



EVALUACIÓN SABER – 2004 MATEMÁTICA Y LENGUAJE ESTUDIO DE CONTRASTE

INFORME FINAL

JULIO DE 2005

Director General del ICFES
Daniel Bogoya Maldonado

Secretario General del ICFES
Genisberto López Conde

EVALUACIÓN SABER – 2004 MATEMÁTICA Y LENGUAJE ESTUDIO DE CONTRASTE

Autores

Carlos Pardo Adames
Martha Jeaneth Castillo Ballén
Cristina Hennig Manzuoli
Jackeline Salamanca Durán
Grace Judith Vesga Bravo

Corrección de estilo
Martha Cecilia Rocha Gaona

Grupo Evaluación de la Educación Básica y Media
Subdirección Académica - ICFES

Diseño, impresión y terminados:
Secretaría General, Grupo de Procesos Editoriales - ICFES

Impreso en Colombia

Presidente de la República
ÁLVARO URIBE VÉLEZ

Vicepresidente de la República
FRANCISCO SANTOS CALDERÓN

Ministra de Educación Nacional
CECILIA MARÍA VÉLEZ WHITE

INSTITUTO COLOMBIANO PARA EL FOMENTO
DE LA EDUCACIÓN SUPERIOR



Director General
JOSÉ DANIEL BOGOYA MALDONADO

Secretario General
GENISBERTO LÓPEZ CONDE

Subdirector de Logística
FRANCISCO ERNESTO REYES JIMÉNEZ

Subdirector de Fomento
LUIS ALFREDO POSADA DELGADO

Oficina Asesora de Planeación
ÁLVARO DÍAZ NIÑO

Oficina Asesora Jurídica
WILLIAM ALBERTO CABALLERO RESTREPO

Oficina de Control Interno
LUIS ALBERTO CAMELO CRISTANCHO

TABLA DE CONTENIDOS

1.	INTRODUCCIÓN	8
2.	SABER 2002–2003	9
3.	LAS SEÑALES DE POSIBLE COPIA	10
4.	DISEÑO METODOLÓGICO DE LA EVALUACIÓN SABER 2004 - CONTRASTE	11
5.	APLICACIÓN DE SABER 2004 – CONTRASTE	14
6.	DEPURACIÓN DE LOS DATOS	14
7.	PROCESAMIENTO DE DATOS	18
7.1	ANÁLISIS DE ÍTEMS Y PRUEBAS	18
7.1.1	ANÁLISIS CUALITATIVO	18
7.1.2	ANÁLISIS CUANTITATIVO	19
7.2	LA CALIFICACIÓN	26
8.	CRITERIOS DE DECISIÓN	27
9.	RESULTADOS	28
9.1	DESEMPEÑO PROMEDIO POR INSTITUCIÓN	28
9.2	DESVIACIÓN ESTÁNDAR POR INSTITUCIÓN	32

TABLAS Y GRÁFICOS

Tabla 1. Número de instituciones y estudiantes participantes en la evaluación SABER 2004 – Contraste, por departamento.	13
Tabla 2. Número de instituciones cuyos datos se eliminaron de la base de datos por superar el índice de similitud en la información.	15

Tabla 3. Número de instituciones y estudiantes participantes en la evaluación SABER 2004 – Contraste en <u>lenguaje</u> , por departamento, después de la depuración por similitud en la información.	16
Tabla 4. Número de instituciones y estudiantes participantes en la evaluación SABER 2004 – Contraste en <u>matemática</u> , por departamento, después de la depuración por similitud en la información.	17
Tabla 5. Información estadística de las preguntas de la prueba de <u>matemática en grado quinto</u> de SABER 2004 – Contraste.	21
Tabla 6. Información estadística de las preguntas de la prueba de <u>matemática en grado noveno</u> de SABER 2004 - Contraste.	22
Tabla 7. Información estadística de las preguntas de la prueba de <u>lenguaje en grado quinto</u> de SABER 2004 - Contraste	23
Tabla 8. Información estadística de las preguntas de la prueba de <u>lenguaje en grado noveno</u> de SABER 2004 - Contraste	24
Tabla 9. Criterio de inconsistencia por prueba y grado SABER 2004 – Contraste	27
Tabla 10. Número de instituciones con disminución en el desempeño promedio de sus estudiantes superior al criterio de inconsistencia.	28
Tabla 11. Número de instituciones con aumento en el desempeño promedio de sus estudiantes superior al criterio de inconsistencia.	28
Gráfico 1. Relación entre el desempeño promedio, de instituciones con resultados consistentes, en SABER 2002–2003 y SABER 2004 – Contraste, en <u>matemática de quinto grado</u> .	30
Gráfico 2. Relación entre el desempeño promedio, de instituciones con resultados consistentes, en SABER 2002–2003 y SABER 2004 – Contraste, en <u>matemática de noveno grado</u> .	30
Gráfico 3. Relación entre el desempeño promedio, de instituciones con resultados consistentes, en SABER 2002–2003 y SABER 2004 – Contraste, en <u>lenguaje de quinto grado</u> .	31
Gráfico 4. Relación entre el desempeño promedio, de instituciones con resultados, consistentes en SABER 2002–2003 y SABER 2004 – Contraste, en <u>lenguaje de noveno grado</u> .	31

Gráfico 5. Relación de la desviación estándar de instituciones con resultados consistentes, en <u>matemática de quinto grado</u> , en SABER 2002–2003 y SABER 2004 – Contraste.	33
Gráfico 6. Relación de la desviación estándar de instituciones con resultados consistentes, en <u>matemática de noveno grado</u> , en SABER 2002–2003 y SABER 2004 – Contraste.	33
Gráfico 7. Relación de la desviación estándar de instituciones con resultados consistentes, en <u>lenguaje de quinto grado</u> , en SABER 2002–2003 y SABER 2004 – Contraste.	34
Gráfico 8. Relación de la desviación estándar de instituciones con resultados consistentes, en <u>lenguaje de noveno grado</u> , en SABER 2002–2003 y SABER 2004 – Contraste.	34

1. INTRODUCCIÓN

La evaluación educativa cobra sentido cuando se aprovecha para comprender el proceso de formación, para determinar sus logros y sus metas por cumplir, con el propósito de orientar efectivamente acciones de cualificación que tendrán como beneficiarios a los estudiantes. Vista de esta forma, la evaluación es un proceso en el cual vale la pena participar de manera transparente porque sus resultados reflejarán la realidad del objeto de evaluación y así, se formularán acciones de cualificación pertinentes. Sin embargo, es frecuente que a la evaluación se le vea como acción fiscalizadora asociada a distintos tipos de consecuencias positivas o negativas y por ello el solo anuncio de su realización genera animadversión, tensión o deseos de un buen resultado, llegando incluso a desencadenar conductas alejadas de principios éticos, tendientes a aparecer mejores de lo que en realidad son.

La presión que pueden llegar a sentir aquellos que son evaluados directamente, así como quienes lo son indirectamente por los resultados de los primeros, puede conducir a unos y a otros a intentar mostrar, por diversos métodos, una apariencia que es lejana de la realidad de lo evaluado, para lo cual pueden valerse de diversos métodos. Queda la preocupación de que si en el proceso de evaluación la institución ha empañado o distorsionado su realidad, los resultados no son reflejo confiable de su calidad y, por ende, el ejercicio evaluativo pierde sentido y se desperdicia la oportunidad de reconocer aspectos fuertes y débiles en los que la institución podría recibir orientación para mejorar. Quizás la única forma de combatir este tipo de situaciones es la demostración de que la evaluación se piensa y se usa en el sentido constructivo.

