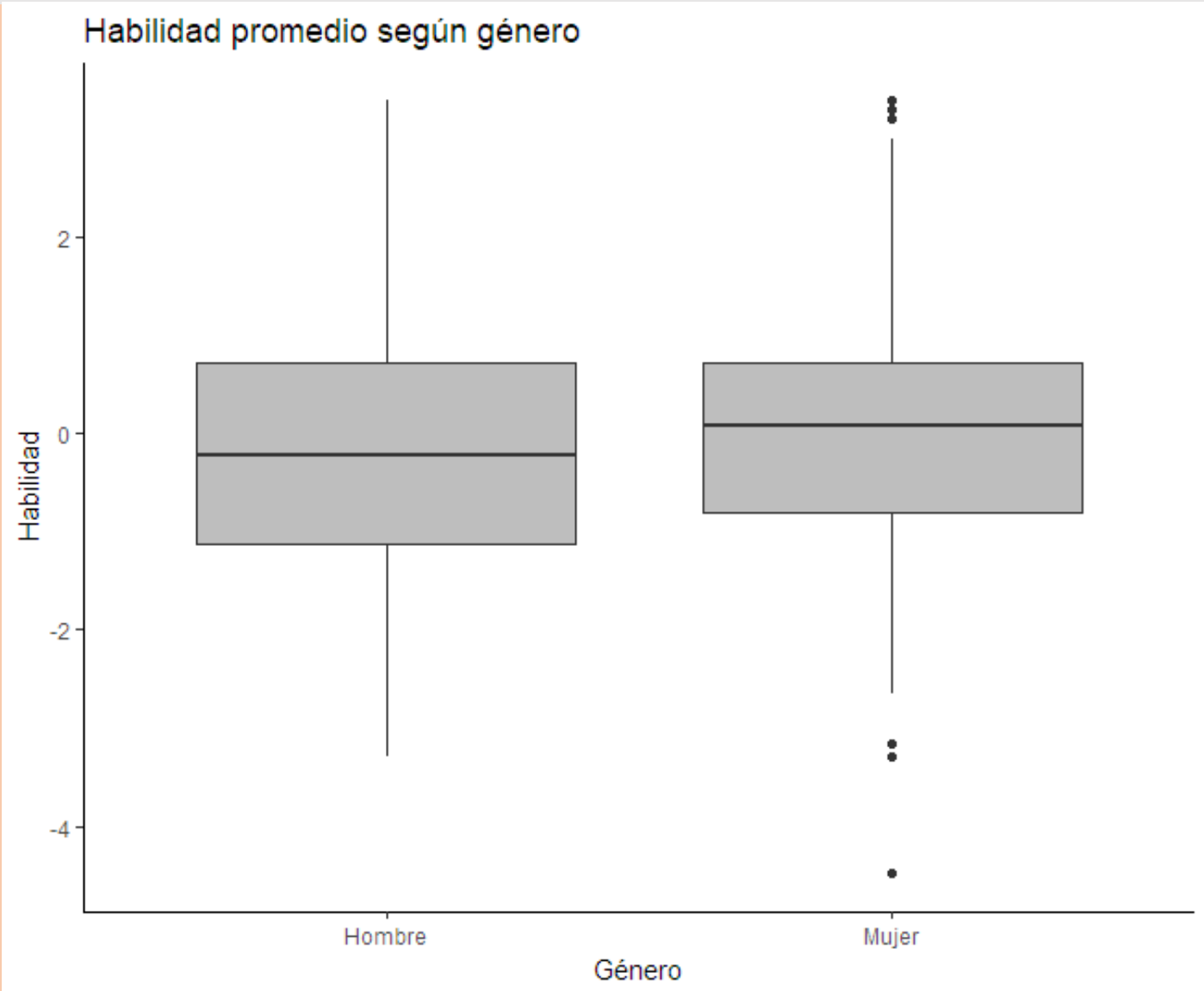
Los resultados del modelo logístico clasifica los ítems en 3 grupos: A cuando el DIF no es significativo, B cuando es moderado y C cuando es grande.

El análisis de facetas entrega coeficientes estandarizados, los cuales se centran en 0. Diferencias de 0.3 (positivas o negativas) se consideran como relevantes de reportar.

