

La verificación de la calidad de la educación en Colombia¹

Ricardo Lucio A.²

Abstract

The document presents an analysis of the massive quality evaluation mechanisms in Colombia's basic and middle educative system. In particular, the study covers international massive evaluation systems applied in the country, the national massive evaluation system, the analysis of associated factors to attainment, qualitative studies made, and the opinion of experts. Also, the results and reach of the education's quality massive evaluation mechanisms are discussed. The main conclusion is that the road the country has to travel is very long, not only towards the improvement of the education's quality, but also to the enhancement of the instruments that are used to verify it.

Resumen

En el presente documento se hace un análisis de las herramientas con que cuenta Colombia para evaluar de manera global la calidad de su sistema de educación básica y media. En particular, el estudio cubre los sistemas internacionales de medición masiva de los resultados aplicados en el país, el sistema nacional de evaluación masiva, el análisis de los factores asociados al logro, los estudios cualitativos hechos y los juicios de expertos. A su vez, se discuten los resultados y el alcance de los mecanismos de evaluación masiva de la calidad de la educación. La conclusión principal es que todavía es muy largo el camino que tiene que recorrer el país, no sólo en cuanto al mejoramiento de la calidad de nuestra educación, sino también para alcanzar el perfeccionamiento de los instrumentos que elaboramos para verificarla.

Keywords: Education, Quality Measurement, Evaluation.

Palabras clave: Educación, Medidas de calidad, Evaluación.

Clasificación JEL: I20, I21, I28, I29

Primera versión recibida en mayo 22 de 2006; versión final aceptada en junio 12 de 2006

Coyuntura Social No. 34, junio de 2006, pp. 113-161. Fedesarrollo, Bogotá - Colombia.

¹ El presente documento buscó inicialmente complementar el estudio que sobre la evolución del sistema educativo en Colombia realizó en 2004 un grupo de trabajo del Programa Nacional de Desarrollo Humano (PNUD), coordinado por Alfredo Sarmiento, su director. Para su elaboración el autor contó con el apoyo financiero de la oficina de la CEPAL en Bogotá. En la versión actual, revisada a comienzos de 2006, se han actualizado algunos datos, especialmente los que tienen que ver con la última aplicación de las pruebas SABER en 2005, a nivel nacional, y la última aplicación de PISA (2003), en lo internacional. Las mediciones y cálculos estadísticos de las pruebas del ICFES fueron realizados por Natalia Ariza, a partir de las bases de datos suministradas por esa entidad.

² Licenciado en Filosofía y Letras, Universidad Javeriana, y Doctor en Educación, Universidad de Bonn, Alemania. Actualmente se desempeña como consultor independiente.

The state ran schools, but could it teach your children to love reading, or was that your job?

Salman Rushdie

I. Introducción

Los alcances, la expansión y la evolución del sistema educativo generalmente se miden y analizan mediante una serie de índices e indicadores que combinan unos cuantos datos estadísticos, que se refieren fundamentalmente a la cantidad y características de los planteles que funcionan en los distintos niveles (preescolar, educación básica, educación media y educación terciaria), a la cantidad y características de los alumnos que cursan los diferentes grados, a los recursos -humanos y físicos- involucrados en el funcionamiento de en dichos planteles y al dinero que cuesta todo el proceso. Los resultados del sistema, a su vez, se analizan a partir de una cifra que expresa el número promedio de años que han permanecido las personas de un país o región en el sistema educativo. Dicha cifra permite no solamente comparar los resultados (el desempeño) de varios sistemas educativos nacionales entre sí, sino también los resultados de varios subsistemas o regiones dentro de un mismo sistema o país, así como su evolución y progreso a lo largo del tiempo.

El problema con este indicador es que se lo trabaja como si fuera una unidad de medida homogénea de la educación recibida en todos los sistemas educativos, a lo largo de períodos prolongados de tiempo. En otras palabras, se asume que, para todas las personas y en todos

los planteles educativos, un año de escolaridad, en términos de educación recibida o cursada, vale lo mismo. Nuestro sentido común, sin embargo, nos impulsa a afirmar lo contrario: sin atrevernos a precisar todavía en qué consiste la calidad de la educación, creemos sin embargo que no es lo mismo cinco años de educación cursados en un colegio bueno -o de buena calidad- que la misma escolaridad obtenida en un colegio malo -o de calidad deficiente-.

Los análisis comparativos sobre la calidad de la educación, tan de moda en las últimas décadas, confirman esa sospecha del sentido común. Para citar un ejemplo, para el caso latinoamericano, tenemos que los estudios comparativos realizados por el LLECE³ a finales de la década de los noventa permiten concluir que las diferencias en la calidad de los aprendizajes escolares realizados en las áreas de lectura y matemáticas (por niños de tercero y cuarto grado) son muy parecidos en los diferentes países latinoamericanos participantes en el estudio y las diferencias son bien poco significativas, si se excluye a Cuba. No obstante, si se incluye a Cuba en el análisis, el cuadro cambia radicalmente, puesto que los resultados obtenidos en el estudio por los niños de este país son significativamente superiores a los del resto de los países participantes. En otras palabras, y para expresarlo en términos cuantitativos: si las tendencias observadas se conservan a lo largo de toda la educación básica, es altamente probable que un grupo poblacional tomado al azar en Cuba con un promedio de educación de 6,7 años tenga bastante más educación que un grupo similar con la misma escolaridad en cualquiera de los países de América Latina.

³ Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación.

De esta reflexión surge la necesidad de ponderar o refinar el análisis del desempeño de los sistemas educativos teniendo en cuenta el factor *calidad*, en particular el papel que ha jugado y las modificaciones que ha sufrido en la expansión y desarrollo de los mismos. Esta consideración es tanto más importante, cuanto que del análisis del comportamiento del sector educativo estamos sacando conclusiones sobre la evolución de la calidad de vida de la población⁴; lo cual es para muchos la razón última de ser de todo esfuerzo colectivo en lo económico, lo social o lo cultural, cualquiera que sea el rasero con que se lo mida.

El presente documento tiene como objetivo recopilar la evidencia empírica disponible que permita diagnosticar la calidad de la educación en el país y su evolución probable en las últimas décadas. Se buscará para ello, inicialmente, realizar un acercamiento a la noción de calidad en el sector educativo (sección II), para pasar en seguida a examinar los instrumentos con que contamos para verificar esa calidad en el país y los principales resultados de su aplicación (sección III). Se mencionan a renglón seguido algunas aproximaciones complementarias a la calidad de la educación (sección IV), para terminar con algunas observaciones críticas (Sección V) y unas anotaciones finales sobre el fortalecimiento de la calidad dentro del funcionamiento actual del sistema educativo colombiano (sección VI).

II. La noción de calidad en el sector educativo

A. ¿Qué es calidad de la educación?

Teóricos, educadores, políticos y los simples ciudadanos tienen con frecuencia posiciones y definiciones (así sean implícitas) muy diversas acerca de la calidad de la educación, posiciones que ni es posible resumir aquí ni, mucho menos, analizar críticamente para lograr proponer una definición común. Busquemos, más bien, una aproximación en términos de expectativas y realizaciones. Es decir, aceptemos que un bien o un servicio es bueno o malo, tiene calidad óptima, aceptable o poca, en la medida en que satisface los objetivos que se le asignaron o las expectativas que se tienen frente a él.

La pregunta sería entonces: ¿qué es lo que la sociedad espera del sistema educativo? o, ¿qué papel, qué tarea, le ha encomendado a dicho sistema? Nos limitamos, por ahora, al sistema educativo visible y formal, compuesto por los planteles de educación preescolar, básica, media y terciaria, dirigidos tanto por el Estado como por instituciones o personas particulares. Por eso, en lo que sigue, el término *sistema educativo* o simplemente *sistema* puede considerarse como el equivalente del sustantivo genérico colectivo singular la escuela⁵. La pregunta inicial por la calidad de la *escuela* se descompone entonces en otras tres:

⁴ El Índice de Desarrollo Humano calculado por el PNUD para las diferentes naciones del mundo, que podría considerarse como una estimación de la calidad de vida de las mismas, resulta de un promedio aritmético simple entre otros tres índices que miden la esperanza de vida, el nivel educativo y el ingreso promedio de la población.

⁵ Usamos en este trabajo el vocablo "escuela" en su acepción técnica para designar cualquier tipo de institución educativa formal, de nivel primario, secundario o terciario, y no en la acepción discriminatoria de uso frecuente en nuestro país, con que se designan exclusivamente las escuelas primarias oficiales ("las escolitas de los pobres...").

- La primera: ¿cuáles son nuestras expectativas con respecto a la escuela? en otras palabras, ¿cuál es el papel o la tarea que esperamos que cumpla la escuela en nuestra sociedad?
- La segunda: ¿cómo hacemos para saber si una escuela determinada está cumpliendo con esa tarea? que es lo mismo que decir ¿cómo verificamos la calidad de esa escuela concreta?
- La tercera: ¿cuál es entonces la calidad del sistema educativo, del conjunto de escuelas, o de la escuela colombiana (genérico, general)? ¿cómo hacemos para verificarla -medirla-? En esta última pregunta nos situamos en un plano diferente: pasamos de la posibilidad de constatar o verificar la calidad de una escuela particular, a la posibilidad de constatar o verificar la calidad de un conjunto de escuelas: ¿es posible dicha evaluación *masiva*? ¿cuáles aspectos de la calidad se pueden verificar en una evaluación masiva, y cuáles no? ¿con qué instrumentos contamos para realizarla? ¿cuáles son los resultados de su aplicación?

B. ¿Qué esperamos de la escuela?

Cualquiera que sea la concepción que se tenga de escuela, nadie pone hoy en duda que ésta gira en torno a dos ejes complementarios, que se entrecruzan y actúan de manera simultánea, pero que podemos distinguir para facilitar el análisis. El primero tiene que ver con el *conocimiento*. Niños, adolescentes y jóvenes en las más diversas regiones y culturas del mundo asisten a la escuela para aprender cosas, y ese aprendizaje tiene que ver con diversos tópicos: matemáticas, deportes, lengua escrita, geografía, cívica, ética o religión, historia y artes manuales, entre otros. En efecto, esos aprendizajes escolares son un poco

más organizados y sistemáticos -disciplinados, para algunos- que los aprendizajes espontáneos que realizan en la calle, en el campo, en el núcleo primario o familiar.

Pero también todos parecen estar de acuerdo en que la escuela no se reduce al conocimiento. Los niños, los adolescentes y los jóvenes que la visitan conviven en ella, se relacionan unos con otros y adquieren habilidades y destrezas sociales -el llamado currículo oculto- muchas veces más importantes o más eficientes para sus vidas cotidianas que lo que puedan aprender en las clases de ética, educación sexual o ciencias sociales. Este es el segundo eje alrededor del cual gira la escuela, que podríamos llamar la *convivencia*, que permite el desarrollo del aprendizaje y las habilidades sociales.

La combinación de los dos ejes, *conocimiento* y *convivencia*, es lo que define la tarea o la función de la escuela contemporánea. Ninguno de los dos ejes separado del otro la justifica: una escuela de sólo conocimiento podría reemplazarse por una red de computadores o por una buena enciclopedia; una escuela de solo convivencia no sería más que un club social. Eso vale para todas las escuelas del mundo; en lo que difieren unas de otras, unos sistemas educativos de otros, es en el tipo de conocimiento (teórico, práctico, crítico, innovador, etc.) y el tipo de socialización (tolerante, disciplinada, competitiva, agresiva, etc.) que en ellos se desarrolla, construye o fomenta. Es decir, los sistemas educativos se diferencian en las expectativas sociales concretas en torno a sus dos ejes básicos.

Es un error considerar que los bienes que produce la escuela -en términos de conocimiento y convivencia- son solamente instrumentos o

medios para alcanzar otros bienes: "estudia para el mañana", como se dice en algunas oportunidades. Sería más ajustado a la realidad decir que el producto de la educación escolar son a la vez bienes que valen por sí mismos y medios para obtener otros bienes. En otras palabras, el conocimiento y la convivencia que se desarrollan en la escuela no son solamente medios para obtener otros bienes, sino que son al mismo tiempo bienes que valen por sí mismos (Cuadro 1).

C. Hacia una aproximación a la calidad de la educación

Siguiendo la reflexión anterior, una educación de calidad sería aquella en la cual el sistema facilita a cabalidad el cumplimiento de sus propósitos, o sea el que sus alumnos se apropien de los conocimientos previstos y logren convivir en una atmósfera propicia para desarrollar habilidades de convivencia.

El siguiente paso consiste en establecer la manera de detectar el cumplimiento de esos propósitos, en otras palabras, verificar o medir la calidad de la educación en una escuela, sub-sistema o sistema educativo nacional.

Esa medición debería permitir idealmente, al menos lo siguiente:

1. Captar qué tan a gusto se encuentran los usuarios en el sistema o escuela, cuál es el nivel de satisfacción -de felicidad- del niño en la escuela. Si los bienes que arroja el sistema no sólo son medios para lograr otros bienes, sino que tienen valor por sí mismos, se espera que los usuarios, es decir los alumnos, se sientan atraídos por ellos, y se sientan a gusto con lo que están desarrollando en la escuela.
2. Detectar si se sabe, si está claro cuáles son esos conocimientos y esos comportamientos

Cuadro 1

EJES O FUNCIONES BÁSICAS DE LA ECUELA

Ejes o funciones básicas de la escuela		
	Conocimiento (Organizado y sistemático)	Encuentro con los demás
Perspectivas de los ejes	Como bien en sí, deseable por sí mismo <ul style="list-style-type: none"> ○ El conocimiento como bien, la satisfacción de la curiosidad natural por descubrir el mundo que nos rodea. ○ El conocimiento no sólo es atractivo, sino que hace mejores a las personas: "qué bueno es aprender cosas". 	<ul style="list-style-type: none"> ○ La tendencia natural a la convivencia: qué bueno es estar con los otros, qué bueno es hacer amigos, deparar y hacer cosas con ellos. ○ La convivencia no sólo es deseable, sino que hace mejores a las personas.
	Como medio para acceder a otros bienes <ul style="list-style-type: none"> ○ Conocimiento en sus diversas acepciones -teórico, aplicado, crítico/analítico- ○ Para que la persona sea útil a la sociedad (utilidad social). ○ Para que la persona pueda acceder a otros bienes (utilidad individual). 	<ul style="list-style-type: none"> ○ El aprendizaje social (valores, destrezas sociales) necesario para el correcto funcionamiento de la sociedad.

de convivencia que la escuela, el subsistema o el sistema pretende que desarrollen los alumnos. Aquí se toca tanto el problema curricular como el de la convivencia deseada en la escuela. La particularización de los propósitos de conocimiento y convivencia puede hacerse a nivel de escuela, a nivel de subsistema o a nivel de sistema nacional.

3. Verificar qué tan pertinentes son esos conocimientos y esas expectativas de convivencia para el hombre y la sociedad del siglo XXI. Vale decir, si están orientados a que el hombre resuelva sus problemas de felicidad, ubicación, pertinencia y funcionalidad a la sociedad del conocimiento y de la producción (lo deseado), pero que también le permita lidiar con la sociedad de la violencia, exclusión y de las carencias y desigualdades que, todo parece indicar, nos depara el siglo XXI (lo real, lo presente).
4. Establecer en qué medida se han apropiado los alumnos de esos conocimientos o de esas habilidades (cognitivas o de convivencia). En el caso del conocimiento esta apropiación puede ser la prevista en el numeral 2, o puede tratarse del desarrollo de conocimiento o de saberes no previstos (el llamado currículo oculto). Otro tanto sucede con el desarrollo de habilidades sociales o de convivencia; en este caso los efectos colaterales del currículo oculto son mucho más importantes; piénsese no más en el tipo de aprendizaje que sobre comportamientos o actitudes sexuales se desarrollan de manera informal, no prevista, en la escuela.

Las verificaciones correspondientes a los numerales 2, 3 y 4 deben hacerse tanto para los propósitos de conocimiento como para los de

convivencia en la escuela. Las cuatro suponen que se disponga de instrumentos de medición que permitan hacer comparaciones -longitudinales o *diacrónicas* y transversales o *sincrónicas*-. En especial, cuando se trata de detectar la calidad de los sistemas nacionales, estas verificaciones deben ser susceptibles de aplicarse de manera masiva.

Ahora bien, no siempre estamos en condiciones de verificar masivamente los cuatro puntos mencionados. La verificación masiva de la calidad de los sistemas educativos suele tener unas características que se desprenden de su índole: entre más amplia quiera ser su cobertura, más tienden a simplificarse los instrumentos de medida para hacer posible su aplicación. En esta media, se reduce o simplifica lo que termina por evaluarse. Quizá por esta razón las mediciones masivas internacionales y domésticas se reducen casi todas a verificar aspectos predominantemente cognitivos, especialmente la apropiación de saberes relativos a las asignaturas académicas o escolares (numeral 4), con algunas incursiones mucho menos significativas en los otros temas, tales como análisis del currículo (numeral 2) y la pertinencia del mismo (numeral 3).

III. La verificación de la calidad de la educación en Colombia

A. Sistemas de verificación

Cuatro sistemas han medido los resultados del aprendizaje realizado en las escuelas colombianas por lo que, con las limitaciones que acabamos de mencionar, podrían considerarse como aproximaciones a una verificación de algunos aspectos de la calidad de la educación en el país: los llamados "exámenes de estado" del ICES y las pruebas aplicadas por el SABER, el TIMSS y el LLECE.