La institución no es la única que requiere resultados precisos de un proceso de evaluación. Tanto el país como las entidades territoriales formulan sus planes y políticas educativas con base en diversa información entre la que se destaca los resultados de la evaluación de la calidad de la educación. Si la información no es confiable, las decisiones que se tomen pueden no ser coherentes con las verdaderas necesidades del sector educativo.

La evaluación SABER 2002–2003, realizada a estudiantes de quinto y noveno grados de educación básica en Colombia, pretendió informar a

distintos usuarios (docentes, institución, región, país y población en general) acerca de la calidad de la educación que se imparte en las diferentes instituciones del país.

En octubre de 2004 se realizó la evaluación SABER 2004 – Contraste, cuyo fin principal fue verificar los resultados de algunas instituciones al identificar casos en los que el desempeño mostrado por los estudiantes en la evaluación SABER 2002–2003 no correspondía con el desempeño observado hasta entonces en algunas entidades territoriales.

2. SABER 2002–2003

La evaluación de lenguaje y matemática realizada por el ICFES y el MEN¹ en 2002 y 2003, es considerada como la primera de carácter censal realizada en el país en educación básica. Con ella se buscó incorporar, cada vez más, a las regiones y a las instituciones en los procesos de evaluación nacional y contribuir, de esa forma, a la construcción de una cultura evaluativa en el país que sensibilizara al sector educativo frente a los beneficios de la evaluación educativa.

La aplicación de las pruebas implicó un trabajo conjunto entre secretarías de educación departamentales y de municipios certificados y el ICFES, de acuerdo con el siguiente procedimiento:

1. Coordinación con las secretarías de educación departamentales y de municipios certificados para la recepción, distribución de material y la organización de la aplicación misma, con apoyo de los jefes de núcleo.
2. Envío del material de examen (pruebas, hojas de respuesta, manuales de aplicación) a las secretarías de educación.
3. Capacitación a los jefes de núcleo para la realización de la aplicación de las pruebas.

¹ El programa Nuevo Sistema Escolar del MEN aplicó la evaluación SABER a algunas instituciones en el País.

4. Entrega del material, por parte de las secretarías de educación, a los jefes de núcleo.
5. Distribución de las pruebas (el día del examen) a directivos docentes y docentes de las distintas instituciones. Las secretarías de educación, con el apoyo del ICFES, resolvieron las problemáticas que se presentaron durante la aplicación.
6. Los estudiantes respondieron las preguntas del examen bajo la supervisión de docentes de las instituciones educativas participantes.
7. Recolección de las hojas de respuesta e informes de aplicación y transporte de este material al ICFES, por medio de una empresa de seguridad.

A partir de este momento se inició el trabajo de obtención y análisis de resultados.

Los resultados alcanzados por algunas instituciones y departamentos desencadenaron hipótesis sobre posibles situaciones irregulares durante la aplicación de los exámenes. Es posible que, partiendo de una equivocada interpretación del propósito de la evaluación, se manipulara para mostrar un desempeño que no era real.

3. LAS SEÑALES DE POSIBLE COPIA

Como se señaló en la sección anterior, la evaluación SABER 2002–2003 de matemática y lenguaje es la primera evaluación censal en educación básica en Colombia, lo que quiere decir que no se cuenta con un registro previo del desempeño de estudiantes para la totalidad de instituciones del país. Es decir, que no se tenían resultados institucionales que hubieran permitido un contraste inmediato entre el desempeño en una evaluación anterior y el obtenido en SABER 2002–2003. No obstante, el país cuenta con dos evaluaciones educativas que permitieron tener información inmediata de contraste, a nivel departamental, para un número significativo de instituciones:

1. Examen de Estado para Ingreso a la Educación Superior. Esta evaluación es respondida por todos los estudiantes que se encuentran

en grado 11 en instituciones de educación media. Es un programa de evaluación que tiene sus orígenes en 1968 y que es de carácter censal desde 1980. Se obtienen resultados por institución participante y por departamento. Al comparar los resultados de SABER 2002–2003, en el contexto departamental con los de este programa, se encontró que algunos departamentos que no presentan buenos resultados en Exámenes de Estado, presentaban buenos resultados en la evaluación SABER, como es el caso de Córdoba, Guainía, La Guajira, Amazonas y Chocó.

2. Evaluación SABER 1997–1999. Esta evaluación fue muestral y representativa a nivel departamental, pero no a nivel institucional. De nuevo, al comparar los resultados con la evaluación SABER 2002–2003, algunos departamentos presentaron resultados sobresalientes en esta última evaluación y no así en la anterior.

Al analizar esta información se planteó la necesidad de verificar, a través de un estudio, la posible manipulación de resultados durante la aplicación de la evaluación SABER 2002–2003.

4. DISEÑO METODOLÓGICO DE LA EVALUACIÓN SABER 2004 – CONTRASTE

En el mundo, los estudios sobre la manipulación de los resultados de las evaluaciones educativas, se han desarrollado con base en programas de evaluación que llevan varios años² y en los que la información permite hacer algún tipo de seguimiento. En estos casos, se establecen las diferencias de resultados de una evaluación a otra para determinar si hay aumentos significativos entre dos evaluaciones.

En este sentido, se parte del supuesto que las transformaciones en educación requieran tiempo para lograrse, que cambios sustanciales no se dan de manera inmediata (de un año a otro) y que los resultados en evaluaciones no aumentan significativamente de un momento a otro.

² Por ejemplo los programas de evaluación del Estado de Kentucky y del Estado de Michigan. Igualmente estudios internacionales como IAP del ETS.

De acuerdo con lo anterior, si en algunas instituciones se manipuló la aplicación de las pruebas 2002–2003 para obtener resultados altos, al realizar una nueva aplicación –de contraste– en la cual se controlen los eventos que pueden alterar los resultados, es de esperar que los resultados de los desempeños de los estudiantes fuesen menores, significativamente, en la aplicación de contraste.

Así, para seleccionar la muestra de la aplicación SABER 2004 – Contraste se establecieron los siguientes criterios:

- Instituciones cuyos resultados en la evaluación SABER 2002–2003 las colocaran en el 5% de mejores resultados en cada departamento del país.
- Instituciones con quinto y noveno grados.
- Instituciones que en SABER 2002–2003 tuvieran un mínimo de 10 estudiantes evaluados en cada grado.

En la tabla 1 se encuentra el número de instituciones por departamento cuyos estudiantes fueron evaluados en SABER 2004 – Contraste, según grado.

El número de colegios participantes y con resultados en la muestra seleccionada para SABER 2004 – Contraste, fue de 907 en grado quinto y 911 en grado noveno, para un total de 85.923 estudiantes que representan aproximadamente el 5% de las instituciones con mejores resultados en la evaluación de lenguaje de SABER 2002–2003.

Tabla 1. Número de instituciones y estudiantes participantes en la evaluación SABER 2004 – Contraste, por departamento.