Resultados y conclusiones

La estimación de habilidad del instrumento posee una confiabilidad de 0.71, la cual es consistente con los resultados obtenidos en el modelamiento original del instrumento que incluye todos los países participantes de la región. Así mismo, la correlación entre la habilidad de los examinados y los valores plausibles de la prueba internacional es de gran magnitud (0.79) y bastante cercana a la reportada en el modelamiento original (0.84).

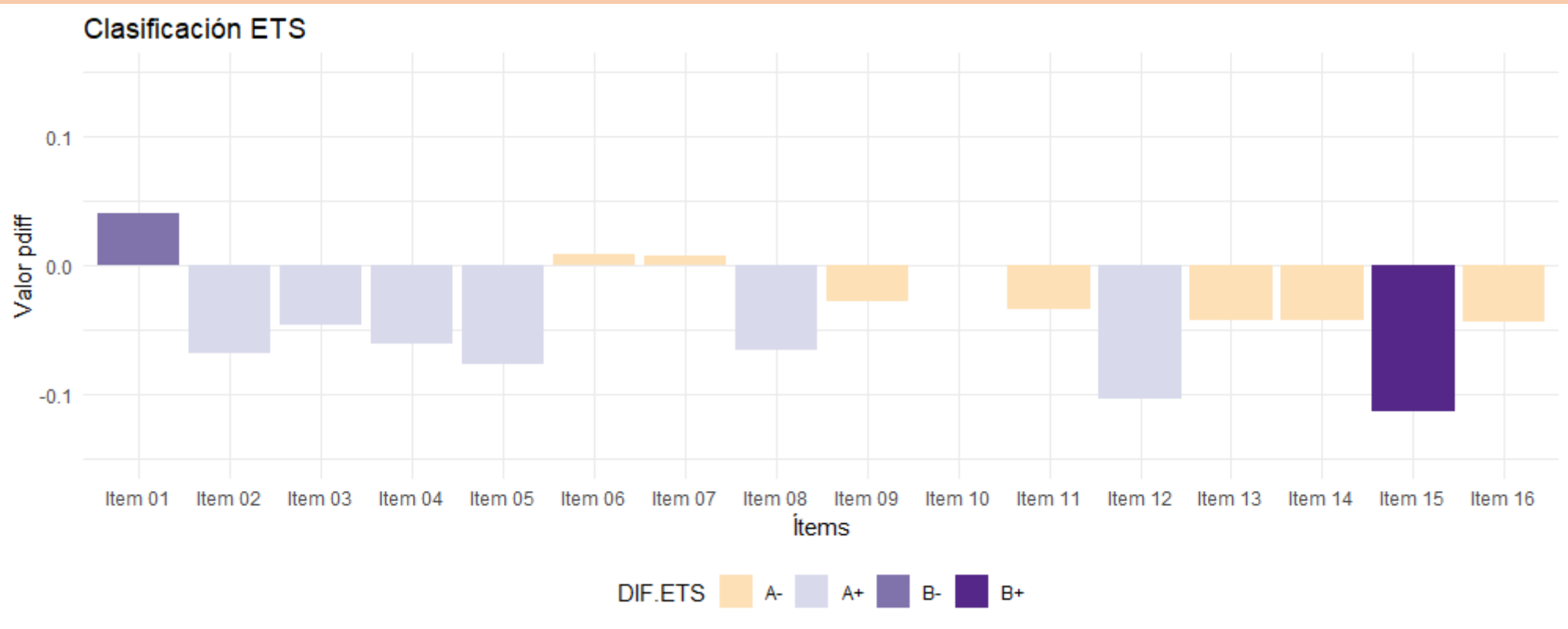
Gráfico 1: Habilidades y género



El grafico 1 muestra que las diferencias en las habilidades promedio de hombres y mujeres son estadísticamente significativas