1. ICFES

El Servicio Nacional de Pruebas del ICFES viene practicando exámenes de conocimientos de manera ininterrumpida a los egresados o a punto de egresar de la educación secundaria (grado 11), dos veces al año desde finales de la década de los sesenta.

El servicio nació como una dependencia semiautónoma coordinada por las autoridades universitarias del país, y después fue integrado al ICFES, institución creada en 1968 para ser la autoridad máxima que fomentara y regulara el funcionamiento de las instituciones de educación postsecundaria del país. Esta articulación explica

una de las características principales de estas pruebas: son inicialmente pruebas de contenido muy académico, elaboradas por docentes universitarios, cuyo objetivo original no era propiamente verificar la calidad de la educación básica, sino examinar la idoneidad de aquellos egresados de la secundaria que aspiraban ingresar a una institución de educación terciaria, y que normalmente son menos del cincuenta por ciento de los que se gradúan como bachilleres (ver recuadro 1).

La evolución del servicio durante las últimas décadas del siglo pasado se realizó en dos aspectos más o menos paralelos: uno de cobertura y otro burocrático. La cobertura de los exámenes se incrementó de tal manera que, hacia finales de la

Recuadro 1
¿Qué tan representativas son las pruebas del Icfes?

Podemos comparar la matrícula al final del bachillerato, según los datos de septiembre de cada año de la Encuesta Nacional de Hogares, los alumnos que se someten durante el mismo año a las pruebas del ICFES, de acuerdo con las estadísticas publicadas por esa entidad, y los que realmente ingresan a la universidad, según los boletines de estadística de la misma entidad, para algunos años de la década pasada, como sigue:

(1) Año	(2) Estudiantes matriculados grado 11	(3) Estudiantes que presentaron el examen del ICFES	(4) = (3)/(2) (%)	(5) Estudiantes matriculados en la universidad	(6) = (5)/(2 _{t-1}) (%)
1992	462.567	253.574	55	170.013	
1993	484.997	265.596	55	187.948	41
1994	550.542	288.571	52	221.360	46
1995	574.683	319.715	56	257.239	47
1996	619.727	369.813	60	263.991	46
1997	660.868	405.223	61	298.128	48
1998	624.878	426.422	68	267.866	41
1999	674.171			230.221	37
2000	715.011			267.950	40

Fuente: ICFES, MEN y Encuesta Nacional de Hogares (DANE). Cálculos del autor.

Aún teniendo en cuenta las inconsistencias que puedan tener las series, podemos aventurar las siguientes conclusiones:

- 1) Efectivamente, la cobertura de las pruebas se incrementó de manera sostenida durante la década (columna 4), aunque es poco probable que hacia el final de la misma hubiera alcanzado el 90 por ciento, como lo afirman algunos funcionarios del ICFES.
- 2) En cambio los ingresos a la educación superior nunca alcanzan el cincuenta por ciento de los alumnos que cursan el último año de secundaria en el año inmediatamente anterior (columna 6).

década de los noventa, más de dos tercios de los alumnos que terminan secundaria se someten a ellos aunque, en principio, sólo están obligados a hacerlo los que aspiran a un cupo en una institución del nivel postsecundario (ver recuadro 1). El estatus social de las pruebas subió: ante la devaluación del cartón de bachillerato al nivel de una simple credencial, empezó a tener más sentido para el nuevo bachiller la presentación del examen del ICFES y la obtención de un buen puntaje en el mismo; los contenidos de la prueba, que no siempre coincidían con las directrices ministeriales relativas al plan de estudios⁶, se convirtieron con el tiempo en el currículo en la sombra del último grado de la secundaria.

Esta transformación se manifestó también en el terreno burocrático: el Servicio Nacional de Pruebas se integró primero como entidad semi-independiente al ICFES, se convirtió posteriormente en una subdirección del mismo, para terminar, a comienzos del nuevo siglo, absorbiendo prácticamente a la institución entera. Efectivamente, las funciones de coordinación del nivel terciario de la educación perdieron la autonomía relativa que tenían en el ICFES, para ser asumidas directamente, en la forma de un viceministerio, por el Ministerio de Educación. El ICFES, por su parte, quedó prácticamente reducido a una entidad técnica

encargada de elaborar y aplicar pruebas masivas de evaluación, precisamente en los momentos en que está en todo su furor el boom -tanto nacional como internacional- de la evaluación masiva de la calidad de la educación.

Las pruebas sufrieron una mutación importante a partir del año 2000. Aunque siguieron teniendo un estilo marcadamente académico⁷, el ICFES insistió en que un nuevo diseño de las pruebas a partir de esa fecha permitiría dar más prioridad al examen de competencias contextuales -basadas en la capacidad de solucionar problemas concretos de la vida cotidiana y de situaciones laborales elementales- que al de las simples competencias académicas que hasta entonces se venían examinando.

La importancia que tienen estas pruebas para cualquier estudio de la evolución en el tiempo del sistema educativo colombiano radica en sus alcances y en su cobertura. Por un lado, es la única prueba con cobertura prácticamente universal que se viene realizando ininterrumpidamente en todo el territorio nacional por más de tres décadas. Por otro lado, los alcances de la prueba, que se desarrolla en tres sesiones durante dos días consecutivos⁸, se extienden a prácticamente todas las áreas, bastante numerosas por cierto,

⁶ En las dos últimas décadas del siglo anterior el Ministerio de Educación Nacional contó con una sola propuesta curricular oficial válida para todo el país, que se comenzó a gestar en la década de los setenta y terminó de publicarse a mediados de los ochenta. Sin embargo, publicó periódicamente (uno por década) unos listados básicos de las asignaturas obligatorias para los diferentes grados de la educación básica y media, llamados planes de estudios, como lo fueron los decretos 45 de 1962 y 1710 de 1963, 80 de 1974 y 1002 de 1982. Debido al tiempo que se tomaba en el ajuste de las pruebas del ICFES, las orientaciones de dichos planes apenas podían verse reflejadas en las mismas cinco o seis años después (véase Cuchimaque 1998, González *et al.* 1998, Rocha 1998 y Romero 1998).

⁷ Véase más adelante, el acápite C: ICFES y SABER: ¿qué están midiendo?

⁸ A partir de 2005 la prueba se realiza en dos sesiones en un solo día (mañana y tarde), aunque la cobertura por áreas del conocimiento se mantiene.

del currículo de la secundaria colombiana: matemáticas, lengua castellana, geografía, historia, biología, física, química, ciencias ambientales, filosofía, psicología, medios de comunicación, conceptualización de la violencia y un idioma lectivo (inglés, francés o alemán).

2. SABER

Las pruebas del SABER se han convertido en Colombia, a partir de la década de los noventa, en el instrumento oficial para medir periódicamente la calidad de la educación básica impartida en las escuelas colombianas.

La elaboración de las pruebas fue encomendada inicialmente a una entidad privada -el instituto SER de investigación- y posteriormente a un grupo de académicos de la Universidad Nacional, que contaban para su realización con el apoyo del Ministerio de Educación y de los municipios interesados. Al final de la década de los noventa las pruebas son elaboradas directamente también por el ICFES, que para ese entonces se estaba convirtiendo en una entidad especializada en la elaboración de este tipo de materiales.

Hasta la fecha se han aplicado cuatro tandas de pruebas: i) 1991-1994, ii) 1997-1999, iii) 2002-2003 y iv) 2005. Las pruebas se aplican a alumnos de los grados tercero, quinto, séptimo y noveno, o sea al promediar y finalizar el ciclo de educación primaria y de educación básica secundaria. Son pruebas sencillas, generalmente constan de un cuadernillo por área, cuyo diligenciamiento toma entre una hora y media y una hora y cuarenta y cinco minutos. Se aplican en las áreas de matemáticas y lenguaje, aunque esporádicamente se han aplicado algunas pruebas en ciencias naturales y educación cívica.

Las dos primeras tandas se aplicaron a muestras nacionales de escuelas, de manera que los resultados pudieran ser significativos por departamento, por sectores rural o urbano y por tipo de planteles (estatales o privados). Simultáneamente se recopiló información sobre los planteles y el contexto socioeconómico de los alumnos, de manera que pudieron realizarse algunos estudios sobre los factores asociados al logro académico. Esta información dejó de recogerse a partir de la tercera tanda. Ya en la segunda tanda Bogotá, que es el municipio más poblado del país, contrató al equipo responsable de las pruebas para diseñar y asesorar una aplicación equivalente en toda la población estudiantil de los grados 3, 5, 7 y 9; aunque la cobertura de las mismas no logró ser total, por desacuerdos con grupos de maestros estatales y la resistencia de algunos colegios privados, se estima que su cubrimiento fue superior al 70%, lo que permitió sacar conclusiones sobre el logro educativo a nivel de colegio y agruparlo por zonas de la ciudad, por estratos socioeconómicos y por tipo de colegios.

La información recaudada permitió igualmente realizar estudios importantes sobre algunos de los factores asociados al logro escolar, así como profundizar en estudios cualitativos -estudios de caso- la situación de algunos colegios con desempeño promedio sobresaliente.

La tercera tanda de evaluaciones se hizo de manera progresiva en diferentes municipios y regiones del país. Esta vez, siguiendo el ejemplo de Bogotá, se buscaron coberturas totales (censales) para los colegios y estudiantes de los grados tercero, quinto, séptimo y noveno en las áreas de matemáticas y lenguaje, aunque no se recogieron datos para indagar sobre los factores asociados al desempeño académico.

La cuarta tanda se realizó en 2005 con una cobertura total (censal) en todas las instituciones oficiales y privadas del país. La comprensión de la prueba se incrementó notablemente, puesto que abarcó no solamente las áreas de matemáticas y lenguaje que se habían aplicado en las tandas anteriores, sino también las áreas de ciencias naturales y ciencias sociales, así como el área “transversal” de competencias ciudadanas. El universo de aplicación, sin embargo, se redujo: esta vez sólo se aplicaron pruebas para todos los alumnos matriculados en los grados quinto (fin de la educación básica primaria) y noveno (fin de la educación básica secundaria).

3. TIMMS

El Tercer Estudio Internacional de Matemáticas y Ciencias -llamado TIMSS, por sus sigla en inglés- realizó una evaluación internacional en matemáticas y ciencias entre 1994 y 1995, a muestras representativas de planteles y estudiantes de los grados séptimo y octavo, de aproximadamente 13 años de edad, en algo más de 40 países. La aplicación en Colombia fue coordinada académica y técnicamente por la Universidad del Valle, con el apoyo y el aval del Ministerio de Educación.

Lo característico de este estudio fue que su aplicación se hizo dentro de un marco conceptual relativo al currículo, que distinguía e interrelacionaba el currículo diseñado (el currículo oficial), el currículo realmente enseñado (en la escuela) y el currículo aprendido (por los estudiantes en cuestión). Por esa razón, su implementación fue algo dispendiosa, pues implicaba el análisis en cada país de los currículos oficiales en matemáticas y ciencias así como de los alcances de la docencia concreta de las escuelas participantes.

Las pruebas, igualmente, requirieron de una aplicación cuidadosa; en ellas no se buscaba indagar solamente por los conocimientos elaborados por los alumnos, sino también por las habilidades desarrolladas, lo que implicaba, en matemáticas, centrarse en el enfoque de solución de problemas y, en ciencias, tener la posibilidad de manipular algunos artefactos para poder responder a las preguntas planteadas.

Las características de la prueba permitieron también que se publicaran no solamente los resultados del desempeño en la misma, comparables internacionalmente, sino que se indagara adicionalmente, de una manera comparativa también, por factores presumiblemente asociados al logro académico, tales como calidad de la docencia (el currículo enseñado) y políticas educativas (el currículo oficial).

4. LLECE

El Laboratorio Latinoamericano de Estudios sobre la Calidad de la Educación -LLECE- fue creado a comienzos de los noventa con el aval de los ministerios de educación del subcontinente y con el apoyo financiero y técnico de la UNESCO. En 1997 realizó una evaluación masiva de la calidad de la educación en los principales países del área. La participación de Colombia fue coordinada directamente por el Ministerio de Educación.

Para la evaluación se aplicaron pruebas únicas de matemáticas y de lenguaje a los alumnos de los grados tercero y cuarto de muestras de escuelas de los países participantes. El diseño de las muestras hizo que los resultados no sólo fueran representativos del conjunto de alumnos de esos grados en los países participantes, sino

también del contexto -megaciudad, urbano o rural- en que se ubicaban las escuelas y del carácter de las mismas -oficial o privado-.

Simultáneamente se levantó información sobre el contexto socioeconómico y familiar de los alumnos participantes, así como sobre aspectos importantes relacionados con el plantel y el aula de clase, tales como ambiente de aprendizaje, clima escolar y clima de aula. Estos datos permitieron que en los informes no sólo se diera cuenta de los resultados académicos por país, totales y discriminados según los estratos mencionados, sino que se extrajeran algunas pistas en cuanto al comportamiento en la región de algunos factores asociados al logro en el aprendizaje.

Como análisis complementario, el laboratorio lideró estudios cualitativos -en forma de análisis de caso- de muestras de escuelas de los países participantes, en las que se obtuvieron los promedios más elevados en los puntajes de las respuestas a las pruebas. Dichos estudios, publicados en un volumen complementario, buscan ayudar a la comprensión, no ya del comportamiento individual de las variables asociadas con el logro, sino del conjunto de los factores y de su mezcla específica que implicaron la producción de esos resultados.

5. PISA

El Programa Internacional de Evaluación de Estudiantes -PISA, por su sigla en inglés-, instituto creado por la Organización para la Cooperación y el Desarrollo Económico, OCDE, viene aplicando desde el 2000 el proyecto más ambicioso de evaluación y comparación internacional de los resultados del aprendizaje. Si bien Colombia no ha participado, es necesario tener en cuenta

este programa, que sin lugar a dudas se está convirtiendo ya en marco de referencia obligado para cualquier esfuerzo que emprendan países o conjunto de países en esa dirección.

Aunque inicialmente se tenía prevista la participación únicamente de los países industrializados (miembros de la OCDE), desde la primera administración de las pruebas en el 2000 participaron países no miembros; en América Latina, fuera de México (país miembro de la OCDE), participaron Argentina, Brasil, Chile y Perú. En 2003, sin embargo, solamente participaron México, Brasil y Uruguay.

El trabajo emprendido por PISA tiene algunas características que le son peculiares y que explican en parte su aporte. La primera de ellas tiene que ver con la intención de continuidad, que permitirá con el tiempo no sólo establecer comparaciones transversales o *sincrónicas* entre países, culturas y situaciones, sino también longitudinales o *diacrónicas*.

En efecto, la evaluación se plantea en las áreas de matemáticas, ciencias y lenguaje, y se hará cada tres años; en cada aplicación se evalúan las tres áreas, pero se hace énfasis, por turnos, en cada una de ellas: la primera evaluación, cuyos resultados ya están publicados, fue en 2000, con énfasis en lectura; en 2003, cuyos resultados comenzaron a publicarse a partir de 2004, el énfasis fue en matemáticas; en 2006 será en ciencias, para regresar de nuevo a lenguaje en 2009, y así sucesivamente. El énfasis consiste en ampliar cada vez la batería de preguntas en el área correspondiente, lo cual permite refinar un poco los niveles de desempeño (las escalas) y detectar factores asociados más explícitamente con el logro en esa área.

A partir de la aplicación de 2003 se produce una innovación. Además de la evaluación en las tres áreas básicas (lenguaje, matemáticas y ciencias), se cuenta con una serie de ítemes específicos que evalúan la capacidad de solucionar otro tipo de problemas. Hasta ahora se había indagado el desarrollo de esta capacidad en los alumnos, pero desde la perspectiva de cada una de las tres áreas específicas; ahora se plantean, adicionalmente, problemas más cercanos a la vida real (en los ámbitos laboral, académico o personal), cuya solución implica la activación de recursos y estrategias cognitivas de diversas áreas. Es lo que llamaríamos en nuestra "jerga" curricular "problemas trasnversales". Los resultados de la evaluación de dichas habilidades, y su correlación con las habilidades más específicas por área, fueron publicados en un informe específico: *Problem Solving for Tomorrow's World* (2004).

Una segunda característica importante es el diseño mismo de las pruebas. No fueron hechas para que se aplicaran a alumnos de un determinado grado escolar, sino a muestras representativas de jóvenes de 15 años de edad. Con ello se evita el eterno problema de la compatibilidad o equivalencia entre los sistemas escolares de los diferentes países y simplemente se indaga, con jóvenes que se encuentran ante el umbral de la edad madura, más o menos hacia el final de la adolescencia y también aproximadamente de la escolaridad obligatoria de casi todos los países, qué tan preparados están para asumir los retos de la edad adulta de una manera autónoma y

competente, en la sociedad postmoderna del siglo XXI. El diseño de la prueba gira entonces en torno a las competencias para los diversos escenarios de la edad adulta, independientemente de que las hayan desarrollado en la escuela o fuera de ella; en este sentido, los ítemes no indagan por acumulación o suficiencia en saberes académicos, sino por la capacidad de utilizar las herramientas que proporciona una alfabetización básica⁹ en lectura, matemáticas y ciencias para desempeñarse en los más diversos escenarios de la sociedad contemporánea. La indagación por factores asociados va más allá de la usual pregunta por las características de la escuela y por el contexto socioeconómico de los hogares de origen: se indaga también por el ambiente cultural (de la escuela y de la familia) en el cual se desenvuelven los encuestados, por las actitudes de éstos ante el aprendizaje y ante la lectura (en el caso de las pruebas de 2000).