DEPARTAMENTO	GRADO QUINTO		GRADO NOVENO	
	Instituciones	Estudiantes	Instituciones	Estudiantes
AMAZONAS	2	207	2	213
ANTIOQUIA	139	7.511	145	8.402
ARAUCA	7	210	7	263
ATLÁNTICO	30	1.162	30	1.068
BOGOTÁ	37	2.061	35	1.544
BOLÍVAR	37	2.339	37	1.832
BOYACÁ	44	2.468	44	2.486
CALDAS	21	1.043	21	932
CAQUETÁ	8	326	8	198
CASANARE	7	578	7	473
CAUCA	23	990	22	663
CESAR	24	1.624	24	1.082
CHOCÓ	12	466	12	548
CÓRDOBA	42	2.416	41	1.805
CUNDINAMARCA	66	3.247	67	3.023
GUAINÍA	1	65	1	40
GUAVIARE	1	11	1	23
HUILA	29	1.251	29	1.017
LA GUAJIRA	11	624	10	491
MAGDALENA	22	923	22	606
META	16	863	16	886
NARIÑO	38	1.452	38	1.505
NTE SANTANDER	28	1.290	28	1.223
PUTUMAYO	7	333	7	213
QUINDIO	11	390	11	353
RISARALDA	16	708	18	806
SAN ANDRÉS	2	147	2	151
SANTANDER	34	1.481	37	1.423
SUCRE	25	1.377	25	1.218
TOLIMA	39	1.854	39	1.715
VALLE	126	5.370	123	4.769
VAUPÉS	1	27	1	43
VICHADA	1	28	1	67
TOTAL	907	44.842	911	41.081

5. APLICACIÓN DE SABER 2004 – CONTRASTE

La aplicación de SABER 2004 – Contraste se diseñó para evitar, en lo posible, la alteración de los resultados durante la aplicación. En este sentido se implementó el siguiente procedimiento:

1. Las pruebas se enviaron a cada secretaría de educación departamental.
2. Un delegado del ICFES viajó a cada secretaría de educación y contrató a los delegados institucionales en cada departamento.
3. Se capacitó a todos los delegados institucionales para controlar la aplicación.
4. Los delegados institucionales llevaron las pruebas a los colegios y, con la colaboración de la institución, las aplicó a los estudiantes de los grados seleccionados.
5. Los delegados institucionales recogieron las hojas de respuesta y pruebas y las llevaron a la capital de cada departamento.
6. Las hojas de respuesta fueron enviadas al ICFES por las secretarías de educación.

6. DEPURACIÓN DE LOS DATOS

Luego de la aplicación de las pruebas y de recoger las hojas de respuesta, se inicia el proceso de lectura de respuestas y depuración de bases de datos. Uno de los fundamentos esenciales de cualquier proceso de evaluación tiene que ver con la calidad de los datos utilizados para la obtención de resultados y su consecuente interpretación. En este sentido se habla de la confiabilidad de la información producida en el proceso de evaluación, es decir, de los resultados obtenidos.

Es común encontrar información que produce ruido en las bases de datos lo que hace necesario diseñar procedimientos y establecer criterios

que permitan identificar este tipo de datos con miras a la depuración de información.

El aspecto que se tuvo en cuenta fue la similitud de información, esto ocurre cuando un estudiante o un grupo de estudiantes copia por lo menos el 80% de todas las respuestas de las preguntas de un examen a otra persona. Se considera que un grupo de estudiantes que presentan pruebas en el mismo lugar, hay similitud en la información si las respuestas equivocadas son idénticas. Esta similitud en la información ocasiona inconsistencia en los datos y disminución de la confiabilidad de los mismos, lo que tiene como consecuencia que el promedio y la desviación estándar de una institución no sean estimaciones confiables del desempeño de los estudiantes.

El procedimiento regular implica que, para la producción de resultados departamentales, se utilicen bases de datos en las que no se incluya información de instituciones que presentan similitud en la información en, por lo menos, el 80% de sus estudiantes.

En la siguiente tabla se indica el número de instituciones cuyos datos se eliminaron de la base de datos según grado y área evaluada.

Tabla 2. Número de instituciones cuyos datos se eliminaron de la base de datos por superar el índice de similitud en la información.

PRUEBA	GRADO	INSTITUCIONES
LENGUAJE	5	20
LENGUAJE	9	193
MATEMÁTICA	5	12
MATEMÁTICA	9	28

Tabla 3. Número de instituciones y estudiantes participantes en la evaluación SABER 2004 – Contraste en lenguaje, por departamento, después de la depuración por similitud en la información.

DEPARTAMENTO	QUINTO		NOVENO	
	Instituciones	Estudiantes	Instituciones	Estudiantes
AMAZONAS	2	207	2	213
ANTIOQUIA	138	7.433	107	5.203
ARAUCA	7	210	7	263
ATLÁNTICO	29	1.153	23	712
BOGOTÁ	37	2.061	12	404
BOLÍVAR	35	2.158	28	1.174
BOYACA	43	2.448	40	1.799
CALDAS	19	945	19	856
CAQUETÁ	8	326	8	198
CASANARE	7	578	6	407
CAUCA	22	950	17	533
CESAR	23	1.611	16	588
CHOCÓ	11	456	10	486
CÓRDOBA	42	2.382	35	1.536
CUNDINAMARCA	66	3.247	54	2.458
GUAINÍA	1	65	1	40
GUAVIARE	1	11	1	23
HUILA	29	1.251	23	672
LA GUAJIRA	10	555	10	491
MAGDALENA	20	879	17	531
META	16	861	15	812
NARIÑO	37	1.430	31	1.234
NTE SANTANDER	28	1.290	20	828
PUTUMAYO	7	333	6	171
QUINDÍO	11	390	10	263
RISARALDA	16	708	15	643
SAN ANDRÉS	2	147	2	151
SANTANDER	34	1.481	29	1.156
SUCRE	23	1.259	19	1.028
TOLIMA	37	1.791	34	1.404
VALLE	124	5.266	99	3.506
VAUPÉS	1	27	1	43
VICHADA	1	28	1	67
TOTAL	887	43.937	718	29.893

Tabla 4. Número de instituciones y estudiantes participantes en la evaluación SABER 2004 – Contraste en matemática, por departamento, después de la depuración por similitud en la información.

DEPARTAMENTO	QUINTO		NOVENO	
	Instituciones	Estudiantes	Instituciones	Estudiantes
AMAZONAS	2	207	2	213
ANTIOQUIA	138	7.433	107	5.203
ARAUCA	7	210	7	263
ATLÁNTICO	29	1.153	23	712
BOGOTÁ	37	2.061	12	404
BOLÍVAR	35	2.158	28	1.174
BOYACA	43	2.448	40	1.799
CALDAS	19	945	19	856
CAQUETÁ	8	326	8	198
CASANARE	7	578	6	407
CAUCA	22	950	17	533
CESAR	23	1.611	16	588
CHOCÓ	11	456	10	486
CÓRDOBA	42	2.382	35	1.536
CUNDINAMARCA	66	3.247	54	2.458
GUAINÍA	1	65	1	40
GUAVIARE	1	11	1	23
HUILA	29	1.251	23	672
LA GUAJIRA	10	555	10	491
MAGDALENA	20	879	17	531
META	16	861	15	812
NARIÑO	37	1.430	31	1.234
NTE SANTANDER	28	1.290	20	828
PUTUMAYO	7	333	6	171
QUINDÍO	11	390	10	263
RISARALDA	16	708	15	643
SAN ANDRÉS	2	147	2	151
SANTANDER	34	1.481	29	1.156
SUCRE	23	1.259	19	1.028
TOLIMA	37	1.791	34	1.404
VALL E	124	5.266	99	3.506
VAUPÉS	1	27	1	43
VICHADA	1	28	1	67
TOTAL	887	43.937	718	29.893

7. PROCESAMIENTO DE DATOS

Con las bases ya depuradas se inician dos procesos, uno enfocado a la verificación estadística de la calidad de los ítems y de la prueba y el otro a establecer las escalas de calificación.

7.1 Análisis de ítems y pruebas

El análisis de ítems y pruebas tiene en cuenta aspectos cualitativos y cuantitativos con el propósito de verificar la calidad de éstos en relación con el propósito de evaluación que se plantea. Es un estudio pormenorizado que vincula los marcos teóricos y las estructuras de las pruebas con la información numérica que proviene del procesamiento de datos.

7.1.1 Análisis cualitativo

Las pruebas de contraste SABER aplicadas en noviembre de 2004 estaban orientadas, de acuerdo con lo establecido en los marcos teóricos, a medir las competencias de los estudiantes en las áreas de lenguaje y matemática, es decir, aquello que los estudiantes hacen con lo que saben en diferentes contextos y situaciones. Para cada área, la prueba de grado quinto constaba de 25 ítems y la de noveno de 35.

