

**Metodología utilizada para
la detección de los casos
de indicios de copia en
SABER 5o. y 9o. 2009**

Presidente de la República

Juan Manuel Santos Calderón

Ministra de Educación Nacional

María Fernanda Campo Saavedra

Viceministro de Educación Preescolar, Básica y Media

Mauricio Perfetti del Corral



Directora General

Margarita Peña Borrero

Secretaria General

Gioconda Piña Elles

Jefe de la Oficina Asesora de Comunicaciones y Mercadeo

Ana María Uribe González

Director de Evaluación

Julián Patricio Mariño von Hildebrand

Director de Producción y Operaciones

Francisco Ernesto Reyes Jiménez

Director de Tecnología

Adolfo Serrano Martínez

Subdirectora de Diseño de Instrumentos

Flor Patricia Pedraza Daza

Subdirectora de Análisis y Divulgación

Maria Isabel Fernandes Cristóvão

Autores (en orden alfabético)

Víctor Hernando Cervantes Botero

Wilmer Osvaldo Martínez Rivera

Colaboración

Nelsy Rodríguez Malagón

Julián Patricio Mariño von Hildebrand

Diagramación

Giovanni Camacho Solorza

Bogotá, D. C., febrero de 2010

Advertencia

Con el fin de evitar la sobrecarga gráfica que supondría utilizar en español "o/a" para denotar uno u otro género, el ICFES opta por emplear el masculino genérico en el que todas las menciones de este se refieren siempre a hombres y mujeres.

ICFES. 2011. Todos los derechos de autor reservados ©.

Todo el contenido es propiedad exclusiva y reservada del ICFES y es el resultado de investigaciones y obras protegidas por la legislación nacional e internacional. No se autoriza su reproducción, utilización ni explotación a ningún tercero. Solo se autoriza su uso para fines exclusivamente académicos. Esta información no podrá ser alterada, modificada o enmendada.

Introducción

El ICFES está comprometido con el mejoramiento continuo de las técnicas de evaluación y medición. Con ello, los resultados de las diversas evaluaciones a su cargo tienen un alto grado de confiabilidad y pueden ser utilizados como insumo para la realización de autoevaluaciones institucionales y para la elaboración de planes de mejoramiento.

En SABER 2009 se utilizó una rigurosa metodología estadística para establecer en qué casos se presentaron indicios de copia en las pruebas. Con ello, fue posible estimar con mayor precisión los resultados de los establecimientos educativos, las entidades territoriales y los diversos grupos de referencia – sector (oficial – privado), zona (urbana – rural) y nivel socioeconómico. El presente documento contiene la descripción de dicha metodología. Además, se incluye un ejemplo de los cálculos realizados, a fin de ilustrar su aplicación.

CONSIDERACIONES INICIALES

SABER 2009 se aplicó a estudiantes de quinto y noveno grados en dos momentos: el primero en el mes de mayo (aplicación M), para quienes asistían a establecimientos educativos de calendario B, y el segundo en octubre (aplicación O), para los alumnos matriculados en instituciones de calendario A. En ambas aplicaciones se llevaron a cabo tres tipos de operativos: censal (S), controlado (C) y estadístico (E).

El operativo censal (S) abarcó la gran mayoría de estudiantes de ambos grados. En el mismo, la coordinación de la aplicación de las pruebas estuvo a cargo del rector (o su delegado) de la institución educativa. Para ello, recogió los materiales de las pruebas (cuadernillos y hojas de respuesta) en un punto de entrega establecido por el ICFES, realizó la aplicación y devolvió las hojas de respuesta en ese mismo punto de entrega. A su vez, en el operativo controlado (C) el personal del ICFES se desplazó directamente a cada una de las sedes de las instituciones educativas seleccionadas, y se encargó de llevar a cabo la aplicación de las pruebas. Finalmente, en el operativo estadístico (E) el ICFES envió un delegado a cada sede y, con el apoyo del personal de la institución educativa, efectuó la aplicación de las pruebas.

En SABER 2009 se evaluaron tres áreas: lenguaje, matemáticas y ciencias naturales. Los cuadernillos de pruebas utilizados en ambos grados fueron diseñados de tal manera que cada uno contenía las preguntas de dos de las tres áreas mencionadas¹. En consecuencia, cada estudiante fue evaluado en dos áreas, pero la aplicación de SABER a nivel de sede cobijó las tres áreas.