Una tercera característica es la suficiencia, en recursos técnicos, financieros y humanos de la mejor calidad disponible en el momento, con que cuenta todo el proceso de diseño, elaboración, aplicación, sistematización, análisis y difusión de estas pruebas. Los informes finales de las aplicaciones de 2000 y 2003 presentan resultados y análisis de factores concomitantes al aprendizaje, respaldados con las últimas técnicas de medición e inferencia estadística, pero expresados en un lenguaje que no se empantana en los tecnicismos con que suelen empantarse los informes de este tipo.

⁹ PISA se refiere a las competencias como "literacy skills for the world of tomorrow", en el sentido de que no se trata de cualquier tipo de competencias, sino de aquellas que se desprenden del aprendizaje o alfabetización (sistemática o no) en conocimientos organizados. Por eso se puede hablar de una alfabetización matemática (mathematical literacy), científica (science literacy) y lectora propiamente tal (reading literacy).

El Cuadro 2 resume las características principales de los cinco sistemas de evaluación citados.

B. Los resultados principales

1. ICFES: la experiencia acumulada

Tanto por la diversidad de áreas evaluadas -prácticamente todas las asignaturas escolares- como por su frecuencia y continuidad, las pruebas del ICFES serían el instrumento más adecuado para elaborar series con resultados promedio nacionales, departamentales e inclusive por tipos de colegios, para observar al menos parcialmente el comportamiento de la calidad de la educación en las últimas tres décadas. Desafortunadamente

esto no es posible, debido a que los datos consolidados que dispone el ICFES se refieren a resultados estandarizados cuya dispersión se establece en torno a medias arbitrarias que no son comparables año a año, y es prácticamente imposible "desestandarizar" los promedios para regresar a los resultados originales. Es decir, podemos observar cómo se comportan los promedios departamentales referidos al promedio nacional, cómo sube o baja una región del país, pero teniendo como referencia el promedio nacional para cada año, sin poder afirmar si hay evolución o no, ni en qué sentido, de un año a otro.

A partir de los noventa el ICFES agrupa los colegios en siete rangos ordinales, de acuerdo con

Cuadro 2
CALIDAD DE EDUCACIÓN - EVALUACIONES MASIVAS DE RESULTADOS

	Programación	Alcance	Cobertura	Estudios de factores asociados al logro detectado	Estudios cualitativos
ICFES	Bianual, a partir de 1968	Grado 11 Todas las áreas académicas	Colombia En incremento, acercándose al 70% al final de los noventa	Setentas: Velandia Noventas: Sarmiento y Parra	
SABER	1. 1992-1994 2. 1997-1999 3. 2002-2003 4. 2005	Grados 3, 5, 7, 9 Matemáticas, lenguaje [una vez cívica y ciencias] Grados 5,9 Matemáticas, lenguaje, ciencias, sociales, competencias ciudadanas	Colombia Primero muestra, luego censo	MEN 1993-94 (grados 3,5) Javeriana (J. Tenjo) 1997-1999 Misión Social Bogotá 1998-1999 C. Uribe Bogotá 1998-1999 Cepeda, Girardot/ Bogotá 1990-2001	1998-1999: estudios de caso realizados en Bogotá por CORPOEDUCACIÓN
TIMSS	1994-1995	13 años, grados 7,8 Matemáticas, ciencias	Internacional (con Colombia) muestra	Los documentos mismos del estudio estudio: análisis de elementos curriculares (currículo diseñado y desarrollado): políticas, escuelas, docentes	
LLECE	1997	Grados 3,4 (misma prueba) Matemáticas, lenguaje	Latinoamérica (con Colombia) muestra	Los documentos mismos del estudio: contexto familiar y escolar	Estudio auspiciado por el mismo LLECE, selección internacional de escuelas con resultados notables
PISA	Cada 3 años, a partir de 2000	Jóvenes de 15 años (sin importar escolaridad) Lectura, matemáticas, ciencias	Internacional (Colombia a partir de 2006) muestra	Los documentos mismos del estudio, clasificación más completa, (inclusive factores actitudinales)	

los promedios de los resultados obtenidos por sus alumnos en las pruebas: *muy superior, superior, alto, medio, bajo, inferior y muy inferior*. Al comparar los años 1998 y 2000, o sea con las pruebas antiguas y con las nuevas, encontramos los porcentajes de colegios que mantienen o cambian su rango, que se presentan en el Cuadro 3.

Como puede observarse, la gran mayoría de los colegios conserva su rango o lo cambia solamente por el inmediatamente inferior o superior. En efecto, no es de esperarse un cambio radical en la calidad (ni por tanto de rangos en los colegios) en tan corto tiempo. La gran coincidencia en los rangos significa entonces que las pruebas siguen midiendo lo mismo, que el cambio tan

radical que se anunció con la aplicación de las nuevas pruebas no lo fue tanto y que es por tanto posible establecer comparaciones longitudinales, al menos desde el punto de vista de los rangos, antes y después de la reforma de las pruebas.

Organizando la información por departamentos, encontramos el siguiente orden descendente en los promedios de sus puntajes, para tres cortes en el tiempo con intervalos de una década (Cuadro 4).

Encontramos en la posición relativa de los departamentos varias tendencias que se mantienen a lo largo de las dos décadas: Bogotá, Santander y Boyacá (en ese orden) ocupan los primeros

Cuadro 3
ICFES - EXÁMENES DE ESTADO, 1998 VERSUS 2000
(Porcentajes de colegios que mantienen o cambian el rango)

1998	2000							Total
	Muy superior	Superior	Alto	Medio	Bajo	Inferior	Muy inferior	
Muy superior	64	36	0	0	0	0	0	100
	82	14	0	0	0	0	0	
Superior	8	78	11	2	0	0	0	100
	17	46	3	0	0	0	0	
Alto	0	23	58	18	1	0	0	100
	1	38	46	5	0	0	0	
Medio	0	1	20	59	19	1	0	100
	0	2	46	50	15	3	50	
Bajo	0	0	2	36	58	5	0	100
	0	0	5	43	62	19	0	
Inferior	0	0	0	4	53	43	0	100
	0	0	0	2	23	72	50	
Muy inferior	0	0	0	0	12	88	0	100
	0	0	0	0	0	6	0	
Total	100	100	100	100	100	100	100	

Fuente: ICFES. Cálculos propios.

Cuadro 4

ICFES - EXÁMENES DE ESTADO

(Rangos (en orden descendente) de los promedios departamentales)

Rango	1983	1993	2003
1	Bogotá	Bogotá	Bogotá
2	Santander	Santander	Santander
3	Boyacá	Boyacá	Boyacá
4	Meta	Nariño	Cundinamarca
5	Risaralda	Cauca	Caldas
6	Cauca	Valle del Cauca	Meta
7	Valle del cauca	Huila	Huila
8	Huila	Meta	Risaralda
9	Caldas	Cundinamarca	Casanare
10	Casanare	Risaralda	Nariño
11	Cundinamarca	Norte de Santander	Valle del Cauca
12	Tolima	Caldas	Quindío
13	Norte de santander	Quindío	Antioquia
14	Nariño	Tolima	Norte de Santander
15	Antioquia	Casanare	Tolima
16	Quindío	Putumayo	Sucre
17	Caquetá	Bolívar	Putumayo
18	Putumayo	Arauca	Arauca
19	Bolívar	Caquetá	Cesar
20	Arauca	Antioquia	Atlántico
21	Atlántico	Atlántico	Cauca
22	Córdoba	Sucre	Bolívar
23	Sucre	Córdoba	Caquetá
24	Magdalena	Cesar	Córdoba
25	Cesar	Magdalena	La Guajira
26	La guajira	La Guajira	Magdalena
27	Chocó	Chocó	Chocó

Nota: no se incluyeron los departamentos en los que en 2003 se presentaron menos de 1000 alumnos a los exámenes: Amazonas, Guainía, Guaviare, San Andrés, Vaupés y Vichada.

Fuente: ICFES. Cálculos propios.

puestos, mientras que Chocó se mantiene siempre en el último. Inmediatamente por encima de éste, y con pocas variaciones, se mantienen además los siete departamentos de la costa (Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena y Sucre) con Arauca, Caquetá y Putumayo.

Hay hechos notorios, como los rangos relativamente altos ocupados por Meta (4, 8 y 6) y por Huila (8, 7 y 6), frente a los rangos relativamente

bajos de Antioquia (15, 20 y 13). Estos resultados indican, si se toma también en consideración el alto puesto ocupado de manera constante por Boyacá, que los resultados medibles de la calidad de la educación no necesariamente tienen que depender de la riqueza o pobreza de la región. En el caso de Antioquia tendríamos que corroborar el hecho, registrado igualmente en otros sectores sociales, de un departamento muy centralizado, donde una producción académica

sobresaliente que se registra en la capital no alcanza a contrarrestar en el promedio el peso que tiene una periferia (en la capital y en el resto del departamento) altamente deprimida.

Variaciones importantes se registraron en Nariño, que sube del rango 14 al 4, para bajar posteriormente al 10; pero, sobre todo, en Cauca, que de ocupar puestos relativamente altos (6 y 5) registra un bajón al puesto 21. Este último caso debería analizarse a la luz del empobrecimiento y la descomposición social que ha vivido el departamento en la última década.

Fuera de esta variación relativa en los rangos ocupados por los departamentos, no es mucho más lo que se puede decir de la evolución en el tiempo de la calidad de la educación, según las pruebas del ICFES. El análisis se puede refinar un poco, sin embargo, si se toman sincrónicamente los resultados departamentales para un mismo año. En el último disponible (2003), estas son las medias en los puntajes departamentales, ordenadas en orden descendente (Cuadro 5).

Si bien es grande la tentación de sacar apresuradamente conclusiones sobre las posiciones y la variación en los rangos por departamentos, como lo presentamos anteriormente, esas conclusiones hay que relativizarlas cuantificando las distancias entre los rangos, como se acaba de presentar. En efecto, la única diferencia entre medias superior a una desviación estándar nacional (3,80) es la registrada entre el primero (Bogotá) y el último departamento (Chocó). Lo que nos lleva a concluir que, a pesar de todo lo dicho anteriormente, la diferencia entre departamentos no es tan grande como parecería insinuarlo un "ranking" puramente cualitativo, y que las enormes diferencias en la calidad de

un sistema educativo tan estratificado como el colombiano habría que buscarlas más bien al interior de los departamentos, entre estratos y, posiblemente también, entre planteles.

Si ordenamos los departamentos por el grado relativo de la homogeneidad en sus promedios,

Cuadro 5

ICFES 2003: PROMEDIOS DEPARTAMENTALES EN LOS EXÁMENES DE ESTADO

Departamento	Alumnos	Desviación	Media
Bogotá	82.449	4,03	41,32
Putumayo	1.683	3,49	39,18
Santander	21.395	4,04	40,39
San Andrés	588	3,76	39,11
Boyacá	15.535	3,60	40,07
Arauca	1.531	3,20	39,06
Cundinamarca	25.590	3,42	39,95
Cesar	7.789	3,63	39,04
Caldas	9.960	3,70	39,88
Atlántico	25.030	3,86	38,97
Meta	7.086	3,54	39,87
Cauca	8.372	3,45	38,97
Huila	8.388	3,53	39,86
Vaupés	132	2,49	38,81
Guainía	124	3,37	39,72
Vichada	258	2,79	38,73
Risaralda	8.246	3,56	39,65
Bolívar	18.417	3,64	38,69
Casanare	2.588	3,36	39,64
Caquetá	2.307	3,10	38,65
Nariño	12.491	3,45	39,61
Guaviare	434	2,83	38,63
Valle del Cauca	36.947	3,55	39,52
Córdoba	13.133	3,32	38,36
Quindío	5.500	3,49	39,48
La Guajira	4.636	3,37	38,26
Antioquia	55.199	3,81	39,46
Amazonas	586	2,83	38,09
Norte de Santander	12.830	3,56	39,43
Magdalena	10.113	3,14	38,01
Tolima	3.307	3,43	39,42
Chocó	3.205	2,71	36,92
Sucre	7.519	3,43	39,34
Total	423.368	3,80	39,76

Fuente: ICFES. Cálculos propios.

medidos por el coeficiente de variación¹⁰, obtenemos los resultados presentados en el Cuadro 6.

En el ordenamiento descendiente del Cuadro 6 -de relativamente más heterogéneos a menos

Cuadro 6

ICFES 2003: COEFICIENTE DE VARIACIÓN EN EXÁMENES DE ESTADO POR DEPARTAMENTO

Departamento	Coefficiente
Santander	0,100
Risaralda	0,090
Cundinamarca	0,086
Atlántico	0,099
Putumayo	0,089
Guainía	0,085
Bogotá	0,098
Meta	0,089
Casanare	0,085
Antioquia	0,097
Huila	0,089
Magdalena	0,083
San Andrés	0,096
Cauca	0,089
Arauca	0,082
Bolívar	0,094
Quindío	0,088
Caquetá	0,080
Cesar	0,093
La Guajira	0,088
Amazonas	0,074
Caldas	0,093
Sucre	0,087
Chocó	0,073
Norte de Santander	0,090
Nariño	0,087
Guaviare	0,073
Boyacá	0,090
Tolima	0,087
Vichada	0,072
Valle del Cauca	0,090
Córdoba	0,087
Vaupés	0,064

Fuente: ICFES. Cálculos propios.

heterogéneos- esperaríamos que aparecieran en primer lugar los departamentos con mayor número de alumnos examinados. Ello ocurre así en departamentos como Antioquia y Bogotá, departamentos poblados y con una desviación estándar relativamente alta. En el polo opuesto (los más homogéneos), esperaríamos igualmente encontrar departamentos menos poblados, como en efecto ocurre con casi todos los nuevos departamentos. Sin embargo, se dan casos en los que la realidad no está de acuerdo con estas expectativas: Cundinamarca es un departamento con un número relativamente alto de examinados y un coeficiente de variación relativamente bajo (homogenización por lo alto); caso contrario ocurre con San Andrés: a pesar de su escaso número de alumnos (menos de mil) presenta una de las variaciones más altas en sus resultados.

Por regla general, los promedios más altos suelen ser los más dispersos, como en los casos de Bogotá y Santander, y los más bajos, más homogéneos, como en el caso de Chocó (homogenización por lo bajo). La gran dispersión relativa de algunos departamentos con promedios altos nos está indicando que, además de las marcaciones altas de un número significativamente elevado de alumnos, existen también marcaciones muy por debajo de los promedios (Bogotá y Santander). En los departamentos con promedios bajos y variación relativamente amplia (Atlántico y Antioquia) hay que concluir igualmente que se deben dar marcaciones muy por encima de los promedios.

Un análisis ulterior, y que puede arrojar resultados todavía más interesantes, es establecer, mediante una descomposición de la varianza (similar a la realizada por PISA para los resultados del 2000), qué porcentaje de ésta se da entre los

¹⁰ Desviación estándar dividida por la media.

establecimientos o colegios, y qué tanto se da al interior de los mismos, por departamento. Los resultados de dicho ejercicio, para 1993, son los que aparecen en el Gráfico 1¹¹.

El rango de la variación entre colegios (entre 3,3 para Guainía y 46,6 para Bogotá), es muy inferior al detectado en PISA 2000, para el análisis entre los países participantes (entre 7,0 y 76,0): allá se analizaron países con un sistema educativo muy estandarizado y homogéneo (los países escandinavos, por ejemplo), frente a países muy estratificados, bien sea por el tipo de colegios (Alemania y Bélgica) o por los desniveles socioeconómicos de los mismos (en general, los países no industrializados participantes). En Colombia, por el contrario, estamos analizando departamentos con comportamientos relativamente similares (salvo unos pocos casos extremos), como se anotó anteriormente.

Este análisis de las variaciones en los puntajes entre los alumnos, descompuesta en la variación entre los colegios y al interior de los mismos, proporciona pistas importantes tanto para el análisis de los factores asociados como para las posibles medidas en política educativa: quiere decir que el estar en uno u otro colegio explica una buena parte de las diferencias en los resultados (*variación entre*, que a nivel del país puede estar alrededor del 40%), y no sólo las diferencias de origen, de empeño en el estudio o de rendimiento observadas entre los alumnos individuales (*variación al interior de*). En términos de política esto quiere

decir que es probable que el mejoramiento en la calidad se logre no solamente incrementando las posibilidades individuales de aprendizaje en los alumnos, sino también mediante acciones orientadas al mejoramiento de las instituciones donde éstos aprenden.