La prueba de matemática se concentró en medir, mediante el enfoque de formulación y resolución de problemas, el uso que el estudiante hace de la matemática para comprender, utilizar, aplicar y comunicar conceptos y procedimientos, que se consideran básicos en la matemática escolar y que están presentes en los estándares de calidad del área. El estudiante debía enfrentarse a problemas rutinarios y no rutinarios que estaban formulados desde lo simple hasta lo complejo. Para llegar a la respuesta esperada, según el nivel de logro de cada ítem (B, C o D para grado quinto y C, D, E y F)³, el alumno debía entender el enunciado,

³ Cada nivel de logro, es un espacio caracterizado por el agrupamiento de preguntas que cumplen con unos rasgos particulares en razón de la dificultad de la misma y la habilidad del estudiante que la responde. En los niveles C y D se encuentran acciones y procesos que permiten relacionar los dos grados (5 y 9), mientras que en los niveles E y F se ubican las preguntas que rastrean particularidades que jalonan lo básico a niveles de acciones y procesos más complejos y elaborados.

reorganizar la información (ya sea que esta apareciera de forma explícita o implícita) y a través del uso del conocimiento adquirido establecer un camino que diera solución al problema.

Por su parte, la prueba de lenguaje rastreaba la competencia comunicativa de los estudiantes a través de la comprensión e interpretación de textos, es decir, no indagaba por contenidos específicos, sino por la manera como el estudiante hace uso de sus saberes previos para interactuar con los textos y develar sus estructuras de sentido. Las preguntas de esta prueba le exigen al estudiante ubicar información que aparece de manera explícita y literal en los textos; realizar un trabajo de selección, omisión y síntesis de información; poner en interacción sus saberes con los saberes del texto para inferir información no dicha y dar cuenta de los tipos de actos comunicativos presentes, de las intenciones, finalidades y propósitos del mismo, se espera además que los estudiantes de grado noveno den cuenta de la funcionalidad semántica y pragmática de los elementos gramaticales en la coherencia y cohesión textual.

El análisis de expertos certificó que la prueba construida para la aplicación SABER contraste cumplía con los requerimientos antes mencionados.

7.1.2 Análisis cuantitativo

Para validar estadísticamente las pruebas, con el modelo usado, se tienen en cuenta los siguientes índices:

Dificultad (δ) = es el grado de complejidad relativa que tiene cada ítem. La unidad de medida es el logit (del modelo de Rasch). El valor de la dificultad varía entre $-\infty$ y $+\infty$.

Índice P = indica el porcentaje de personas que seleccionan en cada ítem la clave.

Índice de ajuste interno (A_i) = indica la cercanía entre los datos de la población contiguos al nivel de dificultad del ítem y los esperados (según el modelo). Se considera que el ajuste interno es aceptable si el valor se encuentra entre 0,7 y 1,3 logitos.

Índice de ajuste externo (A_e) = indica la cercanía entre los datos de la población lejanos al nivel de dificultad del ítem y los esperados (según el modelo). Se considera que el ajuste externo es aceptable si el valor se encuentra entre 0,7 y 1,3 logitos.

Discriminación (D) = indica la diferencia de la proporción de alumnos que responden correctamente el ítem en el grupo de alta habilidad con respecto de los estudiantes que responden correctamente en el grupo de baja habilidad. En matemática y lenguaje se espera que la discriminación sea superior o igual a 0,15.

En las tablas 5, 6, 7 y 8 se muestra esta información para cada prueba, en cada grado.

Tabla 5. Información estadística de las preguntas de la prueba de matemática en grado quinto de SABER 2004 – Contraste

N° ítem	δ	P	A_i	A_e	D
1	0,45	38,99	0,99	1,00	0,40
2	0,48	38,41	1,09	1,12	0,31
3	1,05	27,72	1,01	1,08	0,36
4	-0,64	61,47	1,07	1,13	0,30
5	0,02	47,68	0,95	0,95	0,44
6	-0,65	61,76	0,92	0,89	0,46
7	0,82	31,72	1,10	1,16	0,29
8	0,74	33,23	1,13	1,21	0,25
9	-0,31	54,80	1,12	1,14	0,28
10	1,36	22,69	0,98	1,00	0,38
11	0,68	34,43	1,03	1,07	0,35
12	1,18	25,63	1,03	1,07	0,34
13	0,08	46,51	0,91	0,89	0,48
14	-1,44	76,18	0,90	0,85	0,43
15	-0,52	59,00	0,95	0,94	0,43
16	0,69	34,20	1,10	1,13	0,30
17	-0,53	59,17	1,05	1,07	0,3
18	0	48,28	0,96	0,94	0,44
19	-0,77	64,06	0,92	0,89	0,45
20	-0,25	53,37	0,94	0,91	0,46
21	-2,26	86,95	0,94	0,88	0,33
22	-0,58	60,28	0,87	0,85	0,51
23	0,49	38,06	0,99	0,99	0,40
24	0,06	46,88	1,02	1,02	0,38
25	-0,16	51,51	0,97	0,97	0,42

N° ítem : corresponde al orden de los ítems en cada prueba.

δ = dificultad.

P = índice P.

A_i = ajuste interno.

A_e = ajuste externo.

D = discriminación.

Tabla 6. Información estadística de las preguntas de la prueba de matemática en grado noveno de SABER 2004 - Contraste

N° ítem	δ	P	A_i	A_e	D
1	-3,53	94,26	0,96	0,85	0,22
2	-2,34	84,05	0,96	0,93	0,29
3	-0,48	48,87	1,01	1,01	0,30
4	-0,63	52,17	1,03	1,05	0,26
5	-1,95	78,49	0,90	0,82	0,40
6	-1,55	71,62	0,93	0,88	0,38
7	2,16	7,74	1,01	1,15	0,16
8	1,04	19,26	1,00	1,06	0,25
9	1,01	19,69	1,06	1,15	0,17
10	0,80	22,99	1,18	1,26	0,03
11	0,94	20,76	1,05	1,10	0,20
12	-0,55	50,38	0,93	0,91	0,41
13	0,33	31,34	0,94	0,93	0,38
14	0,08	36,51	0,90	0,89	0,44
15	0,44	29,37	1,03	1,07	0,25
16	1,07	18,91	1,04	1,11	0,20
17	-0,23	43,28	1,16	1,19	0,09
18	-0,01	38,32	1,04	1,04	0,26
19	1,32	15,54	1,06	1,19	0,15
20	1,57	12,75	0,96	0,99	0,28
21	-0,42	47,35	1,00	1,01	0,31
22	-1,46	69,96	0,94	0,91	0,37
23	0,33	31,43	0,84	0,80	0,53
24	-0,83	56,74	0,96	0,95	0,36
25	0,27	32,69	0,98	0,97	0,33
26	0,71	24,52	1,08	1,11	0,17
27	0,32	31,65	0,94	0,95	0,37
28	-0,38	46,45	0,94	0,93	0,39
29	0,35	30,96	0,97	0,97	0,34
30	1,08	18,73	1,08	1,14	0,15
31	1,10	18,46	1,08	1,18	0,15
32	0,78	23,27	1,02	1,08	0,24
33	0,08	36,43	1,07	1,08	0,22
34	-1,19	64,56	0,95	0,93	0,36
35	-0,23	43,30	0,94	0,93	0,40

N° ítem : corresponde al orden de los ítems en cada prueba.

δ = dificultad.

P = índice P.

A_i = ajuste interno.

A_e = ajuste externo.

D = discriminación.

