



La educación  
es de todos

Mineducación

# SABER

AL > DETALLE

EDICIÓN

07

Bogotá D.C.

Abril de 2020

ISSN: 2590 - 4663

Publicación Trimestral

Instituto Colombiano para la Evaluación de la Educación, ICFES

Oficinas: Calle 26 No. 69-76 . Torre 2, pisos 15 -18

Edificio Elemento, Bogotá . Colombia

Directora General: **Mónica Ospina Londoño**

Directora de Evaluación: **Natalia González Gómez**

Subdirectora de Análisis y Divulgación: **Natalia González Gómez (E)**

Subdirectora de Estadísticas: **Jeimy Paola Aristizábal Rodríguez**

Subdirección de Diseño de Instrumentos: **Javier Toro Baquero**

Coordinación General: **Dirección de Evaluación**

**¿CÓMO SE CALIFICAN  
LAS PREGUNTAS DE  
RESPUESTA ABIERTA EN  
EL CONTEXTO DE LAS  
PRUEBAS SABER?**

**icfes**   
mejor saber

## ¿CÓMO SE CALIFICAN LAS PREGUNTAS DE RESPUESTA ABIERTA EN EL CONTEXTO DE LAS PRUEBAS SABER?

En las pruebas Saber Pro y Saber TyT se contempla una metodología de calificación específica para los ítems de respuesta abierta. Este tipo de ítems busca medir habilidades relacionadas con el desarrollo de argumentos y conceptos, como es el caso de la pregunta de habilidades comunicativas o de la prueba específica de arquitectura. A diferencia de los ítems que habitualmente se aplican en el

Icfes, los ítems de pregunta abierta cuentan con evaluadores que utilizan rúbricas de evaluación previamente definidas para ser empleadas como insumo en el proceso de codificación. En esta edición, ahondaremos en la calificación de pregunta abierta, exploraremos qué tipo de ítems corresponden con esta modalidad y expondremos la metodología de calificación que se emplea.

### ¿A qué nos referimos con preguntas abiertas?



Las preguntas abiertas son aquellas en las que el evaluado elabora una respuesta frente a un tema propuesto y en relación con la exposición de un enunciado, que se conoce como estímulo. Por ejemplo: indicar, ¿cuál fue el procedimiento realizado para despejar una variable de una ecuación?, o ¿considera que se deben imponer sanciones a las empresas que contaminen más allá de un límite? Como se observa, este tipo de preguntas son diferentes a las preguntas de opción múltiple con única respuesta y pueden requerir de diversas competencias para su desarrollo.

Las preguntas abiertas han estado presentes en los últimos años en los exámenes estandarizados que aplica el Icfes. En 2014, se incluyeron 8 preguntas abiertas cortas en el examen de Saber 11 en las pruebas de Lectura Crítica, Sociales y Ciudadanas, Matemáticas, y Ciencias Naturales, con el fin de evaluar la capacidad de análisis y argumentación a partir de un contexto (Icfes, 2014). Actualmente, este tipo de preguntas se emplea

para la medición de competencias comunicativas para los exámenes Saber Pro y Saber TyT, y para el módulo específico de proyecto de arquitectura del examen Saber Pro. En esta edición ahondaremos sobre las preguntas abiertas enfocadas en competencias comunicativas.

FIGURA 1. Ejemplo de pregunta abierta

Desde el 1 de julio de 2017, los colombianos deben pagar un impuesto por cada bolsa plástica que usen en el supermercado. Con esta medida, el Gobierno busca desestimular el uso de los empaques plásticos en el país. Sus críticos afirman que esta medida es un negocio que quiere beneficiar a los fabricantes de bolsas.

¿Está de acuerdo con la afirmación de que cobrar impuestos a las bolsas de plástico es una medida que beneficiará a los fabricantes?

Como se observa en la **figura 1**, el módulo de competencias comunicativas busca que los evaluados escriban un texto argumentativo sustentando su posición frente a un tema particular sobre el cual no se requieren conocimientos especializados. El propósito del escrito se centra en la estructuración del argumento y no en la opinión *per se* del evaluado; en este sentido no se evalúa la respuesta sino la coherencia de la exposición de los argumentos en el texto. En tal caso, se contemplan tres aspectos, que consisten en la forma de expresión, el planteamiento expuesto a través

de una postura, y la organización del texto (Icfes, 2019). El primer aspecto está relacionado con la legibilidad del texto, de tal manera que la idea plasmada en el texto sea claramente comprensible a través del uso pertinente de recursos gramaticales y ortográficos; esto es muy importante, pues se evalúa lo que el estudiante escribe. El segundo aspecto se relaciona con una postura clara y definida sobre la tesis presentada en el texto. Finalmente, la organización del texto corresponde con ordenamiento natural de los textos argumentativos.