Hay un detalle importante en los resultados de las pruebas del ICFES. Sarmiento y Parra (2000) detectaron, para las aplicaciones de 1996 y 1999, una correlación alta entre los resultados en lenguaje y en ciencias naturales¹², superior todavía a la encontrada por Velandia (1978) para 1974, lo cual coincide con lo observado en PISA 2000. El detalle es importante, puesto que en el país el rendimiento escolar en matemáticas, o en disciplinas con un alto contenido matemático (como las ciencias), ha sido tradicionalmente el criterio clave con el que los alumnos eligen la carrera a estudiar: un buen rendimiento en matemáticas o ciencias los lleva generalmente a seleccionar ingenierías, carreras técnicas o estudios en ciencias básicas. Un bajo rendimiento, por el contrario, hace que prefieran el derecho, la filosofía o las ciencias sociales. Lo que están mostrando las correlaciones calculadas, es que probablemente un bajo desempeño en matemáticas esté asociado con un mal desempeño también en otras áreas claves del saber académico y, en última instancia, con fallas en el razonamiento lógico en general (verbalismo excesivo con poca coherencia de fondo), lo cual representa un riesgo para el ejercicio profesional, no solamente en las ingenierías sino también en las ciencias sociales y en el derecho.

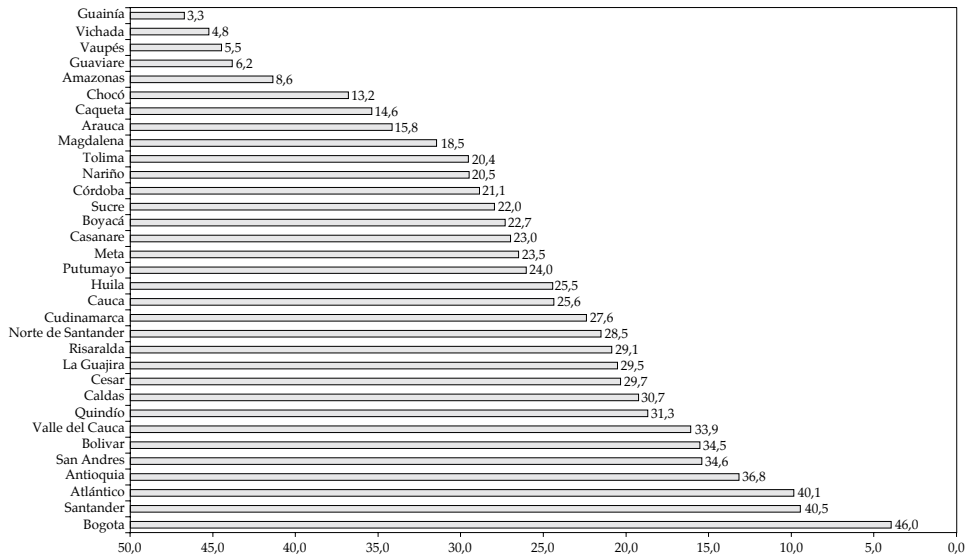
¹¹ En el Gráfico 1 que sigue, los porcentajes están expresados en función de la varianza total del país y no de la de cada departamento. Por eso la suma de los valores para los segmentos de arriba y abajo de cada departamento no necesariamente suman 100. Este procedimiento se realizó siguiendo a PISA 2000 (2003).

¹² Cálculos realizados por Alfredo Sarmiento y Carlos Parra sobre los datos de las pruebas del ICFES en la Misión Social (hoy Programa Nacional de Desarrollo Humano), Departamento Nacional de Planeación, que no fueron publicados.

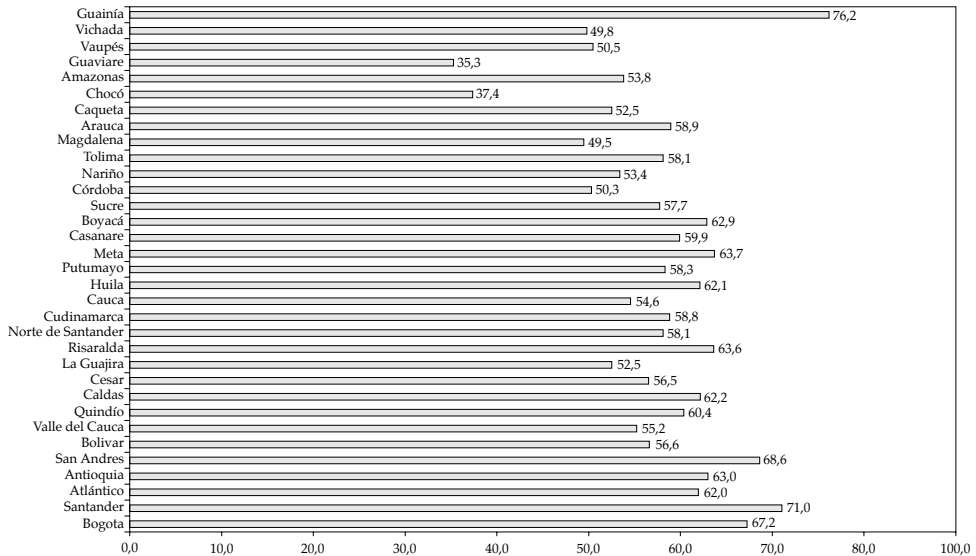
Gráfico 1

ICFES 2003: ANÁLISIS DE DESCOMPOSICIÓN DE LA VARIANZA

Variación entre colegios



Variación al interior de los colegios



Fuente: ICFES. Cálculos propios.

2. SABER, TIMSS, LLECE: la aplicación de los aprendizajes

Las otras tres mediciones -SABER, TIMSS, LLECE- son producto del auge de la evaluación de los resultados del aprendizaje académico que comenzó a tener énfasis a partir de los noventa. Todas ellas tienen en común que cualquiera sea el tipo de medición empleado, no importa si la evaluación es interna o se compara a Colombia con otros países, siempre se llega a la misma conclusión: los niños y jóvenes colombianos están mal equipados con las competencias necesarias, medidas desde los resultados académicos en las áreas de matemáticas, lenguaje y ciencias, para enfrentar satisfactoriamente los retos que a la nación y al individuo les plantea el siglo XXI.

En general, las deficiencias más grandes que los niños muestran en las pruebas son los resultados de un aprendizaje mecánico, que se traduce muchas veces en: a) la aplicación acertada de algoritmos sin que ello conlleve la utilización de herramientas del pensamiento matemático para enfrentar correctamente problemas de la vida cotidiana, o en b) una lectura textual en cierto sentido "correcta", desde el punto de vista de la descodificación de los signos, pero a partir de la cual no se logra reconstruir un sentido, ubicar información útil o pertinente en una situación dada, y mucho menos deducir conclusiones o hacer inferencias no consignadas explícita o literalmente en el texto.

Ambas conclusiones son respaldadas por los documentos "oficiales" de análisis de los resultados. En matemáticas, por ejemplo, el Ministerio de Educación afirma:

"... los porcentajes tan bajos de población que acceden al nivel de resolución de problemas

demandan análisis amplios y correctivos inmediatos" (Sistema Nacional de Evaluación de la Educación 1997c:47).

Por su parte, el ICFES enumera, entre otros, los siguientes problemas en la comprensión de textos: no se reconocen diferentes tipos de textos, no se reconocen las intenciones de la comunicación, hay dificultad para establecer relaciones entre los contenidos de diferentes textos, hay dificultades en la lectura crítica (Pérez 2003b).

Como se indicó arriba, el SABER ha medido los resultados del aprendizaje en las áreas de matemáticas y lenguaje, en cuatro tandas de evaluación, partiendo de comienzos de la década de los noventa hasta el año 2005. En las dos primeras tandas, los datos oficiales publicados por el Ministerio de Educación y por el ICFES se expresan en términos del porcentaje de alumnos que, para cada grado y área, alcanzó al menos uno de los tres niveles de resultados que fueron definidos previamente: B, C y D. Estos tres niveles, en el área de matemáticas, corresponden a la capacidad de resolver problemas que contienen:

- relaciones directas en problemas rutinarios (nivel B),
- relaciones directas en problemas no rutinarios simples (nivel B) o
- relaciones no directas en problemas no rutinarios simples (nivel C).

En lenguaje, por el contrario, aunque se continúa trabajando con tres niveles de resultados, se mencionan sin embargo cuatro tipos de competencias, que parecerían responder a niveles progresivos de logro:

- Lectura literal
- Lectura inferencial
- Lectura crítica
- Lectura intertextual

Los resultados para las dos primeras tandas de exámenes, en términos del porcentaje de alumnos que superan cada nivel de logro, son los siguientes en el Cuadro 7.

Al yuxtaponer de esta manera los resultados de las dos primeras tandas de pruebas, salta a la vista de inmediato que en prácticamente todos los niveles para todos los grados examinados en

las dos áreas los logros alcanzados en la segunda tanda están por debajo de los de la primera, en ocasiones con diferencias muy notorias, como es el caso de las matemáticas para el nivel D en todos los grados. Es poco probable que en el transcurso de cuatro a cinco años se haya producido una modificación tan drástica en los resultados.

En la tercera tanda la situación se complica un poco más, puesto que para secundaria se mencionan otros dos niveles adicionales (E, F) y dejan de aparecer porcentajes de alumnos que sólo han llegado al nivel B (Cuadro 8).

Cuadro 7

RESULTADOS DE LAS PRUEBAS SABER, 1992-1994 Y 1997-1998

	Nivel	Matemáticas				Lenguaje			
		Grado 3	Grado 5	Grado 7	Grado 9	Grado 3	Grado 5	Grado 7	Grado 9
1992-1994	B	95,7	94,4	91,6	94,3	97,0	92,3	79,9	98,3
	C	76,6	58,4	65,7	73,3	80,3	63,8	50,2	76,8
	D	36,7	23,2	25,4	36,4	36,7	20,0	24,7	25,7
1997-1998	B	83,1	82,6	84,3	64,09	5,9	75,7	88,4	92,5
	C	54,8	52,9	34,0	20,27	9,4	33,8	53,1	65,2
	D	17,4	21,5	2,8	1,8	43,7	4,6	11,8	19,6

Fuente: Sistema Nacional de Evaluación de la Educación (1997a) y www.icfes.gov.co (enero 2003).

Cuadro 8

RESULTADOS DE LAS PRUEBAS SABER, 2002-2003

	Nivel	Matemáticas				Lenguaje			
		Grado 3	Grado 5	Grado 7	Grado 9	Grado 3	Grado 5	Grado 7	Grado 9
2002-2003	B	79,3	89,8	-	-	85,8	91,9	-	-
	C	41,4	55,3	68,8	73,5	68,1	84,5	95,0	93,7
	D	19,6	23,6	28,9	32,5	35,6	52,5	79,6	73,4
	E	-	-	7,8	8,8	-	-	41,1	36,3
	F	-	-	2,0	1,3	-	-	3,5	4,0

Fuente: Ministerio de Educación (2002).

De la comparación de los resultados para las tres tandas se puede concluir que las pruebas del SABER pueden ser un instrumento útil para establecer comparaciones entre regiones y tipos de planteles a un nivel horizontal, toda vez que se ha pasado, primero, a una muestra mucho más representativa (en la segunda tanda el total de examinados fue más del doble de la primera) y, luego, a una evaluación casi censal (tercera y cuarta tandas). Segundo, gracias a que se han refinado, además, tanto los criterios de elaboración de pruebas e ítemes de las pruebas, como la medición e inferencia estadística. Sin embargo, la comparación longitudinal a lo largo de la década, como permiten suponerlo los datos así yuxtapuestos de las tres tandas para las cuales se tienen resultados, parece todavía muy poco confiable, en el sentido que permita distinguir tendencias en cuanto a la evolución de la calidad de la educación del país¹³.

Las otras dos mediciones pueden ser importantes para ubicar a Colombia en el contexto internacional. Los resultados, sin embargo, no son muy halagadores. En el TIMSS participaron algo más de 40 países; tanto en matemáticas como en ciencias. Colombia ocupó siempre el penúltimo lugar, superando solamente a Sudáfrica. Además del puesto tan bajo, la diferencia con los países ubicados en las posiciones superiores es notable: en ciencias, los mejores desempeños colombianos apenas son comparables con los 10 percentiles más bajos de Singapur, en matemáticas con los 5 percentiles más bajos del mismo país¹⁴.

Lo novedoso del TIMSS es su concepción curricular a tres niveles, como se explicó arriba. Lo curioso es que, mientras se descubre en el estudio que el currículo oficial colombiano está a la altura de los mejores, el currículo realmente aprendido por los estudiantes arroja los resultados desastrosos ya mencionados. Ello pone en serias dudas la efectividad de las revisiones periódicas al currículo y los planes de estudios oficiales, que por decreto se producen periódicamente en el país.

El LLECE permite situar la calidad de la educación, al menos la que comienzan a recibir niños de tercero y cuarto grado, en un contexto que para muchos es más comparable y, por tanto, más "realista": el contexto latinoamericano. Aunque no todos los países participaron en el proyecto y algunos de los que sí lo hicieron o se desviaron de la metodología prevista o no permitieron publicar los resultados, los datos disponibles permiten hacer comparaciones entre once países de la región. Los resultados más sobresalientes del estudio, publicados ya por el Laboratorio, son cuatro, uno de tipo general y tres específicos por países.

El primer resultado es que los puntajes promedio de todos los países -si se exceptúa Cuba- son bastante comparables (ver Gráfico 2). Prácticamente la mediana de ninguno está a más de una desviación estándar del promedio estandarizado regional, y los rangos para el cincuenta por ciento de los entrevistados, excluidos los cuartiles

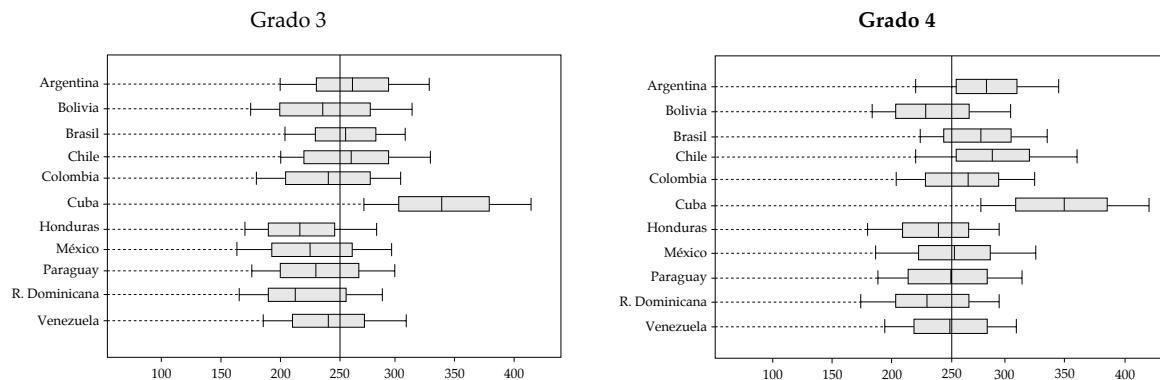
¹³ A comienzos de abril de 2006 el ICFES no había divulgado todavía los resultados de la cuarta tanda de aplicación de las pruebas del SABER. Informes preliminares permiten deducir que se habrían llevado a cabo algunas modificaciones adicionales en la definición de los diversos niveles de logro.

¹⁴ Véanse las cuatro publicaciones del TIMSS/Colombia citadas en la bibliografía.

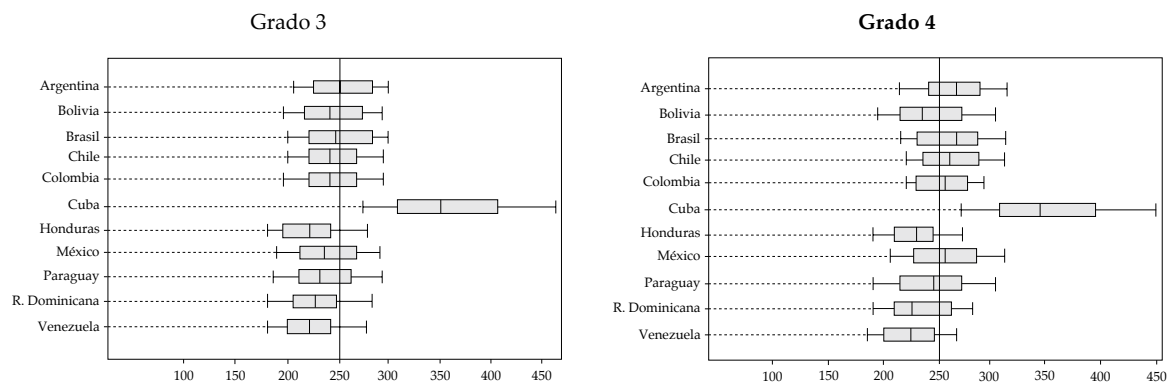
Gráfico 2

LLECE - RESULTADOS GENERALES, MEDIANA, CUARTILES 25% Y 75% Y DECILES 10% Y 90%

A. Lenguaje



B. Matemáticas



Fuente: LLECE 1998.

extremos, se sobreponen en buena parte, como puede verse en los cuatro gráficos que siguen. La tendencia es mucho más marcada en matemáticas que en lenguaje, donde se observan mayores desviaciones del comportamiento promedio.

El segundo resultado complementa el anterior, y es el puesto sobresaliente que ocupan los resultados de Cuba frente al de todos los demás países latinoamericanos. A excepción de un país

en un caso concreto (lenguaje, grado 4) el cuartil inferior de Cuba (excluidos los dos cuartiles extremos) está siempre por encima del cuartil superior de todos los demás países.