Tabla 7. Información estadística de las preguntas de la prueba de lenguaje en grado quinto de SABER 2004 - Contraste

N° ítem	δ	P	A_i	A_e	D
1	-0,51	69,36	0,93	0,89	0,41
2	0,24	53,59	0,97	0,96	0,37
3	0,41	49,90	1,12	1,16	0,19
4	0,95	38,23	1,00	1,02	0,33
5	-1,13	79,79	1,01	1,04	0,26
6	0,14	55,98	0,96	0,94	0,39
7	-0,94	76,92	0,93	0,87	0,39
8	-0,11	61,27	0,99	0,99	0,35
9	-0,13	61,74	1,04	1,06	0,28
10	-0,08	60,54	1,03	1,05	0,29
11	0,85	40,34	1,02	1,04	0,30
12	-1,60	85,96	0,92	0,83	0,36
13	-0,60	71,00	0,94	0,90	0,40
14	0,16	55,38	0,96	0,95	0,38
15	0,02	58,58	0,99	0,98	0,35
16	0,01	58,79	1,06	1,08	0,26
17	1,39	29,57	1,12	1,21	0,16
18	0,85	40,35	1,03	1,04	0,30
19	0,61	45,63	1,03	1,04	0,30
20	0,36	51,12	1,04	1,05	0,29
21	-1,17	80,41	0,94	0,86	0,37
22	0,13	56,21	0,96	0,95	0,39
23	0,70	43,51	1,02	1,03	0,31
24	0,31	52,18	0,95	0,94	0,40
25	-0,84	75,24	0,99	0,98	0,31

N° ítem : corresponde al orden de los ítems en cada prueba.

δ = dificultad.

P = índice P.

A_i = ajuste interno.

A_e = ajuste externo.

D = discriminación.

Tabla 8. Información estadística de las preguntas de la prueba de lenguaje en grado noveno de SABER 2004 - Contraste

N° ítem	δ	P	A_i	A_e	D
1	-0,41	75,92	1,01	0,99	0,33
2	-0,51	77,47	1,08	1,17	0,23
3	-0,15	71,49	1,07	1,09	0,27
4	0,26	63,75	0,94	0,90	0,44
5	1,01	48,11	1,06	1,09	0,31
6	-0,49	77,25	1,00	0,99	0,33
7	0,55	57,84	1,08	1,09	0,29
8	0,15	66,04	0,96	0,93	0,41
9	0,14	66,17	0,93	0,90	0,44
10	2,10	26,53	1,18	1,35	0,13
11	0,42	60,52	0,93	0,90	0,45
12	-0,07	70,16	0,94	0,88	0,43
13	-0,44	76,37	1,01	0,98	0,33
14	-0,74	80,95	0,88	0,74	0,45
15	-0,22	72,85	1,03	1,05	0,31
16	0,24	64,15	0,89	0,85	0,49
17	-0,49	77,19	0,96	0,95	0,37
18	0,42	60,53	1,13	1,18	0,23
19	0,03	68,22	0,90	0,84	0,47
20	-0,94	83,48	0,95	0,86	0,36
21	-0,47	76,92	1,01	1,07	0,31
22	-0,74	80,91	0,94	0,84	0,39
23	0,16	65,72	0,94	0,92	0,43
24	-0,95	83,65	0,91	0,80	0,40
25	0,08	67,32	0,95	0,90	0,42
26	0,28	63,37	1,05	1,06	0,31
27	-0,30	74,15	0,88	0,78	0,48
28	-1,34	87,90	0,92	0,78	0,37
29	0,19	65,16	1,01	0,99	0,36
30	-1,82	91,90	0,90	0,64	0,36
31	0,06	67,78	0,97	0,94	0,39
32	1,36	40,77	1,04	1,10	0,32
33	0,55	57,94	1,19	1,26	0,17
34	1,84	31,20	1,13	1,25	0,20
35	0,22	64,61	1,13	1,17	0,23

N° ítem : corresponde al orden de los ítems en cada prueba.

δ = dificultad.

P = índice P.

A_i = ajuste interno.

A_e = ajuste externo.

D = discriminación.

Cada prueba debe estar conformada por preguntas de diferentes valores de dificultad, una pregunta se considera más fácil en relación con otra si su dificultad es menor, es decir, a menor valor más fácil (mayor porcentaje de personas que seleccionan la clave) y a mayor valor más difícil (menor porcentaje de personas que seleccionan la clave). Se espera, en general que los valores oscilen entre -2 y $+2$, cuando el valor es menor de -2 la pregunta resultó demasiado fácil, en comparación con las demás, mientras que si el valor es $+2$ fue demasiado difícil. Si el nivel de dificultad de un ítem está por fuera del rango esperado, es necesario que los expertos de las áreas revisen esta pregunta para garantizar que, conceptualmente, esta bien elaborada y que quienes la respondieron en forma correcta tienen la habilidad esperada, para dar un concepto en relación con su inclusión o no en la calificación final.

En las tablas puede observarse que en algunas preguntas de las pruebas de lenguaje y matemática de grado noveno el valor de la discriminación está fuera del rango esperado. Se hizo la revisión de estas preguntas por parte de los expertos, se verificó su validez conceptual y por lo tanto no fueron retiradas; es decir, todas las preguntas se tuvieron en cuenta para la calificación.

Se espera, en las pruebas de lenguaje y matemática, que la discriminación tome valores mayores o iguales a $0,15$. Entre mayor sea el valor, mejor es la discriminación de la pregunta; es decir, los estudiantes con las mejores habilidades responden en la forma esperada, esto es abordan correctamente las preguntas con mayores dificultades, mientras que los de menores habilidades las abordan en menor proporción. Cuando el valor de discriminación de un ítem es menor a $0,15$, quiere decir que esta pregunta no permite establecer diferencias entre quienes tienen la habilidad para responderla y quienes no. En las tablas se puede apreciar que en todas las preguntas la discriminación es mayor o igual a $0,15$, es decir que todas permiten, en su nivel de dificultad, discriminar el desempeño de los estudiantes. También se puede observar que los índices más altos se encuentran en las preguntas de matemática.

Los ajustes internos y externos se refieren a la relación que hay entre los datos reales y el modelo empleado para el análisis. Todas las preguntas cuyo índice de ajuste esté por fuera del rango establecido deben revisarse. Si el porcentaje de preguntas con valores no esperados de ajuste supera

el 10%, es necesario revisar toda la prueba. Nuevamente, en las tablas se observa que todas las preguntas cumplen con el nivel de ajuste, tanto interno como externo.

Los análisis presentados en las secciones 7.1.1 y 7.1.2 permiten tener certeza de la validez de los instrumentos utilizados.

7.2 La calificación

Dadas las características del estudio SABER 2004 – Contraste, es necesario calificar estos resultados con la misma escala que se utilizó para los resultados de SABER 2002–2003, de tal manera que los resultados de cada institución sean comparables. Existen diversas formas que permiten calificar con escalas iguales en diferentes aplicaciones; algunas de ellas requieren que un mismo grupo de personas responda las pruebas cuyos resultados se van a comparar y otras implican la utilización de algunas preguntas idénticas en las dos aplicaciones.

Las pruebas de SABER 2002–2003 fueron entregadas a las instituciones, luego de la aplicación, con miras a su utilización pedagógica por parte de los docentes. Debido a lo anterior no fue pertinente utilizar preguntas de esos cuestionarios en la aplicación SABER 2004 – Contraste.

No obstante, la escala de calificación de SABER 2002–2003 está calibrada con la escala empleada en SABER 1997–1999. De acuerdo con lo anterior, en el diseño de la evaluación SABER 2004 – Contraste, se estableció que la calibración de la escala se realizaría, también, con la escala de calificación empleada en SABER 1997–1999. Para tal efecto, se incluyeron preguntas en SABER 2004 – Contraste, que ya estuvieran calibradas⁴ con las escalas de calificación definidas. De esta manera, se puede observar si hay diferencias significativas entre los resultados de cada institución en la aplicación SABER 2004 - Contraste, con respecto a la anterior.