## ¿Cómo se evalúa?



El proceso de codificación de las preguntas de opción múltiple con única respuesta es de tipo binario, en el que la respuesta correcta se codifica con 1 y las respuestas incorrectas con 0. No obstante, la codificación de las respuestas de preguntas abiertas no sigue ese planteamiento, pues requiere la consideración de un elemento adicional en el proceso de codificación llamado evaluador. Los evaluadores son personas externas al proceso de diseño de ítems que conocen del tema y emplean pautas de evaluación definidas por el Icfes para realizar la codificación de pregunta abierta. Estos evaluadores se denominan codificadores, pues codifican la respuesta con una clasificación; por ejemplo, para la prueba de comunicación escrita cada codificador asigna una clasificación de 1 a 4 de acuerdo al cumplimiento de tareas específicas que se definen en la rúbrica. Se codifica en cuatro categorías ordenadas, de tal manera que los textos evaluados en categoría 1 cumplen menos requisitos que en 2, aquellos clasificados en 2 menos que en 3, y los codificados en 3 menos que en los requisitos

de la categoría 4. Para cada evaluado se cuenta con al menos dos evaluadores, con el fin de garantizar confiabilidad y objetividad en el proceso, y se requiere que las respuestas de los evaluados sean legibles, que el texto sea pertinente, y por tanto se dé respuesta sin ambigüedades al estímulo planteado.

Dado que se tienen dos evaluadores independientes, se realiza un seguimiento a la concordancia de las clasificaciones que tienen para un mismo evaluado. Por ejemplo, si los evaluados coinciden en que el texto de un evaluado presenta una clasificación de 3, su codificación corresponderá a esa categoría. Sin embargo, si esto no sucede, se apela a un tercer evaluador, pues la codificación debe reflejar objetividad de las pautas de evaluación. De tal manera que se asigna la codificación más concordante con el tercer evaluador, el cual es un miembro del equipo altamente calificado para realizar estos procesos de codificación.

## ¿Cuál es el modelo de calificación empleado?



Como se ha abordado en publicaciones pasadas, los modelos de calificación empleados en el Icfes buscan establecer una relación probabilística entre las características psicométricas de los ítems y su probabilidad de acierto. En particular, en los modelos de 3 parámetros de Teoría Respuesta al Ítem (TRI) se contemplan la dificultad, la discriminación y el azar. Por su parte, en la calificación de las preguntas abiertas se consideran elementos tales como la severidad del evaluador (codificador), la dificultad de la tarea y la clasificación del evaluado que resultan relevantes para el proceso de calificación. Estos elementos, que se conocen como facetas, son los factores del modelo de calificación de múltiples facetas de Rasch (MMFR), que se utilizan en el procesos de calificación de este tipo de preguntas y por tanto se deben identificar y medir (Bond & Fox, 2007).

La particularidad del MMFR consiste en estimar la habilidad del evaluado, teniendo en cuenta las facetas anteriormente mencionadas. De tal manera que el resultado obtenido por los evaluados aísla el efecto de qué tan severo o laxo es el evaluador y de la dificultad del ítem. Así, el MMFR es una generalización del modelo de Rasch<sup>1</sup>, pues considera el efecto de diferentes facetas para estimar la habilidad de los evaluados (Linacre, 1994).

Los parámetros del modelo de múltiples facetas de Rasch se muestran en la siguiente ecuación:

$$\log \left( \frac{P_{nijk}}{P_{nij(k-1)}} \right) = B_n - D_i - C_j - e_{nijk}$$

Donde,

$P_{nijk}$ : La probabilidad que el evaluado  $n$  obtenga una clasificación  $k$  en el ítem  $i$  por el evaluador  $j$

$P_{nij(k-1)}$ : La probabilidad que el evaluado  $n$  obtenga una clasificación  $k-1$  en el ítem  $i$  por el evaluador  $j$

$B_n$ : Es la habilidad del evaluado  $n$

$D_i$ : Es la dificultad del ítem  $i$

$C_j$ : Es la severidad del evaluador  $j$

$e_{nijk}$ : Es el error de medición del modelo

Brevemente, la ecuación anterior expone que la calificación de un evaluado está relacionada con la probabilidad relativa de ser clasificado entre dos niveles próximos, que depende de su habilidad, la dificultad del ítem y la severidad del evaluador. Teniendo en cuenta la naturaleza del modelo logístico, la escala de medición está en unidades de *logits*. Así, cada faceta se calibra<sup>2</sup> con las clasificaciones dadas por los evaluadores y se llevan a la misma escala en un mapa de facetas, como se presenta en la **figura 2**, de tal manera que es posible analizar todas las facetas en un solo esquema. Vamos a entenderlo.