Una característica peculiar de Cuba, que también puede verse en las gráficas, más notoria en matemáticas que en lenguaje, es la variación tan alta en los resultados, superior a la de los otros países. PISA encontró todo lo contrario: los países

con los puntajes más altos son también aquellos donde la dispersión de los mismos es la más baja. En efecto, en uno de los informes indica que la excelencia en calidad no tiene por qué traducirse en mayor inequidad. En América Latina, por el contrario, parece ocurrir algo diferente: las variaciones más bajas se dan en los países con los menores resultados promedio; lo mismo ocurre en el interior de Colombia, cuando se comparan las variaciones entre departamentos obtenidas en los puntajes del ICFES¹⁵.

Los otros dos resultados sobresalientes tienen que ver con una de las estratificaciones en que organizó el LLECE los resultados. A fin de comprobar más al detalle el efecto de la megaciudad en la educación, la acostumbrada clasificación urbano/rural fue convertida en tres estratos: megaciudad (más de un millón de habitantes), urbano (el resto) y rural. Como era de esperarse, los resultados de la megaciudad fueron superiores a los urbanos y éstos a los rurales en prácticamente todos los países, con dos excepciones notables que analizan los documentos:

- En Chile los puntajes urbanos promedio fueron superiores a los de megaciudad. El documento explica esto como una posible consecuencia de las políticas de regionalización educativa en ese país.
- En Colombia, los puntajes rurales promedio fueron superiores a los urbanos en lenguaje para tercer grado, y superiores tanto a urbanos como a megaciudad en matemáticas. Como el sector rural es en su totalidad oficial, el documento atribuye estos resultados al impacto del

modelo de Escuela Nueva, que se ha implantado en el país desde hace algunas décadas como modelo oficial para la escuela del sector rural. Los resultados rurales en matemáticas, inclusive, son superiores en tercero y cuarto grados a los resultados rurales de los demás países latinoamericanos, excepto Cuba.

Las diferencias en los resultados, sin embargo, pierden mucho en su significación cuando entran los puntajes de Cuba en la comparación: el rango de variación entre los cuartiles primero y tercero en resultados rurales de matemáticas en Colombia -esto es, la mitad de los encuestados excluyendo el 25 % de cada extremo- está entre 221 y 274 puntos para el tercer grado, mientras que en Cuba esa misma variación va de 300 a 400 puntos. En otras palabras, si bien Colombia es ligeramente superior al resto de los países de la región en los resultados de matemáticas de las escuelas rurales, los mejores niños rurales colombianos (excluidos los extremos) tienen puntajes inferiores a los peores niños rurales cubanos (excluidos también los extremos).

3. PISA: un punto necesario de referencia

Como se dijo anteriormente, PISA fue diseñado para aplicarse originalmente a los países industrializados (países miembros de la OECD). Sin embargo, ya desde la primera aplicación de 2000 participaron 15 países no miembros, entre ellos cinco latinoamericanos: México, Brasil, Argentina, Chile y Perú. En la segunda aplicación de 2003 dejaron de participar Argentina, Chile y Perú, se mantuvieron Brasil y México y entró a participar Uruguay.

¹⁵ Aunque no es siempre el caso. Entre los departamentos con mayor variación en los resultados se pueden encontrar tanto los de promedios más altos como los de promedios bajos.

La participación de Colombia está prevista a partir de la aplicación de 2006. Es probable (como puede inferirse a partir de la comparación latinoamericana lograda por LLECE) que nuestro país, en dicha aplicación, obtenga puntajes similares a los obtenidos por México y Brasil, únicos países latinoamericanos que han mantenido su participación en las dos tandas realizadas hasta el momento, cuyos resultados no han sido muy favorables: juntamente con Túnez, dichos países han continuado ocupando los últimos lugares en todos los promedios de puntajes y niveles de logro arrojados por las dos tandas, muy por debajo (con significación estadística) de los promedios logrados por los países industrializados.

No tiene sentido, sin embargo, que Colombia se retire después de unos resultados no muy favorables (como es justificadamente esperable) en la primera participación, tal como lo hicieron Perú, Argentina y Chile después de su participación en 2000. No se justifica la cuantiosa inversión de recursos financieros, humanos y de infraestructura que ello supone, si no se participa al menos en un ciclo completo de tres tandas de evaluaciones, con los énfasis sucesivos en ciencias, lenguaje y matemáticas, en 2006, 2009 y 2012. Y ello, por muchas razones:

En primer lugar, el objeto de dicha participación no debe ser el de la autopropaganda política, sino ponernos más bien bajo la lupa del mejor instrumento disponible en el momento de verificación de los resultados del esfuerzo educativo-o

al menos algunos de esos resultados-. En segundo lugar, la verificación internacional no sólo tiene por objeto comparar resultados entre diversas naciones en un corte de tiempo dado -lo cual es ya de por sí sano-, sino también mostrar tendencias a lo largo del tiempo¹⁶. Y, en tercer lugar, porque participar en el sistema, trabajar con la experiencia mejor y más consistente de medición masiva de la calidad, es una oportunidad única de aprendizaje, para mejorar lo que tenemos.

C. ICFES y SABER: ¿qué están midiendo?

Además de los resultados recogidos en las aplicaciones de las pruebas, y de los instrumentos estadísticos empleados para concluir qué tipo de generalizaciones se pueden hacer y qué comparaciones se pueden establecer entre dichos resultados¹⁷, es necesario analizar también las mismas pruebas, su presentación y el tipo de ítemes o preguntas que las componen (que son producto de consensos de determinado tipo de expertos y de sus expectativas de lo que esperan que los alumnos aprendan) para saber qué es lo que se está midiendo o verificando en el país. Este ejercicio debe centrarse fundamentalmente en las pruebas del SABER y las del ICFES, que son las únicas que se plantean una periodicidad, incipiente o consolidada, en su aplicación. En PISA, desafortunadamente, Colombia comienza a participar sólo hasta 2006, con lo cual los criterios de comparación de que disponemos son referidos al país mismo (salvo las participaciones episódicas en el TIMSS y el LLECE).

¹⁶ Tendencias que, como lo muestra el presente estudio, son difíciles captar con nuestro actual sistema de "evaluación de la calidad de la educación".

¹⁷ Resultados de tipo horizontal (diversas instancias, regiones, zonas o planteles en un momento dado) y longitudinal (tendencias a lo largo del tiempo).

Un análisis dispendioso del tipo de preguntas que se hacen y del saber que estas están en capacidad de detectar -es decir, la calibración del instrumento-es prácticamente imposible dado los alcances de este estudio y la magnitud del objeto de análisis, especialmente en lo concerniente a las pruebas del ICFES¹⁸. No obstante, una revisión parcial permite concluir que las pruebas¹⁹ (especialmente las del SABER) adolecen de muchos defectos de construcción y parecen estar más orientadas a detectar unas competencias demasiado escolares o académicas, que unas competencias más fundamentales para todo ciudadano, académico o no, traducidas en el uso adecuado de herramientas de conocimiento teórico y práctico con que probablemente se enfrentaría con éxito a los problemas y retos que le plantea el siglo XXI²⁰.

Hay dos problemas adicionales que se presentan en las pruebas. El primero tiene que ver con los niveles de logro, que constituyen una especie de grandes escalas para clasificar los resultados individuales y colectivos de los alumnos. En las

aplicaciones modernas de pruebas masivas de resultados del aprendizaje²¹ los niveles de logro se definen para cada área o tópico (equivalentes a los aspectos o líneas básicas de cada área en nuestros estándares curriculares). Un nivel de logro se refiere tanto al tipo de preguntas o ítems de un examen como al nivel de desarrollo del conocimiento en que se encuentra el alumno, en un tópico o línea específica de un área. Se supone que a medida que se incrementa el nivel de logro el alumno es capaz de realizar una serie de tareas cada vez más complejas inherentes a esa línea. El nivel de logro se detecta mediante la selección de una serie de tareas o ítems de prueba, para cuya realización el alumno debe activar las estrategias cognitivas y habilidades inherentes a dicho nivel. Se dice que un alumno ha llegado a determinando nivel de logro cuando tiene una probabilidad alta (usualmente superior al 50 por ciento) de responder correctamente los ítems seleccionados para dicho nivel (y, lógicamente, una probabilidad mayor de responder los ítems de los niveles inferiores).

¹⁸ Son dos aplicaciones anuales realizadas durante más de tres décadas. La aplicación de septiembre de 2003, para citar un ejemplo, constó de 22 pruebas distintas en 16 áreas académicas con un total de 590 ítems o preguntas.

¹⁹ En el estudio que sirvió de base para el presente documento se hace una aproximación a los contenidos de las pruebas y a la estructura de sus ítems con una muestra de las mismas, tanto del ICFES como del SABER, que no pretende ser estadísticamente representativa. Sin embargo, no se puede negar que el análisis del instrumento de medida es un elemento clave para saber lo que significan los resultados de la medición de la calidad de la educación, aunque sea tan sólo desde la perspectiva parcial del conocimiento desarrollado por los alumnos.

²⁰ No se trata de desconocer el valor de las competencias académicas, sino de señalar que no son el objetivo exclusivo de la educación básica. Sin embargo, la tendencia academicista parece imponerse, a pesar de los esfuerzos por señalar que se trata de verificar la construcción de competencias básicas cotidianas; basta con analizar los ítems que, a guisa de ejemplo, aporta el ICFES como material explicativo para la última aplicación de las pruebas (ICFES 2005: *passim*). Esta tendencia indujo, igualmente, a que el aparato conceptual sobre el cual se construyeron las pruebas en sus últimas versiones, así como sus correspondientes niveles de logro, parezca insinuar la peregrina idea de que en el desarrollo del conocimiento tiene lugar, *primero*, el manejo de los códigos y, *luego*, el de los conceptos que expresan dichos códigos.

²¹ En uno de los documentos producidos por PISA 2003 (*Learning for Tomorrow's World*), por ejemplo, después de explicar para cada una de las disciplinas evaluadas las diversas líneas temáticas o de contenidos, sus dimensiones, sus competencias fundamentales y las situaciones en que probablemente se ejercitarán (pp. 25-27), pasa a definir para el área básica de dicho año (matemáticas) unos niveles de logro, no solamente para el área en general, sino también para cada uno de sus tópicos o líneas de contenido (pp. 37-89).

Por otro lado, como los estándares académicos de que dispone el país a partir de 2004 se han especificado en términos de metas académicas (competencias) claras y verificables para cinco grupos de grados (1-3, 4-5, 6-7, 8-9, 10-11) en las cuatro áreas académicas básicas -matemáticas, lenguaje, sociales y ciencias-, con ello se definen automáticamente también cinco niveles escalonados de logro en esas mismas áreas. No se ve, en consecuencia, qué papel tengan que jugar unos niveles de logro paralelos establecidos por los expertos del ICFES -que no siempre coinciden con los de los estándares-.

Para que los resultados de las pruebas puedan ser útiles en los procesos de mejoramiento de la calidad escolar, se necesita, por tanto, que se conozcan no solamente los puntajes brutos por asignatura, sino los niveles de logro alcanzados por los diversos grupos de alumnos en cada uno de los tópicos de las asignaturas. Se necesita, adicionalmente, que en la difusión posterior de los resultados, no solamente se conozcan los ítems aplicados en las pruebas, sino los tópicos y los niveles de logro a que corresponden cada uno de esos ítems. Dicha claridad no ha sido hecha en las pruebas del SABER.

El segundo problema tiene que ver con la evaluación de las llamadas "competencias ciudadanas", área transversal que comenzó a evaluarse a partir de la tanda de 2005. Aun superando el escepticismo ocasionado por querer verificar habilidades casi exclusivamente comportamentales mediante pruebas escritas de tipo cerrado, quedan algunas dudas sobre el tipo de ítems que las constituyen, y lo que en realidad estarían indicando. En el documento preparatorio citado (ICFES 2005:30), se mencionan algunos ítems que no dejan claridad sobre lo que pretenden verificar

ni sobre el tipo de información que se pretende levantar con ellos: el examinado puede aprender fácilmente a responder lo que esperan que responda, como cuando se le pregunta cuántas veces le pegué a alguien (ítem 2), si las reglas se hicieron para romperlas (ítem 4), o si actuó cuando me da rabia (ítem 7).

Sin embargo, con todo y sus defectos de construcción, parece que la característica más importante detectada por las pruebas sigue siendo válida, y es que, en el mejor de los casos, el aprendizaje realizado en Colombia, como tendencia general, es más mecánico que flexible, más memorista que de comprensión, se orienta más a la reproducción rutinaria de conocimientos que a su uso competente en las diferentes formas que puede adquirir la cotidianidad -baja calidad del aprendizaje-. Es muy probable, sin embargo, que dichas deficiencias en el aprendizaje no solamente están siendo producidas por una docencia también deficiente -baja calidad de la enseñanza-, sino que igualmente estén inducidas por unos sistemas de evaluación masiva donde todavía predomina lo mecánico, lo excesivamente académico o la poca claridad en los planteamientos -baja calidad de la evaluación-. Los tres elementos juntos -calidad del aprendizaje, calidad de la docencia y calidad de la evaluación- no sólo se condicionan mutuamente, sino que de su interacción podría generarse una idea de aquello por lo que nos preguntamos inicialmente: ¿Cuál es la calidad de la educación?

IV. Aproximaciones complementarias a la calidad de la educación

A. Los estudios de factores asociados

En las pruebas de evaluación masiva generalmente se incluyen cuestionarios adicionales

destinados a los alumnos mismos, a los rectores o a los docentes, en los que se indaga por diversos tipos de información relativa a circunstancias o procesos que, se presume, es probable que estén asociados a los logros académicos que se están evaluando. Digamos que estos son los factores *asociables* al logro.

Una vez recogida la información, los especialistas usan diversos modelos estadísticos para determinar, con base en la información recolectada, de cuáles de esos factores puede decirse que están efectivamente asociados -positiva o negativamente-, con determinado margen de error, a los logros detectados en las pruebas. Es más, algunos llegan a emplear instrumentos de estimación estadística todavía más refinados que les permiten calcular qué tanto se incrementaría hipotéticamente el logro de los alumnos por cada tanto que se incrementara uno de los factores estudiados. Las herramientas estadísticas han convertido entonces a algunos de los factores asociables en *factores asociados* al logro académico o cognitivo.

El estudio de los factores asociados representa, entonces, otra aproximación -esta vez indirecta- al análisis de la calidad educativa. Ellos permiten analizar en qué circunstancias, en qué contextos y bajo qué condiciones es probable que se obtengan mejores resultados en las pruebas de logro de los alumnos que, en última instancia, están midiendo o verificando uno de los ejes importantes con que se determina la calidad de la escuela o del sistema educativo, cual es el del conocimiento.

Como no se trata de hacer una discusión conceptual sobre la metodología y los resultados de la literatura disponible, que va en aumento

constante, sino más bien de recoger los resultados más notables de los estudios que han acompañado los sistemas de evaluación masiva que mencionamos arriba, plantearemos a continuación tres preguntas que pueden ordenar la exposición: i) ¿de qué tipo de factores se trata? ii) ¿cómo se pueden agrupar esos factores? iii) ¿qué tan asociados están?

Respecto a la primera pregunta -*de qué tipo de factores se trata*- hay que admitir que prácticamente todas las circunstancias que rodean los procesos de docencia y aprendizaje son susceptibles de ser codificadas u operacionalizadas en forma de variables, por lo que podría afirmarse que todas pueden ser trabajadas como factores asociados o, al menos, asociables.

- En general, algunos distinguen entre factores directos (como sería la didáctica empleada por el maestro) e *indirectos* (como podría ser la presencia de campos de juego en el colegio), según sea su manera de incidir en los resultados (en este caso académicos) del proceso educativo.
- Desde el punto de vista de la pragmática en la política educativa, el LLECE distingue, además, entre factores más *estructurales* (las condiciones socioeconómicas de los alumnos, por ejemplo), y factores que pueden ser *modificados o afectados por la política educativa* (como sería el caso de las dotaciones escolares).
- Están además los factores asociados al *rendimiento individual* de los alumnos, tales como su edad, sexo, historia educativa o actitud hacia el aprendizaje, y factores que sólo pueden estudiarse desde su asociación probable al *rendimiento promedio de los alumnos de un grupo o colegio*, tales como el clima de aula,

o la administración y la dotación académica del plantel.

- Tanto los factores relativos a personas, como los referentes a procesos o circunstancias, pueden ser además de dos tipos. El primer conjunto de factores son fácticos o *descriptivos*, los cuales son más fáciles de detectar. Tales son, por ejemplo, las características sociodemográficas del rector, de los alumnos o de los docentes, para hablar de personas, y la cantidad de libros en la biblioteca, para hablar de cosas. Los otros factores son más *evaluativos*, de comportamiento o procedimentales, como serían el autoconcepto de los alumnos, el ambiente de aprendizaje en el aula, la calidad de los textos o la asiduidad del trabajo en la biblioteca. Estos últimos factores son más difíciles de detectar, requieren de una operacionalización o indexación para ser trabajados como variables, lo cual puede prestarse a simplificaciones exageradas.

