

# Estimación del error de equiparación para las pruebas saber 3°, 5° y 9° entre los años 2009 y 2012.

Diana C. Gamboa  
Subdirección de Estadística  
Instituto Colombiano para la Evaluación de la Educación - ICFES  
*dgamboa@icfes.gov.co*

Noviembre 10, 2013

- 1 Presentación
- 2 Objetivo
- 3 Descripción de los datos
- 4 Estimación de constantes de equiparación
- 5 Resultados  
Discusión
- 6 Referencias

## Presentación

SABER 3°, 5° y 9° es una evaluación nacional de carácter externo presentada periódicamente por estudiantes de educación básica de todo el país con el fin de conocer el desarrollo de las competencias básicas en las áreas de lenguaje, matemáticas, ciencias naturales y competencias ciudadanas a través de los años.

Además, permite orientar los planes de mejoramiento de los establecimientos educativos y la toma de decisiones por parte de las instituciones del sector educativo.

## Equiparación

Es un conjunto de procedimientos psicométricos que permiten ajustar los puntajes de las formas de una prueba de modo que sus puntajes puedan ser intercambiables (Kolen & Bremman, 2004). El procedimiento empleado para esta equiparación en SABER 359 es el de transformación lineal. Este se define como:

$$\theta_{Ji} = A\theta_{Ii} + B$$

donde:  $\theta_{Ji}$  y  $\theta_{Ii}$  son valores de los individuos en las escalas  $I$  y  $J$ .

Los parámetros  $A$  y  $B$  se estiman como:

$$A = \frac{\sigma(\theta_J)}{\sigma(\theta_I)}$$

$$B = \mu(\theta_J) - A\mu(\theta_I)$$

## Objetivo

Estimar el error en la equiparación entre las aplicaciones 2009 y 2012 de las pruebas SABER para los grados 5<sup>o</sup> y 9<sup>o</sup> en las áreas de Ciencias naturales, Lenguaje y Matemáticas.



## Descripción de los datos

- La aplicación SABER 5° y 9° del año 2009 se compone de 8 bloques de preguntas, de los cuales cuatro fueron liberados durante la aplicación censal.
- La aplicación SABER 3°, 5° y 9° del año 2012 se compone de 6 bloques de preguntas, 4 de los cuales provienen de la aplicación controlada 2009.

HABILIDADES  
CIENCIA

## Estimación de constantes de equiparación

- 1 Se realiza una calibración conjunta de las aplicaciones 2009 y 2012 de forma independiente en cada grado y prueba.
- 2 Usando los parámetros generados a partir de la calibración conjunta se estima la habilidad de los estudiantes evaluados en la aplicación 2009.
- 3 Se busca la transformación lineal que permita enlazar los puntajes obtenidos bajo la calibración conjunta para la aplicación 2009 y los puntajes originales.
- 4 Se aplica esta transformación a los puntajes originales de la aplicación 2012.

## Estimación del error

Para la estimación del error se trabaja con los establecimientos educativos pertenecientes a la muestra controlada (unidades primarias de muestreo). Se utiliza la técnica de remuestreo *Jackknife*. El estimador de la varianza para cada constante de la transformación se define como:

$$v(\hat{A}) = \sum_{h=1}^L \frac{n_{h-1}}{n_h} \sum_{i=1}^{n_h} (\hat{A}_{(hi)} - \hat{A})^2$$

donde,

$n_h$  es el total de establecimientos en el estrato  $h$ ,  $\hat{A}$  corresponde a la estimación del parámetro de la población, y  $\hat{A}_{(hi)}$  a la estimación del parámetro al eliminar el establecimiento  $i$  del estrato  $h$ .

## Resultados

Pruebas	Ciencias		Lenguaje		Matemáticas	
	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )
Quinto	0.01452 (0.00019)	1.02888 (0.00028)	0.02585 (0.00016)	1.01425 (0.00020)	0.02552 (0.00016)	1.00797 (0.00017)
Noveno	0.06369 (0.00016)	1.02503 (0.00026)	0.05844 (0.00059)	1.01326 (0.00041)	0.03692 (0.00027)	1.00984 (0.00025)

## Sensibilidad de las estimaciones

Como método exploratorio se decide estimar las constantes de la transformación y sus respectivos errores muestrales desanclando aleatoriamente el 25% y 50% de los ítems comunes en las aplicaciones.

Pruebas	Ciencias		Lenguaje		Matemáticas	
	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )
Grado						
Quinto	0.01044 (0.00016)	1.02462 (0.00022)	0.02312 (0.00014)	1.00812 (0.00016)	0.04409 (0.00021)	1.00821 (0.00018)
Noveno	0.03553 (0.00027)	1.01274 (0.00028)	0.05004 (0.00051)	1.00994 (0.00037)	0.033339 (0.00022)	1.00743 (0.00021)

Tabla Desanclando el 50%

## Sensibilidad de las estimaciones

Pruebas	Ciencias		Lenguaje		Matemáticas	
	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )	Intercepto(B) ( $Ee_{mues}$ )	Pendiente(A) ( $Ee_{mues}$ )
Quinto	0.01201 (0.00019)	1.02782 (0.00026)	0.02772 (0.00017)	1.01428 (0.00021)	0.03718 (0.00020)	1.00776 (0.00018)
Noveno	0.04344 (0.00034)	1.01666 (0.00034)	0.05912 (0.00062)	1.01380 (0.00043)	0.04224 (0.00031)	1.01018 (0.00029)

Tabla Desanclando el 25%

## Discusión

- Se propuso una metodología para estimar la variabilidad de las constantes de equiparación entre las aplicaciones SABER 5° y 9° del año 2009 y del año 2012.
- El error fue estimado a partir de la técnica de re-muestreo jackknife utilizando las unidades primarias de muestreo (establecimientos).
- Se concluye que las estimaciones de la pendiente (A) y el intercepto (B) presentan una baja variabilidad.

## Referencias y bibliografía

- Cervantes V.H. & Lopera, C. (Eds.). (2011). Informe técnico SABER 5o. y 9o. Instituto Colombiano para la Evaluación de la Educación, ICFES. Bogotá: ICFES.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*. New York, NY: Springer.
- ETS (2010). DESI. Efficacy of effect size measures in logistic. *Direct estimation software interactive v3.2.7*. (Software de cómputo). ETS Educational Testing Service, Princeton, NJ.

## Referencias y bibliografía

- Martin, M., Mullis, I. V. S., Foy, P., Brossman, B, & Stanco, G. M. (2012) Estimating linking error in PIRLS. . *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5, 35–47.
- Von Davier, M., Gonzalez, E., & Mislevy, R.J. (2009). What are plausible values and why are they useful?. ?. *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments*, 2, 9-36.
- Wolter, K. M. (2007). *Introduction to Variance Estimation*. 2nd Edition. Springer, Inc.