De manera general se aprecian 5 columnas que, de izquierda a derecha, dan información de 1) la escala de medición, 2) la severidad de los evaluadores (codificadores), 3) la habilidad de los evaluados, 4) la dificultad de la tarea y 5) las clasificaciones de los evaluados dadas por los evaluadores. Para todos

1. El modelo de Rasch puede ser visto como un caso particular del modelo 1PL de TRI. Para profundizar sobre los modelos de calificación y sus supuestos refiérase a la Edición 1 del Boletín Saber al Detalle ¿Cómo se generan los puntajes en las pruebas Saber del Icfes?

2. Para profundizar sobre los modelos de calibración y sus supuestos refiérase a la Edición 1 del Boletín Saber al Detalle ¿Cómo se generan los puntajes en las pruebas Saber del Icfes? Ejemplo de pregunta abierta.

los casos, el asterisco corresponde a un conjunto de evaluados. En la primera columna se aprecia la escala de medición en unidades *logits* que permite tener una regla de medición común de los parámetros de las facetas definidas. Así, por ejemplo, la categoría 1 de la última columna corresponde con los valores bajos de la primera columna, señalados con signo negativo, y conforme aumenta el valor de la última columna, los valores de la escala de *logits* de la primera columna también se incrementan en magnitud.

En la segunda columna, se observa la severidad de los evaluadores, es decir qué tan estricto es el codificador a la hora de clasificar a un evaluado. En esta columna se muestra la distribución de la severidad medida en la escala *logits*, y en la columna siguiente se observa la distribución de la habilidad de los evaluados en la misma escala. En la columna 4 se observa la faceta de tarea donde se muestra numéricamente la cantidad de estímulos propuestos y, de acuerdo con su ubicación en la escala, se puede determinar el nivel de dificultad de cada uno. Finalmente, en la última columna se observa la codificación realizada por los evaluadores.

En el caso del ejemplo de la **figura 2**, se puede observar que la severidad de los codificadores es más alta que la habilidad de los evaluados, debido a que la distribución de la habilidad de los evaluados se encuentra desplazada hacia la parte inferior del eje. Adicionalmente, las dos tareas tienen el mismo nivel de dificultad, que obedece a los criterios de construcción de ítems. Finalmente, con el uso de la columna 5 se pueden establecer niveles de desempeño<sup>3</sup> de acuerdo con la habilidad estimada de los evaluados.

FIGURA 2. Mapa del modelo de múltiples facetas de Rasch

Measr	-Codificador	+Evaluados	-Tarea	RS1
16		.		(4)
15		.		
14	.	.		
13		.		
12	.	.		
11	.	.		
10	.	.		
9	.	.		
8	*.	*.		
7	*.	*.		---
6	****.	*.		
5	*****	**.		
4	*****.	**.		3
3	*****.	**.		
2	*****.	**.		
1	***.	*****.		---
0	**.	*****.	1 2	
-1	*.	*****.		
-2	.	**.		
-3	.	**.		
-4	.	***.		2
-5	.	****.		
-6		**.		
-7		*.		
-8		.		---
-9		.		
-10		.		
-11		.		
-12		.		
-13	.	.		
-14				
-15				
-16				
-17		.		
-18				
-19				
-20				
-22				
-23		.		(1)
Measr	* = 21	* = 1202	-Tarea	RS1

Fuente: Icfes, 2020

3. Recordemos que corresponde a una clasificación de 1 a 4 de acuerdo con el cumplimiento de tareas específicas que se definen en la rúbrica; estos son los requisitos que cumple los textos evaluados por los codificadores.

---

## Bibliografía

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Icfes. (2014). *Sistema Nacional de Evaluación Estandarizada de la Educación. Alineación del examen Saber 11°. Lineamientos generales 2014-2*. Obtenido de <https://www.icfes.gov.co/documents/20143/177687/Guia+lineamientos+generales+Saber+11+2014-2.pdf/1c306ebd-3885-2695-4670-d133dc86ffd9>

Icfes. (2019). *Saber TyT. Módulo de comunicación escrita. Ejemplos de preguntas explicados*. Obtenido de <https://www2.icfes.gov.co/documents/20143/1476660/Preguntas+explicadas+comunicacion+escrita+saber-tyt.pdf/471a1df4-1439-73de-eace-6c54d5540bb9>

Linacre, J. M. (1994). *Many-Facet Rasch Measurement*. Chicago: MESA Press.

---

# SABER AL > DETALLE

**¿CÓMO SE CALIFICAN  
LAS PREGUNTAS DE  
RESPUESTA ABIERTA EN  
EL CONTEXTO DE LAS  
PRUEBAS SABER?**



La educación  
es de todos

Mineducación



@icfescol



ICFES



icfescol



YouTube: ICFES

Instituto Colombiano para la Evaluación de la Educación, ICFES

Oficinas: Calle 26 No. 69-76 . Torre 2, pisos 15 -18

Edificio Elemento, Bogotá . Colombia