En cuanto a la segunda pregunta *-agrupación posible de los factores-* las respuestas son también diferentes: PISA organiza los factores asociados de una manera, los estudios realizados a partir de las pruebas del SABER²² de otra y el LLECE de otra manera diferente. La agrupación que sigue procura dar cabida a todos los factores estudiados en los diferentes tipos de evaluación y a los estudios concomitantes. Esta agrupación es independiente del tipo de asociación, observable u observado; se trata por tanto de una agrupación posible de factores asociables, no necesariamente asociados:

- Un primer grupo de factores estaría relacionado con las políticas educativas globales del país, que de alguna manera puedan estudiarse en asociación con el logro académico. Ejemplos de este grupo pueden ser los currículos oficiales o el impulso de estrategias determinadas, como por ejemplo la implementación del modelo de Escuela Nueva, para el sector rural colombiano. Este tipo de factores solamente aparece en los estudios del LLECE.
- Están, en segundo lugar, los factores que tienen que ver con la escuela. Los grandes grupos de factores son los siguientes:
 - El rector (algunos lo consideran como un grupo aparte de factores), tanto sus características demográficas (edad, experiencia, sexo, formación, etc.) como sus actitudes (liderazgo pedagógico vs. liderazgo administrativo, principalmente). La presencia de liderazgo pedagógico en la escuela, que puede estudiarse también como factor asociable, no necesariamente debe radicar en la persona del rector.
 - La administración y organización de la institución o, como se llama a menudo, la "gestión escolar".
 - Los recursos humanos, especialmente los maestros. Se trata de analizar aquí tanto su cantidad como sus características más descriptivas: edad, sexo, experiencia, formación, remuneración. El comportamiento más específicamente pedagógico de los

²² Las pruebas del SABER recogieron datos adicionales, que permiten hacer un análisis de los factores asociados al logro en el aprendizaje, solamente en las primeras dos tandas aplicadas hasta el momento.

maestros se analiza, generalmente, como un factor de tipo didáctico, dentro del grupo de factores relacionados con el aula.

- La dotación, tanto de tipo general (espacios culturales, recreativos, etc.) como específicamente académica (aulas, bibliotecas, laboratorios, etc.) de la escuela.
 - Finalmente, hay un factor que se investiga tanto a nivel de escuela como a nivel de aula, y tiene que ver con el clima escolar. A nivel de escuela se refiere más que todo a los aspectos institucionales de las relaciones interpersonales, a la presencia de conflictos o de problemas de convivencia que tienen que ver con la escuela en su conjunto.
- Un tercer grupo de factores se relacionarían directamente con el aula, como lugar donde se realizan concretamente los procesos de docencia y aprendizaje. Aquí se incluyen en su mayoría factores de tipo actitudinal / procedimental y, aunque teóricamente serían los más decisivos en los procesos de aprendizaje, son los menos estudiados, al menos de manera directa. Este tipo de factores asociables se podrían organizar en dos subgrupos:
- El primer subgrupo tiene que ver con el ambiente del aula. Se podría distinguir entre un ambiente general de convivencia y un ambiente específico de aprendizaje. En la convivencia habría que considerar tanto las relaciones entre alumnos y docentes como las relaciones horizontales (convivencia, conflictos) entre los alumnos del grupo. En el ambiente de aprendizaje se estudian, por regla general, situaciones de aula que influyen en el aprendizaje: alumnos por grupo, condiciones generales de comodidad, conformación del grupo (el grupo de pares, según el estudio de Uribe), dotación de materiales.
 - El segundo subgrupo, más actitudinal, comprende los factores de tipo didáctico. Se refieren, generalmente, a la organización, metodología y rutinas seguidas en las sesiones de clase (ejercicios, trabajo en grupos/trabajo individual/clase magistral, enfoques pedagógicos, etc.). A juicio de los pedagogos y profesionales de la educación es el factor más importante; no obstante, es el menos estudiado.
- Un cuarto grupo de factores se relaciona con los alumnos. Estos factores, igualmente, pueden ser de dos clases:
- El primer grupo son características puramente descriptivas de los alumnos: edad, sexo, historia escolar, historia (si se tiene) laboral.
 - Un segundo grupo tiene que ver más directamente con las actitudes de los alumnos: hacia sí mismos (autoconcepto, principalmente), hacia los demás (actitudes de convivencia), hacia la escuela en general (identificación, satisfacción), hacia el aprendizaje y hacia el conocimiento específicamente. En este último tipo de factores PISA tiene anotaciones interesantes, referentes a factores asociados a la capacidad lectora, tales como gusto / placer por la lectura, diversidad del material de lectura, estrategias metacognitivas frente a la lectura y al aprendizaje.

○ Por último, los factores que tienen que ver con la familia del alumno, se podrían agrupar de la siguiente manera:

- El nivel socioeconómico de la familia, para lo cual se pueden recabar datos relativos al estrato, el tipo de vivienda, los ingresos y la educación de los padres.
- La estructura de la familia: extensa, monoparental, número de hijos.
- Las posesiones, especialmente de tipo cultural (libros, prensa, computador).
- La comunicación del alumno con su familia (especialmente los padres), tanto de tipo general (temas socioculturales) como relativo al aprendizaje (tareas, dudas). Un factor especial estudiado en este grupo es el de la actitud del alumno y de su familia ante las tareas asignadas en la escuela (si le ayudan, si las hace solo, si consulta, si la familia se margina del problema...).
- Relaciones de la familia o los padres con la escuela.
- En PISA se pregunta además si el idioma "oficial" de la familia es el del país, como indicador para tratar de calificar el factor migración externa. En Colombia la migración externa no es tan significativa, pero sí podría serlo la migración interna, o su agravante el desplazamiento forzado.

Finalmente, la tercera pregunta *-qué tan asociados están-* se refiere al grado de correlación (positiva o negativa) detectado con el aparato estadístico entre esos factores mencionados y los

resultados del logro; vale decir, cuántos de esos factores asociables pueden ser calificados, desde el punto de vista de la probabilidad estadística, como verdaderos factores asociados.

Ante la imposibilidad de mencionar al detalle todos los estudios de factores asociados, nos referiremos en lo que sigue al resumen elaborado por el profesor Cepeda (s.f.), de la Universidad de los Andes, sobre los factores asociados encontrados para Bogotá; puntualmente lo complementaremos con los hallazgos de las otras fuentes. Estos son los factores a los que les detectó una asociación más significativa con los resultados académicos mostrados por los alumnos:

○ En los factores relacionados con las políticas educativas nacionales, LLECE encontró una asociación positiva entre los resultados de los alumnos y la política nacional colombiana de implementar de manera masiva el modelo de Escuela Nueva para el sector rural. En contra de lo esperado, por otra parte, de los resultados obtenidos por Colombia en el TIMSS (buenos currículos, malos resultados) se deduce que la oficialización del currículo no es relevante para los resultados académicos. LLECE encontró una ligera correlación positiva entre el logro y el hecho de haber cursado algún tipo de preescolar; si bien esta asociación se relaciona con las características individuales (historia educativa) de los alumnos del numeral 4), la existencia misma del preescolar en el sector oficial es una política del gobierno central que, en casos como el colombiano, es reciente y todavía no ha logrado su completa implementación.

○ Entre los factores relacionados con la escuela, cabe mencionar los siguientes:

- No es muy claro el significado que tienen las características personales (formación, edad, sexo) del rector; por otro lado, una asociación positiva se encontró con su liderazgo o interés de tipo pedagógico, más que con el de tipo administrativo.
 - Hay una correlación positiva entre el logro y la calidad de las instalaciones educativas en general, y especialmente las pedagógicas. Se menciona, en particular, la dotación y los servicios de la biblioteca. No aparecen aún estudios sobre la correlación entre logro y dotación de equipos de cómputo, aunque algo se puede inferir de los estudios del TIMSS (ver recuadro 2).
 - En cuanto a las características de los docentes, Claudia Uribe (2003) encontró positiva la asociación de su formación (más la licenciatura que los postgrados) y de su experiencia con el logro de los alumnos. El incremento en los resultados asociado al incremento en la experiencia, sin embargo, es más notable en las primeras fases de la carrera docente y disminuye después de algunos años; LLECE encontró también asociaciones bajas entre experiencia docente y logro escolar.
 - Hay asociación positiva con el espíritu colaborador de los docentes, con las prácticas de evaluación a los mismos por parte de los colegios, y con la carga laboral: los rendimientos más altos se asocian a 20-25 horas semanales de clase, y disminuyen con rubros superiores o inferiores a este rango. En general, LLECE encontró también correlaciones positivas con la satisfacción del docente, tanto con su salario como con la autonomía de que goce para realizar su labor pedagógica.
 - En el clima escolar se detectó asociación positiva con la existencia de actitudes positivas y buenas relaciones entre los diferentes estamentos de la comunidad educativa.
 - Los estudios realizados por la Misión Social con los datos de Bogotá muestran que, como tendencia general, la asociación entre las características del plantel y el logro académico es más estrecha en primaria que en secundaria.
- Entre los factores relacionados con el aula, aquellos más notables por su asociación significativa con el logro fueron los siguientes:
- LLECE encontró que el factor con mayor grado de asociación entre los estudiados fue el del clima general de convivencia en el aula. Sin embargo, la variable se refiere a la convivencia en cuanto a percibida por los alumnos, por lo cual bien puede confundirse o sobreponerse con factores más subjetivos relacionados con los alumnos, del grupo siguiente de factores.
 - En cuanto al clima de aprendizaje, se encontró una asociación negativa con un número muy grande de alumnos por aula, especialmente cuando pasan de cuarenta (ver recuadro 2).
 - Una serie de prácticas pedagógicas indagadas muy al detalle en las pruebas del SABER, no mostraron correlación alguna (ni positiva ni negativa) con el logro; las principales, enumeradas por Cepeda,

Recuadro 2
Rendimiento académico y tamaño de los grupos de clase

Existen posiciones encontradas sobre la relación entre el tamaño de las clases y el rendimiento académico: tradicionalmente los pedagogos han sostenido que un número menor de alumnos por grupo redundaría en mayor motivación de estudiantes y docentes y en mayor disponibilidad de éstos últimos para resolver problemas individuales, mientras que los diseñadores de política educativa sostienen que aumentar el número de alumnos por grupo no afecta para nada el rendimiento y que, por el contrario, se traduce en una mayor eficiencia del gasto en educación.

Los resultados del TIMSS parecen darle la razón a esta segunda posición: algunos países asiáticos, que ocuparon los primeros rangos en la prueba, se caracterizan por tener aulas con promedio de alumnos bastante elevado, como Singapur (36), Hong Kong (40) y Corea (46), mientras que nuestro país, cuyos resultados ocuparon los últimos lugares, ostentó un número significativamente menor de alumnos por grupo en promedio (33). En diferentes países, además, los mayores porcentajes de puntajes altos se registraron en las escuelas más grandes (Martin 1999 y 2000).

Con cierto despliegue la revista *Semana* hizo eco en nuestro medio de esos hallazgos, al citar en su edición del 18 de abril de 2002 los estudios del profesor Ehrenberg, de la universidad de Cornell, en los que afirma que la mayoría de los estudios aportados para defender el mayor rendimiento académico de las clases pequeñas no son serios ni consistentes.

Sin embargo, este tipo de conclusiones apresuradas puede también ser leído de manera diferente, al ubicar los datos en su contexto. En efecto, los estudios citados del TIMSS afirman también, primero, que en muchas escuelas se observa la tendencia a ubicar a los alumnos con problemas académicos en grupos más pequeños, con lo cual se estaría sesgando la relación entre rendimiento y tamaño del grupo; y, segundo, que en los mismos países donde se detectaron rendimientos sobresalientes y tamaños grandes de grupos se observó también una dotación significativamente mejor de las aulas, que pudo haber contrarrestado el impacto negativo de los tamaños de las clases. Para citar un solo ejemplo, mientras en Singapur, Hong Kong y Corea el porcentaje de escuelas sin computadores es 9, 17 y 0 respectivamente, en Colombia es del 68 por ciento; en Singapur el 39 por ciento de sus escuelas dispone de un computador para menos de 15 alumnos, en Colombia sólo el uno por ciento de las suyas.

En una revisión de la literatura efectuada por Blachford (1998) se concluye que, aunque la investigación al respecto es insuficiente, ésta se centra casi siempre en medir la relación entre tamaño de clases y rendimiento académico, y pocas veces entre tamaño y procesos pedagógicos. Una revisión más detenida de la literatura, sin embargo, le lleva a concluir que se dan relaciones verificables entre el tamaño y una serie de procesos pedagógicos, tales como los criterios para la organización de subgrupos en las clases, la naturaleza y calidad de la docencia, la atención de los alumnos, la adaptación de los alumnos a la escuela, además de la moral, el estrés y el entusiasmo con que trabajan los docentes. En Colombia habría que medir un factor adicional, que se presenta cuando las decisiones de política obligan a incrementar indiscriminadamente el número de alumnos por grupo, y es el del hacinamiento de los niños en espacios muy reducidos.

Las conclusiones de Claudia Uribe (2003:122) en su tesis doctoral son muy tajantes al respecto: "my study finds a substantial effect of class size on students' fifth grade mathematics achievement"; según su estudio, se espera que los logros de los estudiantes en una clase de 39 alumnos de matemáticas de quinto grado estén 0.4 desviaciones estándar por debajo de los de una clase de 29 (controlando las características de los docentes).

Más cautelosas son las conclusiones que arroja el LLECE (2000:18): hay una correlación negativa entre tamaño de las clases y rendimiento, pero ésta es "de proporciones sumamente pequeñas". Cepeda (s.f.) puntualiza un poco más, al afirmar que la correlación negativa encontrada comienza a ser de importancia solamente con tamaños muy grandes de las clases: a partir de los 40 y, más todavía, a partir de los 47 alumnos por clase.

Una vez más, hay que concluir que el análisis descontextualizado de cierto tipo de factores asociados al resultado académico puede apresurar conclusiones -y decisiones de política- poco acertadas. Como lo afirman los estudios de PISA (2003b:21), "no hay un factor que explique por sí solo por qué algunas escuelas o algunos países tienen mejores resultados que otros. El desempeño exitoso se atribuye a un conjunto de factores, incluyendo los recursos escolares, la política y práctica de la escuela, y la práctica de clase".

son las siguientes: elaborar un plan diario de clase, planear ejercicios, desarrollar experiencias fuera del aula, fomentar actividades de integración con otras áreas, realizar anotaciones de seguimiento a los alumnos, anunciar el tema o relacionarlo con los hechos cotidianos para captar la atención de los alumnos, realizar durante las clases demostraciones, ejercicios de observación, ejercicios para resolver, solicitar a los alumnos que presenten aplicaciones concretas de lo tratado en clase. Tampoco se hallan asociaciones significativas con el enfoque didáctico declarado por los docentes (conductista, culturalista, integrador, constructivista), ni con organizar el trabajo en grupos, autónomos o dirigidos, homogéneos o heterogéneos. Tan sólo se encontraron asociaciones entre el logro y el tiempo dedicado al trabajo individual (frente al colectivo) y negativas con respecto al tiempo dedicado a escuchar (pasivamente) al maestro. Una correlación positiva fue encontrada igualmente por LLECE entre los resultados académicos y el hecho de que se cuente con un sistema concreto y explícito de evaluación de los alumnos.

- Entre los factores relacionados con los alumnos, cabe destacar:
 - Por un lado, la asociación negativa con la extraedad. Por otro lado, la ventaja de los varones sobre las mujeres en matemáticas, y de éstas en lenguaje, aunque no tan pronunciada en América Latina ni en

Colombia como la encontrada por PISA en los países industrializados.

- En cuanto a las actitudes generales, se detectó una asociación negativa con la preocupación por los problemas económicos y familiares, y con la angustia producida por la situación social del país, especialmente en lo tocante a la violencia y el desempleo. Hay una correlación positiva con las posibilidades de esparcimiento y diversión y con la afición no excesiva a la televisión.
- El estudio de PISA encontró que la autoestima (también detectada en los estudios colombianos), la actitud positiva ante el estudio, los hábitos de lectura y el acceso a material diverso de lectura (no solo libros y no solo textos) tienen una correlación positiva con los logros académicos.
- Igualmente, hay una correlación positiva con el estudio individual (frente al estudio acompañado por compañeros).
- Finalmente, los factores relacionados con la familia son quizá los más mencionados en todos los estudios. Muchos de ellos tienen un carácter estructural, que los hace muy difíciles de modificar desde la perspectiva del sector educativo. Los más importantes fueron:
 - Correlación negativa con el tamaño de la familia, medido en el número de hermanos²³. PISA, por su parte, no encontró asociación entre el logro y la característica monoparental de las familias.

²³ En países no industrializados se ha comprobado una correlación alta entre pobreza y tamaño de la familia; es posible, por lo tanto, que sea la variable pobreza la que en última instancia explique la correlación negativa encontrada entre tamaño de la familia y rendimiento académico.

- Existe una fuerte correlación significativa entre el logro y el contexto socioeconómico y cultural de las familias, medido en términos de estrato económico, nivel educativo de los padres y posesiones de la familia, especialmente en el ámbito cultural: libros, computadores, obras de arte.
- Con respecto al ambiente familiar se encuentran relaciones positivas con la atención recibida de los padres, saberse querido por ellos y poder contar con ellos en las dificultades. LLECE, que evaluó a niños más pequeños, encontró asociación positiva entre los resultados y el hecho de que los padres lean a sus hijos. PISA encontró, adicionalmente, una correlación positiva con la frecuencia con que se tratan en familia temas de interés económico y social generales. Quizá tenga eso alguna relación con la asociación encontrada en Colombia con el tiempo de permanencia de los padres en la casa, en las horas no laborales.
- En relación más explícita con lo académico, hay correlación positiva con el interés de los padres por el estudio de sus hijos, por revisar y corregir (no hacer) sus tareas, por informarse y recibir información de lo que pasa en la escuela.