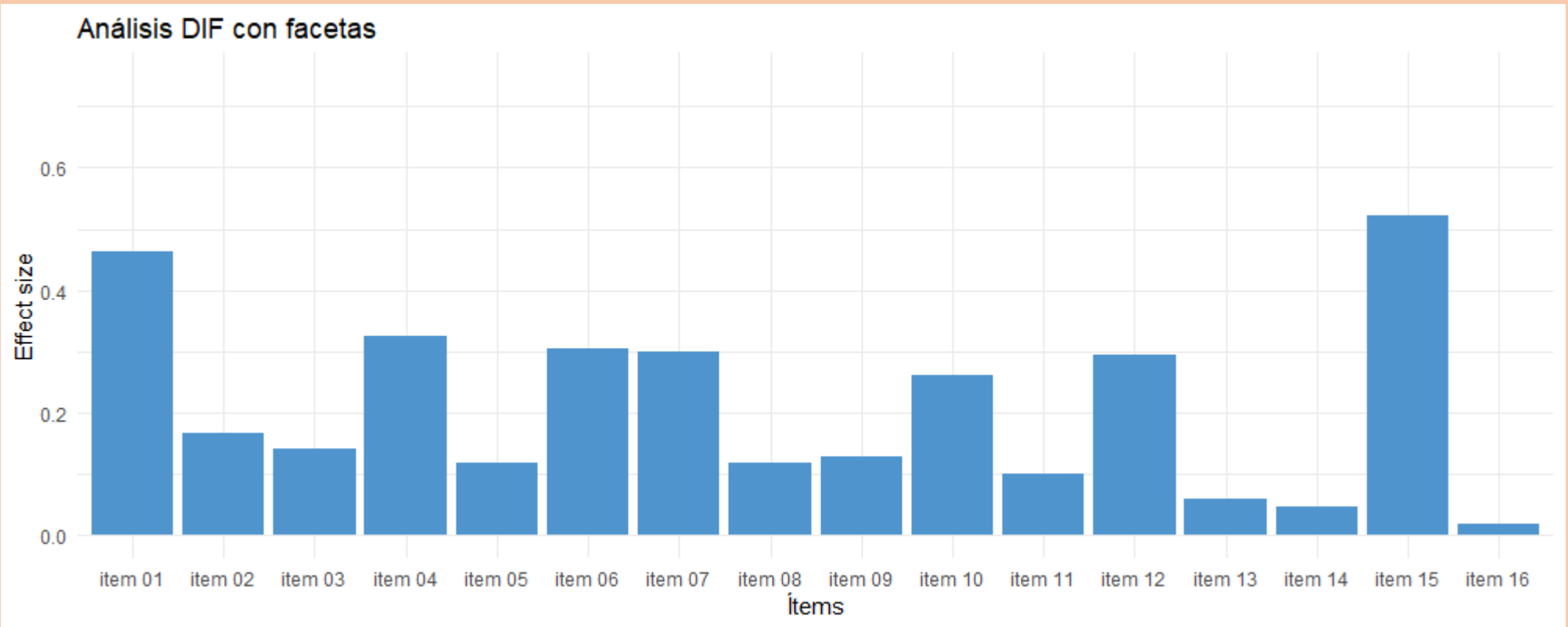
El gráfico 2 muestra los 16 ítems de la prueba de conocimiento cívico y sus categorías de DIF en base a la estimación logística. Aquí se observa que no existen ítems con alerta de DIF, lo que indica que, a partir de este método, no hay ítems que presenten sesgos atribuidos al género.

Gráfico 2: Clasificación de ítems a partir de criterios ETS



El gráfico 3 muestra los 16 ítems de la prueba de conocimiento cívico y sus tamaños de efecto estimados a partir del análisis con facetas. Este modelo incluye una interacción entre el ítem y el género, reportando el coeficiente estandarizado. Dado que este coeficiente está estandarizado y centrado en 0, se calcula el tamaño del efecto a partir de la multiplicación del valor absoluto por 2; diferencias mayores a 0.6 se consideran importantes para entregar una alerta de DIF.

Gráfico 3: Tamaño del efecto del DIF a partir de análisis con facetas



- A pesar de que son significativas las diferencias de habilidades no hay sesgo por género en los ítems
- Los resultados obtenidos a partir de ambos métodos son consistentes y no reflejan alerta de DIF en ninguno de los ítems que componen el instrumento.
- Dado que el análisis logístico es muestra dependiente, el análisis con facetas se considera un método que permite amortiguar el error de los ítems al comparar respuestas y no grupos.
- A partir de estos resultados es posible concluir que, para el caso de Chile, este instrumento es válido e imparcial, pues sus ítems no tienen sesgo asociado a la pertenencia a un grupo, que en este caso corresponde al género de los examinados.
- Finalmente, el análisis DIF debe ser considerado como un componente relevante en cualquier evaluación con consecuencias y que tengo como propósito reportar resultados.

mlmunoz1@uc.cl ; ccmiranda@uc.cl ;
ijrossi@uc.cl