El procedimiento de detección de copia empleado en el análisis de los datos de las pruebas aplicadas en SABER 2009 consistió en el cálculo de la probabilidad de encontrar un patrón de respuestas con cierta similaridad al dado por cada estudiante evaluado en la misma sede-jornada, para cada área. Para evaluar esta probabilidad se propuso la metodología que se describe más adelante, teniendo como base lo señalado por Linacre (1992)² en el sentido de que “el criterio está dado por una gran cantidad de respuestas compartidas (correctas o incorrectas) (§5)”³ y que “la prueba debe considerar [...] qué tan aceptables son las similitudes a la luz de la distribución empírica de las respuestas (§4)”. Sin embargo, no se consideró factible la aplicación en su totalidad del procedimiento planteado por este autor, dado que requería obtener todas las comparaciones dos a dos entre estudiantes evaluados para determinar el número de respuestas compartidas, lo que en el caso de los evaluados en SABER 2009 produciría un número de comparaciones demasiado grande (aproximadamente $1.400.000^2 =$

¹ Como resultado, se obtuvieron tres formatos de cuadernillos, de acuerdo con la combinación de dos de las tres áreas en cada uno, a saber: lenguaje – matemáticas; lenguaje – ciencias naturales y matemáticas – ciencias naturales. Esta combinación se utilizó tanto para quinto como para noveno grado.

² Linacre, J. (1992). *Catching Copiers: Cheating Detection*. Rasch Measurement Transactions, 6(1), p. 201. Recuperado de <http://www.rasch.org/rmt/rmt61d.htm>, el 01/08/2009

³ Traducción de los autores.

1.960.000.000.000) para poder ser realizado en un tiempo razonable dentro del procedimiento de análisis y calificación de las pruebas. Por lo anterior, se consideró una aproximación a la distribución empírica y el uso de la distribución binomial de forma similar a la planteada por Holland (1996) y Sotaridona y Meijer (2002)⁴, pero a diferencia del indicador empleado por estos autores, se usó el patrón de respuestas propiamente dicho, en lugar de utilizar solamente el patrón respuestas incorrectas.

METODOLOGÍA PARA DETECCIÓN DE COPIA

Se definió la similitud entre el patrón de respuestas de dos estudiantes evaluados por $d =$ “el número de respuestas diferentes entre los dos evaluados”, de modo que entre menor es el valor, más se parecen los dos patrones; por ejemplo, un valor de cero quiere decir que ambos patrones de respuesta son exactamente iguales en cuanto a la opción elegida por cada evaluado, sin importar si esas respuestas son correctas o incorrectas. Si consideramos la distribución empírica del número de evaluados que presenten valores de d tan grandes como los encontrados entre un par de evaluados en una muestra obtenida del conjunto completo, es posible ver que se trata de la distribución de la cantidad de evaluados que cumplen cierta característica, en una muestra de tamaño (n) obtenida de una población finita de tamaño (N), donde existe una cierta cantidad de evaluados con dicha característica (D). Esta distribución corresponde a un proceso hipergeométrico con parámetros (n, D, N)⁵ cuando la muestra es elegida de forma aleatoria.

Puesto que obtener la cantidad D que define la distribución del número de evaluados considerada implica obtener todas las distancias dos a dos, se decidió aproximar la distribución hipergeométrica mediante la distribución binomial de parámetros n y $p = D/N$ ⁶, y estimar p (\hat{p}) de una muestra aleatoria de M evaluados, obteniendo la distancia a todos los patrones de respuesta incluidos en la muestra, y calculando la razón entre la cantidad de evaluados con una $d \leq d_{min} + 2$, donde d_{min} es la menor distancia en la que se encuentra algún evaluado en el sitio para un patrón específico de respuestas. Así, si en un sitio el patrón de respuestas compuesto por elegir A en todas las preguntas se encontró k veces, se considera sospechoso si de acuerdo con la distribución binomial de parámetros n y \hat{p} , la probabilidad de encontrar $k+1$ o más evaluados a una $d=0$, es menor a 0,000001.

Dadas las características de SABER 2009, se definieron sitios (unidades de análisis de copia) en los cuales fuese coherente analizar copia. Para ello, como la unidad de aplicación fue la sede-jornada y en ella se dispuso de varios formatos, el sitio está

⁴ (1) Holland, (1996). *Assessing unusual agreement between the incorrect answers of two examinees using the K-index*. Research report 96-7. Educational Testing Service. (2) Sotaridona, L. & Meijer, R. (2002). “Statistical properties of the K-index for detecting answer copying”. *Journal of Educational Measurement*, 39(2), 115-132.

⁵ Blanco L., (1996). Probabilidad. Universidad Nacional de Colombia, Facultad de Ciencias sede Bogotá.