B. Los estudios cualitativos

El estudio de los factores asociados es ante todo un ejercicio analítico, que tiende a dar cierta claridad sobre el grado probable de incidencia de una serie de elementos y procesos diversos (desglosables, agrupables y vueltos a desglosar, cuantificables y medibles casi indefinidamente) en aquello que finalmente pretendemos cap-

tar o detectar, como es la calidad del sistema educativo. El resultado final es un mapa más o menos coherente, más o menos agregado, de una serie cada vez más grande de causalidades o asociaciones probables. Las conclusiones de los estudios casi siempre van igualmente en el mismo sentido: uno sólo de esos factores no explica el comportamiento del sistema, lo más probable es que éste se deba a una combinación de los mismos. ¿Cómo es esa combinación, cómo se comportan unos factores con otros en la vida real de las instituciones educativas, del aula de clase, o de la salida de la escuela donde el alumno pone en práctica u olvida sistemáticamente lo aprendido en las aulas?

Los estudios cualitativos pretenden responder a ese interrogante. Son por eso complementarios y en ningún momento sustitutivos de la detección del logro y del análisis estadístico de los factores asociados al mismo. En tales estudios pretende acercarse a escuelas concretas, donde se ha medido el nivel de logro de los alumnos y donde se ha indagado por el grado de asociación de aquellos factores que se prevé inciden o acompañan al logro observado, para ver en la práctica cómo se comportan, vale decir, cómo es la combinación concreta de factores de la que, se sospecha, provienen los resultados observados. En otras palabras, se busca pasar de la *explicación* de las relaciones entre factores convertidos en variables, dependientes o independientes, intervinientes o asociadas, a la *comprensión* concreta de cómo funcionan los hechos o procesos que prefiguran dichas variables en la cotidianidad de hacer educación, en el funcionamiento concreto de escuelas determinadas. Es decir, se busca complementar la visión *analítica* que arroja el análisis estadístico con la visión *sintética* que da el estudio cualitativo.

Los estudios cualitativos generalmente se emprenden con la metodología del estudio de caso²⁴. Cuando se toman como complemento de las evaluaciones masivas, el "caso" se refiere concretamente a una escuela o grupo de escuelas cuyos alumnos participaron en las pruebas. Es útil seleccionar los sujetos de estudio (escuelas o conjuntos de escuelas) de los rangos extremos, bien sea de entre aquellos cuyos alumnos mostraron los mejores resultados en promedio, o de los peores.

En los cinco sistemas de evaluación masiva estudiados encontramos dos que contaron con estudios de caso complementarios realizados con escuelas exitosas. El primero de ellos se refiere a las pruebas del SABER²⁵: valiéndose de la información suministrada en sus resultados, CORPOEDUCACIÓN seleccionó de entre las escuelas exitosas bogotanas un grupo de doce con las que realizó estudios de caso. El estudio buscaba la comprensión concreta de cómo funcionaban dichas escuelas, particularmente desde el punto de vista de la organización y gestión escolar. Aunque esta investigación (no publicada) no pertenece oficialmente al sistema SABER, sí puede tomarse indirectamente como un complemento y secuela de la misma, particularmente de la tanda de evaluación realizada en Bogotá en 1998.

Los estudios de caso realizados por el LLECE tienen un interés adicional por dos motivos: en primer lugar, porque fue una investigación complementaria patrocinada y coordinada directamente por el mismo Laboratorio responsable de

la evaluación masiva. En segundo lugar, porque además de tratar de comprender el funcionamiento de la escuela como tal, su organización y su gestión (a lo que se limitó CORPOEDUCACIÓN), se metió en el aula, realizó indagación y buscó comprensión de lo que sucede en ella, de los comportamientos y las relaciones entre sus actores, de lo que se pretende lograr, de la manera como se trabaja y de lo que de hecho se logra, elementos éstos que han sido los que menos se han estudiado y analizado en las investigaciones de los factores asociados.

Las instituciones seleccionadas por el LLECE para los estudios de caso fueron un total de 35 de siete de los países participantes, entre ellas cinco de Colombia. Hay escuelas de diverso tipo: urbanas y rurales, grandes y pequeñas, oficiales y privadas. El grupo total no puede tomarse como una muestra, en el sentido estadístico del término, y la validez de los resultados no está en la representatividad estadística de un universo investigado, sino en la comprensión lograda de cómo funciona la mezcla de factores (investigados o no) en situaciones concretas que están detrás de un tipo de resultados detectados (en este caso promedios altos de acierto en las respuestas a las pruebas planteadas). El ideal sería entonces reconstruir el funcionamiento de estos factores (la mezcla que produce los resultados) en forma de un relato²⁶ coherente para cada institución. Por organización del tiempo y de los resultados, en ambos casos se optó más bien por presentar los datos de manera agregada, por capítulos o por

²⁴ La observación se refiere a los estudios cualitativos cuyo objetivo es una escuela o grupo de escuelas (ver metodologías y clases de estudios en Senhouse 1988). La indagación de los comportamientos en el aula, sin embargo, puede hacerse con otro tipo de metodologías cualitativas, compiladas en diversos artículos del manual editado por Keeves (1988).

²⁵ Véase Camargo (2000).

temas. Cada capítulo o cada tema se soporta en el documento final con observaciones o testimonios realizados en las instituciones.

En el estudio del LLECE se analizaron los siguientes tópicos o temas relacionados con la gestión escolar: los procesos relacionados con la construcción del ambiente de aprendizaje en la escuela, el ámbito organizacional de la misma, la forma peculiar de hacer las cosas en ella, la administración de recursos humanos y financieros, las autoridades de la escuela, las normas, los modelos de decisiones académicas e instancias relevantes de toma de decisiones, la cultura organizacional. Como queda dicho, varios de estos tópicos contienen factores cuya asociación con los resultados del logro cognitivo de los alumnos ha sido estudiada por diferentes estudios de tipo analítico.

Los tópicos relacionados con la práctica pedagógica, que fueron estudiados desde el acontecer del aula, fueron los siguientes: el repertorio de estrategias y principios de acción sostenidos por los docentes, la administración de los recursos del aula, del currículum, del tiempo asignado, la perspectiva o enfoque pedagógico global, el nivel de compromiso profesional de los docentes, las expectativas de los docentes sobre sus alumnos, los recursos materiales disponibles para los procesos del aula. Aunque parece un listado de temas significativamente menor, el tema primero

(repertorio de estrategias didácticas) es el más extenso del estudio, y se desagrega en tópicos tan diversos (cerca de 20) como las rutinas de trabajo, el manejo no tradicional del error, lo acogedor del ambiente de aula, la integración de las prácticas a la vida del alumno, la integración del juego como recurso didáctico, la comunicación entre los actores, el trabajo de análisis individual, grupal y del conjunto de la clase, etc. Muchos de estos tópicos o "factores", al analizarse individualmente pueden no mostrar un grado de correlación alto con un nivel alto o aceptable de resultados en las pruebas pero, al tomarse en su conjunto, nos están informando sobre cómo funcionan las escuelas cuyos alumnos, precisamente, muestran buenos resultados en esas pruebas.

Finalmente, el estudio concluye con el análisis de algunos tópicos no previstos inicialmente, pero que resultaron importantes para la comprensión de lo que ocurre en las escuelas seleccionadas, sobre temas relativos al clima escolar y a la interacción de la institución con el contexto.

Un ejemplo concreto de cómo los estudios de caso pueden aportar niveles de comprensión difíciles de alcanzar por los estudios analíticos es el de la variable "clima de aula", que en el estudio del LLECE mostró el mayor grado de asociación con los resultados académicos de las pruebas. La definición operativa de la misma, sin embargo, fue muy pobre, ya que se la construyó a partir

²⁶ Se trata de un relato que involucre testimonios, integre situaciones de la cotidianidad, incorpore perfiles de los actores, adjunte material adicional (gráfico, auditivo, documental), etc., de manera que el lector pueda reconstruir situaciones con coherencia y verosimilitud. Puesto que los estudios de caso no tienen la pretensión de validez estadística, no tendría sentido reducir su informe (como desafortunadamente muchas veces se hace) a un simple listado de categorías (en vez de variables), y sintetizar sus rasgos comunes (en vez del comportamiento promedio o estadístico de las variables), en un esfuerzo vano por remedar los estudios analíticos más cuantitativos. Es lo mismo que pasa con la historia, cuando se quiere superar el listado anodino de fechas, personajes, lugares y batallas no queda más remedio que acudir al relato, igualmente consistente, verosímil e integrador.

de tres variables dicotómicas que expresaban el sentir de los alumnos interrogados sobre si los otros los molestan (sí/ no), si son buenos amigos en el grupo (sí/ no) y si ocurren peleas frecuentes en el mismo (sí/ no). Como consecuencia, a pesar de tener un gran poder explicativo la medición de esta variable, su grado de simplificación es tal, que si no se enriquece con el conocimiento *ad hoc* de cómo se comportan los ambientes reales de clase, cuál es la interacción real de los actores en los mismos, qué tipo de conflictos se presentan y cómo los manejan las escuelas exitosas (y cómo las escuelas deficientes), es poco lo que se gana en la comprensión de la misma como factor o contexto importante de la producción de determinados resultados académicos.

C. El criterio de los expertos

Un acercamiento adicional a la situación educativa en el país es el juicio o criterio de expertos sobre aspectos importantes de la misma, que complementan tanto los estudios cuantitativos como los cualitativos y definen, a partir de una ponderación más o menos juiciosa de las buenas prácticas educativas (reales o pensadas), lo que podríamos llamar la "ortodoxia" educativa del país.

Acertados o no, los criterios de los expertos han sido los responsables de muchas decisiones concernientes a la calidad de la educación del país, tales como el diseño de los currículos oficiales, la definición de las áreas del conocimiento objeto de la docencia escolar, la determinación de lo que se espera de esas áreas y las pautas para la elaboración de pruebas que midan si se logró lo

esperado. En pocas palabras, los "expertos" han definido siempre estándares o metas, el camino (currículo sugerido) para lograrlos y los criterios para verificar si se lograron o no.

El juicio de los expertos, respaldado generalmente por criterios pedagógicos y educativos, no siempre está exento de la "contaminación" de algunos intereses que pueden estar latentes debajo de la mesa. Un ejemplo muy concreto de ello fueron las discusiones en torno a las llamadas "áreas del conocimiento" definidas en Colombia -caso insólito- por la ley²⁷, cuyo producto final, más que las buenas prácticas y los consensos aterrizados en torno a una estructura curricular básica, fue más bien resultado de transacciones y estrategias de negociación, en ocasiones por debajo de la mesa, entre los intereses defendidos por grupos de mandos medios del gobierno y las presiones de los representantes sindicales en las mesas de negociación.

Sin embargo, no todos los resultados de este tipo de decisiones tienen que ser objeto de transacciones parroquiales de intereses creados. El criterio de los expertos sigue siendo válido en la toma de muchas decisiones que tienen que ver con el futuro de la calidad educativa. En PISA, por ejemplo, fue el criterio de los expertos el que determinó por consenso lo que se esperaba medir (el logro deseado para los distintos niveles de competencias científicas, matemáticas y de comprensión de lectura) y los instrumentos (ítemes de las pruebas) mismos con que se midieron.

Mencionemos, a guisa de ejemplo, tres instancias críticas para la calidad de la educación

²⁷ Ley General de Educación (ley 115 de 1994).

contemporánea en general y colombiana en particular, cuyo énfasis y llamado de atención, más que de resultados empíricos de medición de la calidad, provienen del buen juicio de expertos que vienen trabajando el tema de tiempo atrás.

El primer ejemplo tiene que ver con la problemática de las bibliotecas. Jorge Orlando Melo, desde su posición de director de la principal biblioteca del país, y por lo tanto en su calidad del principal experto colombiano en el tema, viene poniendo el dedo en la llaga desde hace un buen tiempo. Por un lado, está su análisis (Melo 2001) de la pobreza histórica de las bibliotecas universitarias del país, lo que lo lleva a concluir que varias generaciones de profesionales colombianos se formaron, no en la consulta y análisis de fuentes diversas, sino en el estudio mecánico de fotocopias del libro o texto favorito de cada docente y de sus apuntes de clase. Su análisis también se ha referido (Melo 1998) a los centros de educación básica: independientemente de lo que concluya cualquier investigación empírica, es evidente que no puede desarrollarse la comprensión de la lectura con unos docentes que no leen y unos alumnos que no tienen acceso a los libros; y, para los alumnos de más bajos recursos, el acceso a los libros significa la existencia de bibliotecas escolares. Lo que existe, desafortunadamente, no son bibliotecas escolares donde se facilita el acceso de los alumnos a diversos tipos de materiales textuales, sino simples anaqueles donde se guardan libros y textos viejos que ya nadie usa.

Un segundo ejemplo se refiere a la integración curricular o integración de áreas, tan de moda

en los últimos años. Un grupo de expertos, en su mayoría antiguos colaboradores de los grupos disciplinares de desarrollo curricular patrocinados por el Ministerio de Educación, resolvieron acogerse al patrocinio de una ONG nacional y lanzarse, junto con algunos maestros, a la aventura de proponer alternativas novedosas de integración curricular para los colegios colombianos. Aunque su trabajo está todavía inconcluso, algunas publicaciones son ya un adelanto del mismo, en especial el libro *El saber tiene sentido* (Vasco et. al. 2001), donde sitúan el problema curricular dentro del contexto de una cosmovisión y de una visión epistemológica que organiza el conocimiento, construye un relato de fondo serio que le presta sentido y no cae en el verbalismo pseudocientífico vacío que prolifera en muchos de los escritos pedagógicos contemporáneos que se ven en el país.

Para el tercer ejemplo situémonos en la reflexión que han suscitado los resultados del PISA en ciertos grupos de expertos educativos de Alemania, país que no salió muy bien librado en los resultados de la primera aplicación del 2000, frente al comportamiento promedio del resto de los países industrializados. Si bien algunos se rasgan las vestiduras y tratan de desencadenar una especie de "síndrome de Sputnik" con los resultados, hasta el punto de proclamar la necesidad de pasar a una nueva era, de comenzar de cero la reconstrucción del sistema -*Lernen nach Pisa*, aprender después de Pisa²⁸-. Otros son más críticos, y se plantean interrogantes como el siguiente, válido también para nosotros: en un mundo altamente competitivo y globalizado, donde las evaluaciones masivas convierten el

²⁸ "Lernen nach Pisa", por ejemplo, es el tema central del número 2 de la revista *Pädagogik* en febrero de 2003.

aprendizaje en un proceso efectista que tiene que mostrar resultados inmediatos, ¿dónde queda la escuela de la reflexión, de la meditación, de la contemplación activa, del profundizar y ver las cosas desde diferentes perspectivas? La escuela de la reflexión, de la *Nachdenklichkeit*²⁹, podría llamarse ese conjunto convergente de pensamiento, que estaría de alguna manera emparentado con las reflexiones de un pedagogo colombiano, convertido después en político, quien abogaba por unos currículos intensivos (de profundización, de reflexión, hasta cierto punto de contemplación), frente a unos currículos extensivos, enciclopedistas, de datos y respuestas inmediatas, que prepara para responder a preguntas cerradas de selección múltiple, como son las que se plantan en las pruebas para las 22 asignaturas con que examina el ICES a los que terminan la secundaria.

Igualmente, podrían mencionarse otros temas críticos abordados por diferentes grupos de expertos, tales como las opciones metodológicas básicas o las reflexiones en torno a escuela y convivencia o escuela y conflicto, que no aparecen con mucha nitidez en los análisis estadísticos de factores asociables a unos resultados medidos con unas pruebas de evaluación masiva.

V. Análisis final: ¿hay respuestas a las preguntas iniciales?

Si regresamos a las cuatro preguntas sobre la calidad de la educación en Colombia, planteadas

al inicio³⁰, y queremos responderlas con base en las reflexiones hechas hasta aquí, tendremos que decir lo siguiente:

- La primera pregunta -qué tanto les atrae la escuela a niños y jóvenes, qué tanta es la satisfacción que les produce lo que hacen en ella- partía de la consideración de la escuela como un bien en sí, de lo que la escuela hace como algo deseable por sí mismo y no sólo por lo que ayuda a alcanzar otros bienes. En este sentido, es muy poco lo que se puede afirmar en el momento. Plantearse si están contentos los niños en la escuela es tanto como preguntarse por la felicidad de los niños, en qué consiste y cómo se logra, tema poco trabajado en la literatura relativa a educación y pedagogía³¹. Aunque sería posible medir el grado de satisfacción construyendo algunas variables a partir de ciertas preguntas, las conclusiones serían muy pobres y habría que echar mano de las investigaciones cualitativas; lo poco que se sabe con estas es que satisfacción personal y logro de objetivos cognitivos pueden estar estrechamente correlacionados, y que es posible que niños que no aprendan o vivan un clima conflictivo en la escuela, no se sientan muy felices en ella.
- Para responder a las otras tres preguntas que siguen, contamos con instrumentos de verificación en lo relacionado con el conocimiento y bien poco en lo relacionado con la convivencia. En la primera de ellas, sobre si se

²⁹ El número doble (7-8) de la misma revista en 2002 (julio-agosto) traía como uno de sus temas centrales "Schule der Nachdenklichkeit".