⁴ El procedimiento regular de calibración, para comparar los resultados de dos pruebas distintas, implica utilizar algunas preguntas iguales en los dos exámenes y comparar los valores de dificultad que tienen en las dos poblaciones en que se aplican dos pruebas. De esta manera se constituye un índice que calibra las escalas de calificación.

8. CRITERIOS DE DECISIÓN

La decisión más importante en este estudio corresponde a determinar si la disminución en el resultado de los estudiantes de una institución es significativa al comparar los resultados de la evaluación SABER 2004 – Contraste con los obtenidos en la evaluación SABER 2002–2003, lo que indicaría inconsistencia de los datos.

En diferentes estudios realizados en el mundo sobre este tema se utilizan distintos criterios. Por ejemplo en el Michigan Educational Assessment Program (MEAP), establecen la diferencia de porcentaje de estudiantes que obtienen calificación satisfactoria en las evaluaciones estatales. Si el aumento de porcentaje es notorio, se considera que hubo alta probabilidad de copia y se inicia una investigación pormenorizada.

Un criterio de tipo estadístico considera que cambios superiores a dos veces el error de medición de una evaluación, indican diferencias significativas en los resultados de dos aplicaciones distintas y, en razón de esto, inconsistencia de los datos.

Para la evaluación SABER 2004 – Contraste, se tomó el segundo criterio, es decir, el de dos veces el error de medición. En la siguiente tabla se reporta el datos correspondiente al doble del error de medición en puntaje de la escala de calificación, para cada prueba y grado, tomado como criterio de inconsistencia.

Tabla 9. Criterio de inconsistencia por prueba y grado
SABER 2004 – Contraste

PRUEBA	GRADO	CRITERIO DE INCONSISTENCIA
LENGUAJE	5	9,00
LENGUAJE	9	8,40
MATEMÁTICA	5	9,40
MATEMÁTICA	9	8,20

De acuerdo con la tabla anterior, por ejemplo, si en la evaluación SABER 2004 – Contraste, una institución disminuye el resultado promedio de sus estudiantes de quinto grado en lenguaje en 9 puntos, se considera que tiene datos no confiables y resultados no válidos.

9. RESULTADOS

9.1 Desempeño promedio por institución

Se contrastó el desempeño promedio de cada institución en SABER 2004 – Contraste en relación con el promedio en SABER 2002–2003. En la tabla 10 se relaciona el número de instituciones en las que el desempeño promedio de sus estudiantes disminuye en mayor valor que el criterio de consistencia.

Tabla 10. Número de instituciones con disminución en el desempeño promedio de sus estudiantes superior al criterio de inconsistencia.

PRUEBA	GRADO	INSTITUCIONES
LENGUAJE	5	81
LENGUAJE	9	93
MATEMÁTICA	5	153
MATEMÁTICA	9	27

Tabla 11. Número de instituciones con aumento en el desempeño promedio de sus estudiantes superior al criterio de inconsistencia.

PRUEBA	GRADO	INSTITUCIONES
LENGUAJE	5	8
LENGUAJE	9	3
MATEMÁTICA	5	26
MATEMÁTICA	9	125

En los gráficos 1, 2, 3 y 4 se muestra, para las instituciones con resultados consistentes, la relación que hay entre sus resultados promedio en SABER 2002–2003 y SABER 2004 – Contraste, en cada grado.

En cada gráfico se han dibujado dos líneas de tendencia. La línea marcada con la letra I, es una línea con una inclinación de 45° con respecto al origen de la gráfica; las instituciones que se encuentran sobre ella son aquellas cuyos promedios son iguales en las dos evaluaciones.

La línea marcada con la letra R es la de tendencia real, obtenida a partir de los datos. Las instituciones que están por debajo de ésta obtuvieron mejores resultados en la evaluación SABER 2004 – Contraste que en SABER 2002–2003 y, las que se encuentran por encima, mejores resultados en la aplicación SABER 2002–2003 que en SABER 2004 – Contraste.

Las instituciones localizadas en la parte superior izquierda son aquellas con resultados altos en SABER 2002–2003 pero bajos en SABER 2004 – Contraste. Las que se encuentran en la parte inferior derecha son aquellas que aumentaron promedio en SABER 2004 – Contraste, comparado con el obtenido en SABER 2002–2003.

Se puede observar que en los gráficos 1 y 2 la línea de tendencia real (R) indica que, en términos generales, los promedios en matemática mejoraron en la aplicación SABER 2004 – Contraste. En los gráficos 3 y 4, también se observa que la línea de tendencia real (R) indica que los promedios en lenguaje aumentaron, pero menos que en matemática (tienen pendiente menor).

Gráfico 1. Relación entre el desempeño promedio de instituciones con resultados consistentes, en SABER 2002–2003 y SABER 2004 – Contraste, en matemática de quinto grado.

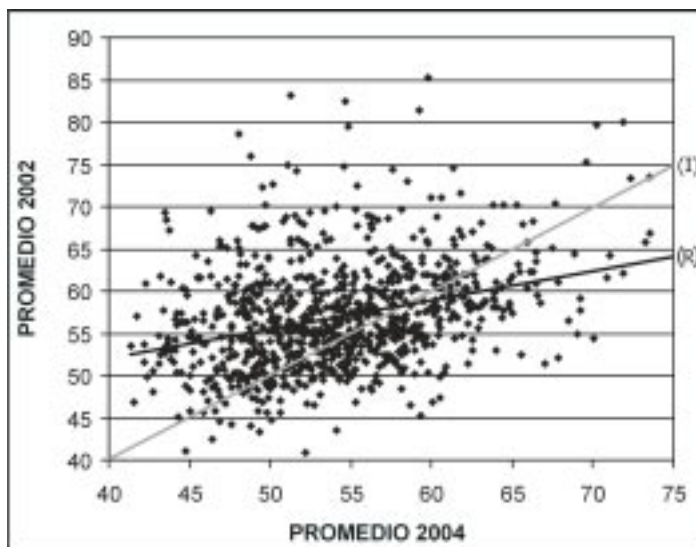


Gráfico 2. Relación entre el desempeño promedio de instituciones con resultados consistentes, en SABER 2002–2003 y SABER 2004 – Contraste, en matemática de noveno grado.

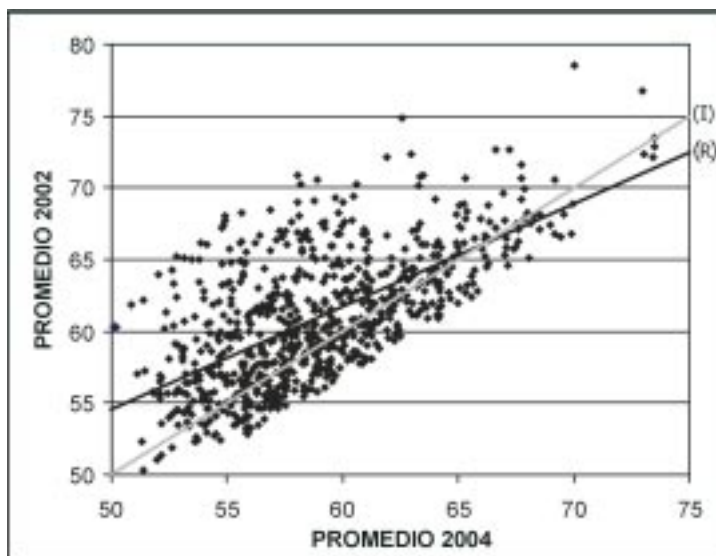


Gráfico 3. Relación entre el desempeño promedio de instituciones con resultados consistentes, en SABER 2002–2003 y SABER 2004 – Contraste, en lenguaje de quinto grado.