⁶ Blanco L., (1996) Op.cit., p. 110.

definido como sede–jornada área–formato; por ende para los individuos en cada sitio se hicieron comparaciones dos a dos entre sus patrones de respuesta, garantizando así que los estudiantes comparados están respondiendo la misma prueba con las mismas preguntas. Luego, para cada estudiante en un sitio se calcularon las distancias a las que se encontraban los patrones de respuestas de los demás. Para cada individuo se calculó cuántos individuos se encontraban a 0, 1, ..., 11 diferencias de su patrón de respuestas. Por cada diferencia se construyó una variable, la cual dice cuántos individuos se encuentran a dicha distancia. Ahora, como la cantidad de preguntas en cada área–formato no fue la misma, se estableció el siguiente criterio: si la suma de las primeras k distancias es mayor a cero, donde k es el entero siguiente al 15% del número total de preguntas en cada área–formato, entonces ese estudiante es denominado sospechoso de copia. Por ejemplo, si en un sitio la cantidad de preguntas fue 54, el 15% es 8,1, siendo 9 el entero siguiente, luego si para un estudiante la cantidad $d_0 + d_1 + \dots + d_9$ es mayor a cero individuos este estudiante es sospechoso de copia, es decir, si la suma es 1, en el sitio existen al menos dos personas que presentan en su patrón de respuestas a lo más 9 diferencias.

Puesto que el objetivo era calcular la probabilidad de encontrar un patrón de respuestas dado en la población y así construir un índice de copia, en este caso el patrón de respuestas de los sospechosos, se utilizó un modelo binomial siguiendo el siguiente procedimiento: a partir del conjunto de sospechosos que presentaron la prueba del área c ($c = 1, 2, 3$ donde 1 es ciencias, 2 es lenguaje y 3 matemáticas) del formato i ($i = 1, \dots, 114$), se tomó una muestra aleatoria, de la población que presentó el área c del formato i sin los sospechosos de copia, de tamaño n_i , contra los que se calculó nuevamente las d_k , $k=0, 1, \dots, 11$ de cada uno de los individuos sospechosos. Con estas distancias y con las encontradas en cada sitio se calculó la probabilidad P_s dada por:

$$P_s = 1 - Pba(x_a, p_m, n_a)$$

Donde Pba es la probabilidad binomial acumulada, x_a corresponde al número de individuos que se encuentran a una distancia $d_{\min_a+2} + 1$, (d_{\min_a+2} es la menor distancia en la que se encuentra algún evaluado más los individuos a dos distancias adelante, en el sitio a , ($\min_a = 0, 1, \dots, 11$); p_m es el cociente entre las distancias acumuladas hasta $d_{\min_b+2} + 1$, pero ahora dentro de la muestra, y el tamaño de muestra menos los individuos (no sospechosos) del sitio que pertenezcan a la muestra; es decir, $p_m = \#indv(d_{\min_b+2} + 1) / (m - \#indv_a)$ con $\#indv(d_{\min_b+2} + 1)$ corresponde a la cantidad de estudiantes que se encuentran a la distancia acumulada $d_{\min_b+2} + 1$ en la muestra b , m . El tamaño de muestra y $\#indv_a$ es el número de individuos no sospechosos del sitio a que pertenecen a la muestra; n_a , es la cantidad de individuos en el sitio a –ésimo. Así, el índice de copia por persona está dado por:

$$I_s = \text{redondear} \left(\log_{10} \left(\frac{1}{P_s} \right) \right),$$

donde I_s se aproxima (redondea) al entero más cercano; además, se tiene que $0 \leq I_s \leq 14$.

En los casos en que P_s es nula el índice I_s es igualado a 14, indicando que la probabilidad de encontrar dicho patrón de respuestas en la muestra es cero. Puesto que, como ya se mencionó, el procedimiento no solamente consideró respuestas incorrectas sino también aquellas correctas, se hizo un ajuste por la proporción de respuestas correctas, esto es:

$$C_s = I_s - 14 \left(\frac{rc}{tr} \right)^3,$$

siendo C_s el índice de copia ajustado, rc la cantidad de respuestas correctas del individuo s y tr el número total de respuestas por sitio, luego si $C_s \geq 6$, se considera que el individuo s tiene indicio de copia.

Finalmente, la sede-jornada que presentó el área c del formato i (sitio a) es denominado como sitio con copia masiva en el área c si más del 40% de los estudiantes tiene indicios de copia.

Este procedimiento fue llevado a cabo de manera independiente para cada una de las aplicaciones.

Es importante tener en cuenta que los individuos que en su patrón de respuestas presentaron más de 5 omisiones o multimarca no fueron considerados en el análisis.

Se considera que hubo copia masiva en una sede-jornada para un área determinada cuando más de 40% de los estudiantes presentaron indicios de copia. Cuando se encontró copia masiva en un área y grado, no se calcularon ni se publicaron los resultados de esa sede-jornada para dicha área y grado. Además, si se encuentra que hubo copia masiva en dos o más áreas de un determinado grado, no se estimaron ni se publicaron los resultados correspondientes a ese grado.