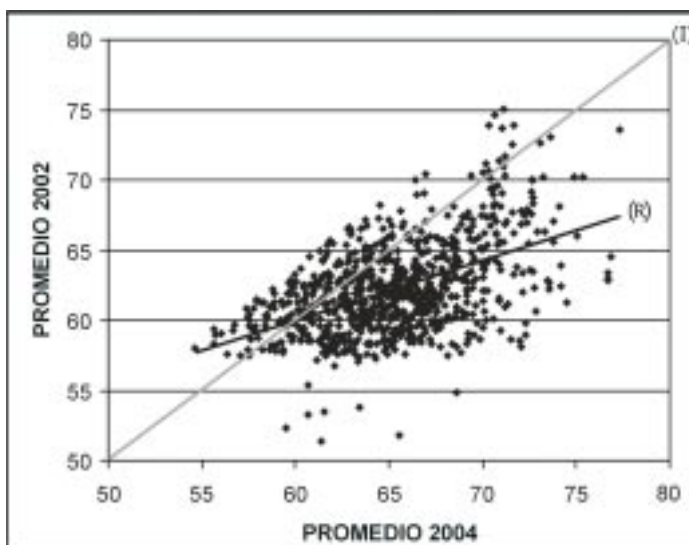
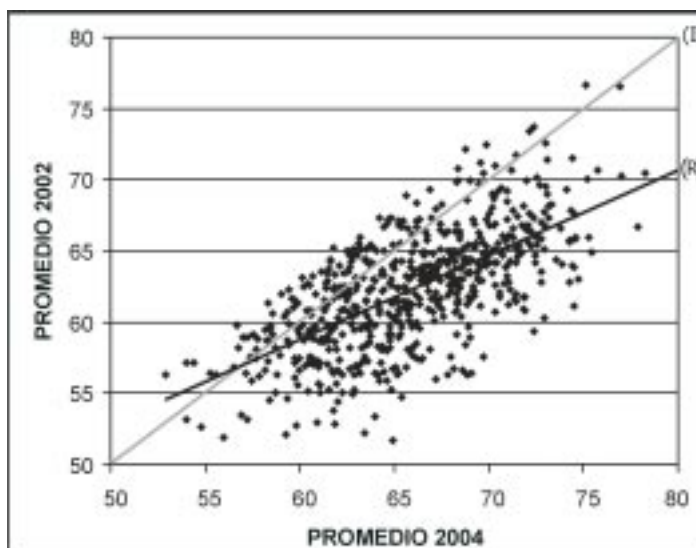


Gráfico 4. Relación entre el desempeño promedio de instituciones con resultados consistentes, en SABER 2002–2003 y SABER 2004 – Contraste, en lenguaje de noveno grado.



9.2 Desviación estándar por institución

Las diferencias en el desempeño promedio de las distintas instituciones participantes en las dos evaluaciones SABER proporcionan información acerca de lo que pudo ocurrir en cada evaluación. No obstante, es posible complementar esta información para observar el efecto de todo un proyecto educativo en los estudiantes.

La desviación estándar es un índice del grado de dispersión de una población y, como tal, indica el grado de homogeneidad de un grupo de personas, con respecto de lo evaluado en una prueba. Es importante insistir en que cada institución debería mejorar su promedio sin aumentar la dispersión.

En los gráficos 5 a 8 se encuentra la relación de las desviaciones estándar, de instituciones con resultados consistentes, en la evaluación SABER 2002–2003 y SABER 2004 – Contraste.

Al igual que en los gráficos anteriores, se han dibujado dos líneas de tendencia. La línea marcada con la letra I, es una línea con una inclinación de 45° con respecto al origen de la gráfica, las instituciones que se encuentran sobre ella, son aquellas cuyas desviaciones son iguales en las dos evaluaciones.

La línea marcada con la letra R es la de tendencia real obtenida a partir de los datos. Las instituciones que están por debajo de ésta tuvieron dispersiones más altas en la evaluación SABER 2004 – Contraste que en SABER 2002–2003 y, las que se encuentran por encima, más altas en la aplicación SABER 2002–2003 que en SABER 2004 – Contraste.

Las instituciones localizadas en la parte superior izquierda son aquellas que presentaron en SABER 2002–2003 mayores desviaciones que en SABER 2004 – Contraste. Las que se encuentran en la parte inferior derecha son aquellas en las que la dispersión aumentó significativamente en SABER 2004 – Contraste, en comparación con la de SABER 2002–2003.

Observando los gráficos 5, 6, 7 y 8, se puede concluir que la dispersión es mayor en los resultados del área de lenguaje que en matemática, es decir que en general, en SABER 2004 – Contraste, el incremento en la heterogeneidad es mayor en esta área. En matemática, es mayor la dispersión en grado quinto que en noveno.

Gráfico 5. Relación de la desviación estándar de instituciones con resultados consistentes, en matemática de quinto grado, en SABER 2002–2003 y SABER 2004 – Contraste.

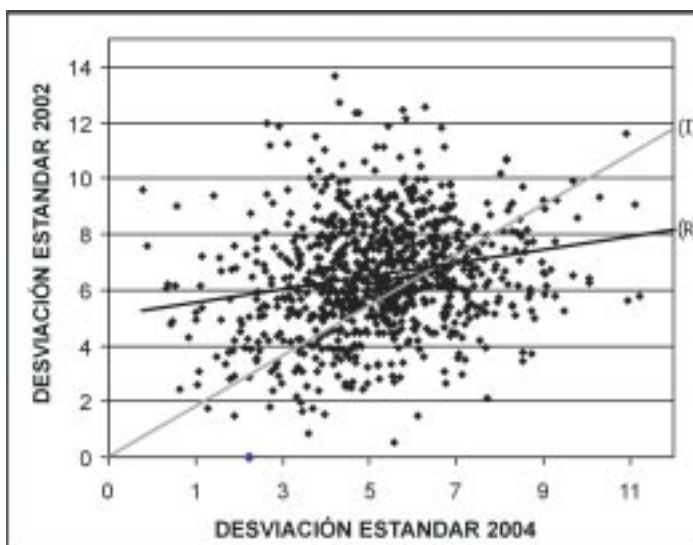


Gráfico 6. Relación de la desviación estándar de instituciones con resultados consistentes, en matemática de noveno grado, en SABER 2002–2003 y SABER 2004 – Contraste.

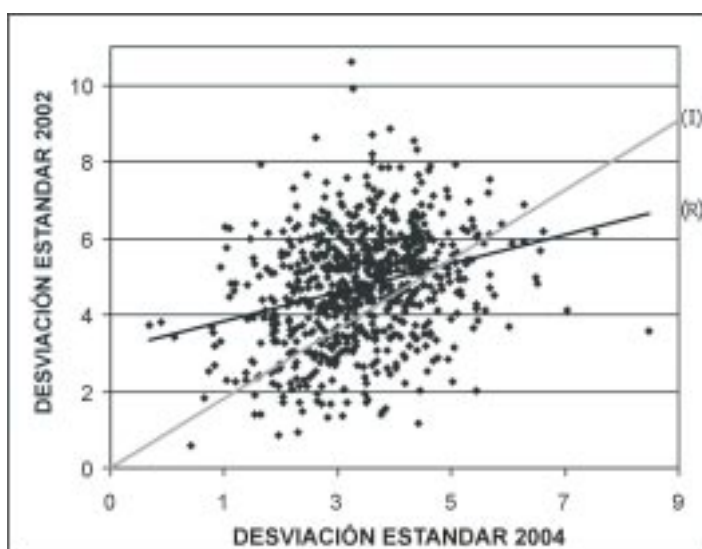


Gráfico 7. Relación de la desviación estándar de instituciones con resultados consistentes, en lenguaje de quinto grado, en SABER 2002–2003 y SABER 2004 – Contraste.