EJEMPLO DE APLICACIÓN DE LA METODOLOGÍA

Con el fin de ilustrar el procedimiento descrito en el apartado anterior se tomaron 40 estudiantes de una sede-jornada cualquiera, en la que se aplicaron tres formatos (formatos a, b y c). Luego, se tienen 6 sitios en esta sede-jornada, dado que cada formato está compuesto por dos áreas. Por sitio se calcularon las distancias dos a dos y se encontró que 24 individuos son sospechosos de copia. Para cada uno de los 6 sitios se tomaron muestras aleatorias y se calcularon las distancias de los sospechosos contra las muestras, respectivamente. De esa manera, se determinaron las distancias mínimas a las cuales se encontraron individuos (véase **Tabla 1**). A partir de estos valores y

teniendo en cuenta los tamaños de los sitios sospechosos y de las muestras se calculó el valor P_s , y los índices I_s y C_s . Luego, para este ejemplo se tiene que en matemáticas más del 40% de los estudiantes tienen indicio de copia; por lo tanto, esta sede-jornada presenta indicios de copia masiva en el área de matemáticas (véase **Tabla 2**).

Tabla 1. Ejemplo de resultados por estudiante

Estudiante	sitio	na	nmta	dmin_a2_1	dmin_b2_1	codpba	formato	Ps	Is	rc	tr	Cs	indic_copia
5	1	7	6338	5	11	cien	a	0	14	46	54	5,345933039	0
7	1	7	6338	2	0	cien	a	0	14	41	54	7,872288269	1
2	2	6	6333	2	0	cien	c	0	14	43	54	6,931095361	1
8	2	6	6333	2	0	cien	c	0	14	42	54	7,412894376	1
11	2	6	6333	3	0	cien	c	0	14	43	54	6,931095361	1
2	5	6	6332	2	2	leng	b	6,29779E-10	9	45	54	0,898148148	0
8	5	6	6332	2	1	leng	b	7,87508E-11	10	43	54	2,931095361	0
3	6	7	6339	3	0	leng	c	0	14	43	54	6,931095361	1
6	6	7	6339	4	10	leng	c	2,05058E-13	12	48	54	2,167352538	0
9	6	7	6339	2	2	leng	c	1,09821E-09	8	44	54	0,426357771	0
12	6	7	6339	2	1	leng	c	1,37341E-10	9	43	54	1,931095361	0
1	3	7	6336	5	0	mate	a	0	14	41	54	7,872288269	1
4	3	7	6336	4	0	mate	a	0	14	42	54	7,412894376	1
5	3	7	6336	2	0	mate	a	0	14	48	54	4,167352538	0
7	3	7	6336	3	0	mate	a	0	14	40	54	8,309810496	1
10	3	7	6336	3	0	mate	a	0	14	43	54	6,931095361	1
13	3	7	6336	2	0	mate	a	0	14	47	54	4,769178987	0
15	3	7	6336	6	0	mate	a	0	14	45	54	5,898148148	0
1	4	7	6332	5	10	mate	c	0	14	44	54	6,426357771	1
3	4	7	6332	5	2	mate	c	0	14	44	54	6,426357771	1
6	4	7	6332	5	0	mate	c	0	14	44	54	6,426357771	1
9	4	7	6332	6	15	mate	c	0	14	48	54	4,167352538	0
12	4	7	6332	4	0	mate	c	0	14	43	54	6,931095361	1
14	4	7	6332	5	3	mate	c	0	14	46	54	5,345933039	0

En la Tabla 1 se presentan los resultados del ejemplo tomado para ilustrar la aplicación de la metodología. Cada columna está identificada como sigue:

- **Estudiante:** código del estudiante
- **Sitio:** código del sitio
- **na:** tamaño del sitio
- **nmta:** tamaño de la muestra restado la cantidad de individuos de cada sitio no sospechosos que pertenecen a la muestra
- **dmin_a2_1:** es la menor distancia en la que se encuentra algún evaluado más los individuos a dos distancias adelante + 1 dentro del sitio
- **dmin_b2_1:** es la menor distancia en la que se encuentra algún evaluado más los individuos a dos distancias adelante + 1 dentro de la muestra
- **codpba:** nombre de la prueba
- **formato:** código del formato
- **indic_copia = 1** si el estudiante tiene indicio de copia y 0 en caso contrario
- P_s, I_s, rc, tr y C_s corresponden a lo expuesto en las expresiones anteriores.

En resumen, por cada área se tiene lo siguiente:

Tabla 2. Resultados ejemplo por área en la sede - jornada

	Cantidad estudiantes con indicio de copia	Total estudiantes evaluados	% copia
Ciencias	4	13	30,8
Matemáticas	8	14	57,1
Lenguaje	1	13	7,7



Calle 17 No. 3-40 • Teléfono:(57-1)338 7338 • Fax:(57-1)283 6778 • Bogotá - Colombia
www.icfes.gov.co