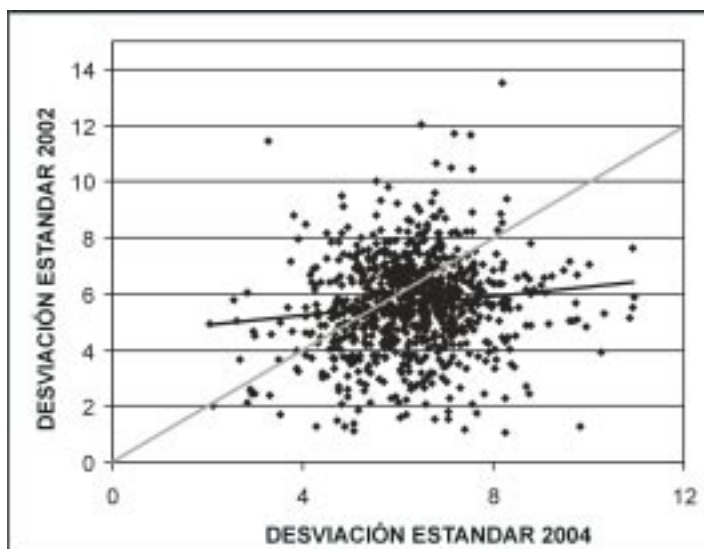
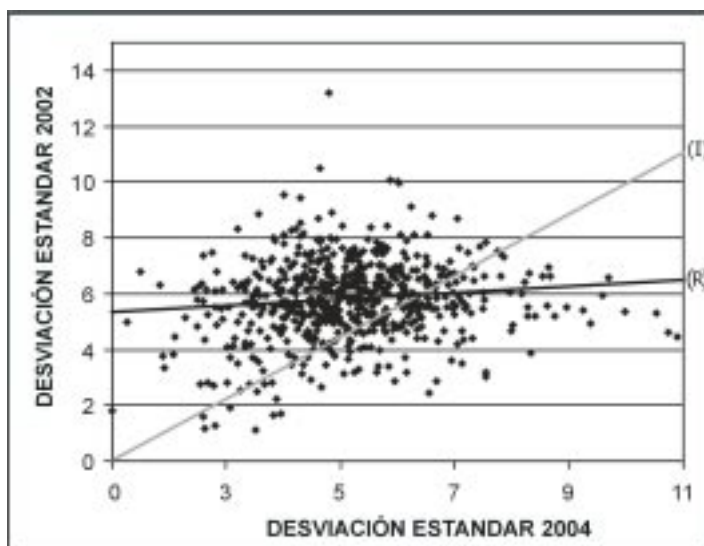


Gráfico 8. Relación de la desviación estándar de instituciones con resultados consistentes, en lenguaje de noveno grado, en SABER 2002–2003 y SABER 2004 – Contraste.



10. CONCLUSIONES

Al revisar los resultados de las instituciones que participaron en la evaluación SABER 2002–2003 y SABER 2004 – Contraste, se encontraron tres casos: instituciones que mejoran en su desempeño promedio, superando incluso el criterio de inconsistencia establecido; otras que alcanzaron resultados semejantes en las dos evaluaciones y aquellas cuyo desempeño promedio disminuyó, en algunos casos en más del criterio de inconsistencia establecido. En ningún caso se pueden decir, con total certeza, las razones para que haya sucedido cualquiera de las situaciones planteadas.

Es posible que el avance en algunas instituciones se deba, en alguna medida, a las acciones emprendidas por el gobierno nacional relacionadas con la política de mejoramiento de la calidad de la educación colombiana, como el desarrollo y divulgación de los estándares en diferentes áreas de la educación básica, la orientación a secretarios de educación y directivos docentes y docentes en la implementación de planes de mejoramiento, el acompañamiento a instituciones para dar mayor solidez a la fundamentación de su trabajo en el aula; en fin, son varias las acciones que al momento de realizar las evaluaciones pueden haber fructificado y mostrar, en los resultados de las mismas, una mejoría gradual. También puede deberse a que algunos directivos o docentes hayan realizado diferentes acciones para lograr, así, mejorar el desempeño, como por ejemplo, hacer cambios en los currículos producto de analizar con rigurosidad y detalle los resultados de evaluaciones como las de SABER o las que la institución realice.

11. REFLEXIONES Y RECOMENDACIONES FINALES

Desde hace varios años el ICFES ha promovido que la evaluación forme parte integral de los procesos educativos y que no sea considerada como un elemento separado, con connotaciones de tipo negativo. El propósito ha sido transformar la cultura de la evaluación que existe en nuestro país, tratando de mostrar sentidos y significados positivos y

constructivos de la evaluación, que se articulen mucho mejor con las propuestas y metas actuales de la educación.

Se trata de cambiar la visión punitiva y estigmatizadora de la evaluación por una más amable que permita reconocer en la evaluación la oportunidad de propiciar el mejoramiento de la calidad de la educación y, por ende, de la calidad de vida de los colombianos. En este sentido, es importante que las instituciones educativas:

- Participen en el desarrollo de la evaluación SABER, coordinada por El MEN y el ICFES. Los directivos docentes y los docentes de todas las instituciones educativas del país pueden aportar sus reflexiones en torno a los aspectos disciplinares, pedagógicos y sociales que se encuentran implícitos o explícitos en las preguntas de las pruebas aplicadas. De esta forma estarán contribuyendo al perfeccionamiento de los instrumentos que serán utilizados en futuras aplicaciones.
- Apoyen los operativos de aplicación. La logística de la aplicación, coordinada entre el ICFES y las secretarías de educación, requiere del compromiso de jefes de núcleo, directivos docentes y docentes quienes, con su participación transparente en el proceso, ayudan a construir el verdadero sentido y significado de la evaluación educativa, asumiéndola como una oportunidad de autoreconocimiento para orientar acciones de mejoramiento.
- Analicen y reflexionen constructivamente los resultados de las evaluaciones. De nuevo, son los docentes, los directivos docentes y la comunidad en general, quienes, demostrando su compromiso con el mejoramiento de la calidad educativa, pueden aprovechar los reportes de resultados, confrontándolos con los resultados esperados, de acuerdo con el Proyecto Educativo Institucional, local y nacional, para comprender las particularidades de los procesos pedagógicos y vincular dichos resultados con planes de mejoramiento.

La evaluación SABER es de todos, no es un proceso ajeno al que debemos mostrar un rostro mejor. La evaluación permite reconocernos mutuamente y contribuir a la comprensión del proceso educativo con miras a su reorientación para lograr las metas que requiere el país.

Los resultados de la evaluación SABER 2004 – Contraste, además de evidenciar que aún no se ha construido totalmente una cultura de la evaluación en Colombia, indica que, en algunos casos (instituciones y departamentos), se ha logrado una mejoría, un avance importante, que demuestra el compromiso con el mejoramiento de la calidad de la educación. Es importante reconocer los cambios que han llevado a ese mejoramiento para fortalecer los procesos que los acompañan y valorar la pertinencia de los mismos.

Si se parte de entender la evaluación como el proceso que permite el reconocimiento del grado de desarrollo de una persona (o institución), como consecuencia de sus interacciones educativas, y que tiene como propósito orientar transformaciones para responder a las exigencias de calidad de los contextos socioculturales en el cual se está inmerso, se entenderá que la evaluación educativa forma parte integral de los procesos pedagógicos que se desarrollan en la institución y por lo tanto debe ser permanente, participativa y basada en información útil, sistemática y precisa, cuyos resultados sean comprensibles y aprovechables por toda la comunidad, con miras al mejoramiento del ser humano.



Se terminó de imprimir en
el Grupo de Procesos Editoriales de
la Secretaría General del ICFES
en julio de 2005