³⁰ Los cuatro planteamientos iniciales para verificar la calidad de la educación, que fueron mencionados en uno de los acápites iniciales de este trabajo (II.C. Hacia una aproximación a la calidad de la educación).

sabe o no qué es lo que se pretende enseñar, el TIMSS avanzó notablemente en el estudio de los currículos. Mencionábamos atrás que en Colombia se legisló sobre plan de estudios prácticamente cada década y se produjo una renovación curricular importante en la década de los ochenta, y que después de cierta anarquía derivada de la autonomía curricular establecida por la ley general de educación, se ha regresado a ciertos parámetros nacionales con el establecimiento de unos estándares curriculares mínimos, de reciente divulgación³², tanto en las áreas básicas de matemáticas, lenguaje, ciencias y sociales como en el área transversal de competencias ciudadanas. El TIMSS, sin embargo, encendió una especie de alerta roja al afirmar que no es solamente el currículo oficial lo que importa, sino la porción de éste que realmente se enseña y, de esto último, la porción que a su vez aprenden los alumnos. Es más, si bien en Colombia se detectaron avances significativos en el currículo oficial, el currículo aprendido estuvo por debajo de casi todos los países participantes en esa misma evaluación masiva.

Quizá esta distancia entre el currículo oficial y el currículo realmente aprendido resida en el nivel intermedio: el currículo enseñado. No sólo las condiciones de enseñanza distan mucho de ser las ideales que supone el currículo oficial, sino que habría que preguntarse

hasta qué punto los docentes, responsables del currículo realmente enseñado, llegan a conocer de verdad (y no por terceras interpretaciones) el currículo oficial, cuando de éste se imprimen cuando mucho 10.000 ejemplares de cada edición para los cerca de medio millón de docentes existentes en el país, cuando la mayor parte de los colegios no disponen ni de la facilidad de fotocopiado ni -modernamente- del acceso masivo a Internet.

- La tercera pregunta se refería a la pertinencia de los conocimientos que se pretende enseñar y de las habilidades sociales que se pretende construir. Ya el TIMSS dio la respuesta: tal vez los currículos son pertinentes, lo que no está claro es si lo son los instrumentos que se adoptan a) para enseñarlos (la docencia) y b) para lograr que los alumnos se los apropien (el aprendizaje). Un análisis de la pertinencia de los currículos, por otro lado, tendría que involucrar la pregunta de hasta qué punto hemos logrado avanzar desde la definición de estándares cognitivos de una pertinencia marcadamente "escolar" (competencias académicas), válida para los representantes de cada parcela del conocimiento que elaboran los currículos, hacia una relevancia más centrada en los problemas y retos que plantea la cotidianidad de la sociedad del conocimiento en el siglo XXI (competencias para la cotidianidad, tanto privada como pública y laboral).

³¹ Fuhr (2002), por ejemplo, insiste en la necesidad de investigar más el tema de la felicidad de los niños y de tener más en cuenta sus repercusiones en la pedagogía.

³² En 2002 se publicaron unos estándares provisionales, grado por grado, para las áreas de matemáticas, lenguaje y ciencias. En 2003 se publicaron estándares definitivos para matemáticas y lenguaje, en 2004 para ciencias, sociales y competencias ciudadanas. Los estándares definitivos ya no están definidos grado por grado, sino para los grupos de grados (1-3), (4-5), (6-7), (8-9) y (10-11).

En esto hay debates, detractores y defensores, pero una mirada desapasionada a los ítemes de las pruebas actuales, tanto del ICFES como del SABER, permite sospechar que la indagación todavía se orienta preferencialmente desde la visión un poco cerrada y hasta cierto sentido autista de los grupos académicos que manejan cada área del conocimiento. Si bien sólo los miembros de una comunidad científica validan en la sociedad postmoderna los avances en dicha área, en la educación básica tiene que ser una discusión interdisciplinaria la que los valide, puesto que su razón de ser está por fuera de cada área del conocimiento. En otras palabras, para citar un ejemplo, se estudian matemáticas en la educación básica no para ser matemáticos ni para responder a la racionalidad del conocimiento matemático, sino para responder a una serie de expectativas y retos ciudadanos que trascienden el mundo de las matemáticas y de los matemáticos³³.

- Finalmente, la cuarta pregunta se orienta hacia el grado en que los conocimientos elaborados o contruidos por los alumnos responden realmente a los planteados por la sociedad para la escuela. Las evaluaciones masivas disponibles permiten responder que lo logrado por la escuela, si podemos hablar de promedios o tendencias nacionales, se centra mucho en el aprendizaje mecánico, de códigos en el caso del lenguaje y de algoritmos en el caso de las matemáticas.

Sin embargo, es posible que en el cuartil superior del rendimiento promedio por escuelas

podamos encontrar situaciones interesantes que se puedan profundizar mediante estudios orientados a la mayor comprensión, como lo son los estudios de caso. Igualmente sería válido el estudio de experiencias de aula más exitosas que otras dentro de una misma escuela. En esos casos convendría hallar respuestas al interrogante de por qué, en una situación general de resultados muy deficientes en cuanto al aprendizaje de conocimientos y habilidades útiles para el mundo de hoy algunas escuelas, públicas o privadas, urbanas o rurales, o algunas experiencias de aula logran que sus alumnos, ricos o pobres, como tendencia se aparten del comportamiento global y muestren resultados satisfactorios en sus aprendizajes.

- Queda por plantear una última pregunta que en cierto sentido condiciona las respuestas que hemos dado a las preguntas anteriores, y sería aproximadamente del siguiente tenor: ¿qué tan acertados o adecuados son los instrumentos con que estamos verificando la calidad de nuestras escuelas? O mejor aún, ¿qué tan útiles son como instrumentos de información en el momento en que se tengan que diseñar políticas tendientes a elevar esa calidad de la educación que ellos miden? En la respuesta tendrían que salir a flote dos escollos ya mencionados, sin que por ello tengamos que negar que es mejor medir algo a medir nada. El primer escollo tiene que ver con la insistencia marcadamente "escolar" con que se continúan elaborando las pruebas, que implica una concepción de la escuela al

³³ Esto implicaría igualmente, que en el "criterio de los expertos" a que nos referimos en el acápite anterior, habría que pasar también de un criterio meramente disciplinar especializado o "endogámico" a otro de tipo más integrador, como el logrado por Carlos Vasco y su grupo en la obra citada (2001).

servicio del conocimiento académico, y no al servicio del ciudadano común y corriente, que necesita desarrollar unas herramientas para vivir mejor el siglo XXI, y no necesariamente para ser un científico -al menos en lo que a la educación básica se refiere-. El segundo escollo tiene que ver con la falta de consistencia de unos instrumentos que, como en el caso del SABER, si bien son útiles para medir o comparar particularidades regionales o institucionales en un momento determinado, se muestran inoperantes para analizar la evolución misma de la calidad del sistema educativo en el tiempo, ya que los resultados de sus medidas, aún en el corto lapso que lleva su aplicación, no son consistentes entre sí.

Resumiendo este acápite, tendría aquí también que concluirse que una visión completa de la calidad de la educación tiene que ser el resultado combinado -la mezcla- de una aproximación masiva con pruebas donde se verifiquen resultados concretos de los aprendizajes, más una serie de estudios que nos permitan sospechar qué factores o conjunto de factores puedan estar asociados a ese tipo de resultado, más la constatación en la práctica sobre cómo se combinan dichos factores en las escuelas y en las aulas que producen resultados notables, bien sea porque sobresalen por encima -las mejores prácticas-, bien sea porque sobresalen por debajo -las situaciones catastróficas que valdría la pena evitar-. Finalmente, en temas específicos habría que tener en cuenta los consensos o los criterios de aquellos expertos que, por su vinculación al tema y a experiencias exitosas dentro de ese tema, tienen mucho que aportar, así sus afirmaciones sean todavía tentativas y no puedan aducir todas las veces una fundamentación empírica cuantitativa completa que las soporte.

VI. Para terminar: la calidad y la organización del servicio educativo colombiano

La organización del sistema educativo colombiano con miras a desarrollar su calidad, se ha ido perfilando alrededor de tres polos institucionales con funciones muy específicas:

- El Ministerio de Educación, que produce unos estándares nacionales de calidad, en el sentido de unas metas mínimas, claras y verificables que expresan los consensos, tanto nacionales como internacionales, sobre los resultados deseables del aprendizaje y de la convivencia escolar.
- Los planteles educativos (oficiales o no) que, coordinados y apoyados por las autoridades educativas locales y regionales, definen e implementan,
 - Lo que se va a enseñar y las reglas básicas de convivencia escolar,
 - La manera como van a hacerlo (metodología de la docencia y de la convivencia),
 - La verificación permanente (evaluación) de lo que se va logrando, tanto en términos de conocimiento como de convivencia.
- El ICFES, que evalúa de manera masiva, periódica y consistente los resultados de la actividad educativa de los planteles en sus alumnos; es decir, la calidad de la educación.

Este esquema institucional puede funcionar y sería bueno para el fortalecimiento de la calidad de la educación en Colombia si se cumplen, al menos, ciertas condiciones básicas que se pueden resumir en las tres siguientes: a) si hay coordina-

ción y al mismo tiempo autonomía entre las tres instancias, b) si hay claridad con respecto a la calidad de la educación y c) si hay consistencia en la estructura planteada. En particular:

a) Coordinación y al mismo tiempo autonomía de las instancias:

- Es función exclusiva del Ministerio la producción de unos estándares, como objetivos nacionales mínimos y medibles en conocimiento y convivencia. Los estándares deben tener difusión suficiente entre los involucrados de las otras dos instancias, especialmente los maestros de los planteles. No basta con "colgar" unos estándares en una página Web, cuando la mayoría de las escuelas del país ni tienen acceso a Internet, ni disponen de impresoras o fotocopiadoras. La difusión masiva de los estándares por el medio impreso parece que sigue siendo ineludible, si no se quiere que se conviertan en unos documentos cuasi-clandestinos, como sucedió con los currículos. Los estándares deben consultar tanto los consensos internacionales en materia de currículo escolar como la realidad nacional, sin llegar nunca al parroquianismo académico: el conocimiento es cada vez más universal.
- Los planteles deben tener plena autonomía para definir lo que van a trabajar (currículo y convivencia), la manera como van a hacerlo (metodología) y los procesos de verificación permanente de lo que van logrando (evaluación), con el compromiso de que ello vaya orientado a cumplir, como mínimo, con los estándares nacionales. Para ello deben contar con el

apoyo (financiero, logístico y pedagógico) de las autoridades educativas locales.

- El ICFES, como entidad especializada en la evaluación masiva, es autónomo para definir las técnicas y procedimientos de la misma. No es autónomo para definir lo que se va a evaluar (los resultados concretos del trabajo educativo de los colegios), ni el marco de referencia contra el cual va a evaluar (los estándares publicados por el Ministerio). En pocas palabras, el ICFES ni debe inmiscuirse en el currículo o la metodología de enseñanza, ni en los estándares. Es decir, sólo es una entidad técnica evaluadora, que debe diseñar sus pruebas, aplicarlas, analizarlas y publicar sus resultados, y que tiene la responsabilidad de adecuarlas cada día más para que verifiquen, de la manera más acertada y comparable posible, el cumplimiento en los colegios de los estándares. Sin embargo, aunque es autónomo en producir los resultados (de las evaluaciones), no es autónomo en definir sus características: los resultados de las evaluaciones tienen tres tipos de destinatarios, con necesidades específicas y no siempre coincidentes: el Ministerio, para definir sus políticas educativas; la escuela y sus docentes, para mejorar sus resultados; y la sociedad en general, última veedora y supervisora de las políticas públicas y de sus resultados.

b) Claridad con respecto a la calidad de la educación:

En dos aspectos fundamentales las tres instancias tienen que tener claridad para que el esquema funcione:

- Hay dos tipos de evaluación, con dos alcances perfectamente diferentes. El primero se refiere a la evaluación como proceso permanente de acompañamiento a la labor educativa, que es responsabilidad ineludible de la comunidad educativa, de cada docente en particular, en cada institución; se trata del monitoreo permanente que debe acompañar a cualquier actividad humana intencional y sistemática: ¿cómo hacemos para saber si lo que estamos haciendo está dando resultados, si estamos logrando lo que pretendíamos? La segunda es la evaluación masiva de tipo diagnóstico, que se realiza periódicamente con el fin de calibrar los resultados del sistema, de comparar los de los diversos subsistemas locales, de visualizarlos frente al panorama internacional y de observar su evolución a lo largo del tiempo. La evaluación masiva detecta situaciones y resultados, especialmente relativos al conocimiento, de políticas educativas en conjunto; pero la evaluación acompañante de la docencia verifica y se informa sobre procesos y actitudes, tanto cognitivos como comportamentales, que se escapan a la masiva. El sistema nacional de educación necesita de ambas evaluaciones, que no se superponen porque tienen objetivos y razones de ser diferentes.
 - El Ministerio, el ICFCES y, sobre todo, los colegios, deben tener en claro las diferencias entre estándares, evaluación y calidad de la educación. Los estándares, por sí mismos, no garantizan la calidad, solo ayudan a tener metas comunes y mínimas. La evaluación tampoco es en sí calidad, sólo verifica en qué medida nos estamos acercando o no a la calidad deseada (Lucio, 2002). La calidad, querámoslo o no, se produce en los colegios, con la interacción de convivencia y de conocimiento que éstos propician entre los alumnos y entre alumnos y docentes. La razón de ser de los estándares y de la evaluación masiva radica en ayudar a que se produzca o se incremente en los colegios esa calidad. Esto es importante porque muchas veces, con intenciones de propaganda política que es fácil entender, se proclaman los avances en la producción de estándares o en la evaluación masiva como logros en la calidad, cuando son sólo instrumentos que pueden ayudar a promoverla, si van acompañados con acciones reales destinadas a mejorar lo que se hace en los colegios. Si no se da este segundo componente, es lo mismo que creer que los casos de fiebre elevada se solucionan solamente con tener buenos termómetros y saber medir bien la temperatura corporal del paciente.
- c) Consistencia en la estructura planteada:
- Las políticas educativas deben ser políticas de Estado, y no se pueden estar modificando en cada gobierno los esquemas de funcionamiento, ni arrancar de cero con cada nuevo ministro o secretario de educación. Esto quiere decir que, para que el sistema funcione, tiene que garantizarse cierta consistencia en el tiempo. En particular, es necesario que no se cambien cada año los estándares (como pasó con los estándares publicados en 2002 y 2003); que no se vayan improvisando cada vez diferentes sistemas de evaluación masiva, que a la larga hacen imposible la verificación a lo largo del tiempo del comportamiento de la calidad (como está pasando con las cuatro

tandas de aplicación del SABER); que los colegios tengan la posibilidad de aplicar correctivos y redireccionamientos a lo que hacen (en términos de docencia y de convivencia), orientados por los estándares y por los resultados observados en

sus alumnos con respecto a esos estándares con cierta consistencia en el tiempo, lo que significa consistencia en el mejoramiento de sus recursos institucionales (tanto físicos como financieros y humanos).

Bibliografía

Fuentes "oficiales" consultadas de los cinco sistemas de evaluación masiva de la educación

ICFES, Instituto Colombiano para el Fomento de la Educación Superior. Exámenes de Estado.

(1974-2003) Los resultados de las pruebas de estado han sido tomados directamente de las bases de datos suministradas por dicha entidad.

(2003) Pruebas completas aplicadas en septiembre, con sus respectivos resultados.

ICFES (2000), *Exámenes del ICFES. Evolución histórica 1968-1998*.

Cuchimaque, Ernesto (1998), *Exámenes de Estado para el Ingreso a la Educación Superior. Pruebas de Lenguaje*.

González, Myriam et al. (1998), *Examen de Estado para ingreso a la educación superior. Pruebas de ciencias naturales*.

Rocha, Martha Cecilia (1998), *Exámenes de Estado para ingreso a la educación superior. Pruebas de Matemáticas*.

Romero, Víctor Manuel (1998), *Exámenes de Estado para ingreso a la educación superior. Pruebas de Ciencias Sociales*.

Varios Autores (2002), *ICFES 2002: La prueba*. Bogotá, Casa editorial El Tiempo [estos materiales no son publicación oficial del ICFES; sin embargo, a juicio de funcionarios del Instituto son una muestra bastante aproximada de las pruebas e ítemes que se aplican a partir del 2000]:

Vol. 1: Lenguaje e Inglés

Vol. 2: Historia y Geografía

Vol. 3: Matemáticas y Física

Vol. 4: Química y Biología

Vol. 5: Filosofía y Pruebas interdisciplinarias

Vol. 6: Orientación Profesional

SABER, Sistema Nacional de Evaluación de la Educación

www.icfes.gov.co (enero 2003), resultados de las evaluaciones nacionales de 1997-1999.

Bases de datos de las aplicaciones en Bogotá (1998-1999) y nacional (2002-2003).

Sistema Nacional de Evaluación de la Educación (1997a), *Evaluación de logros, áreas de lenguaje y matemáticas, resultados en grados 3, 5, 7, 9. 1992-1994*.

Sistema Nacional de Evaluación de la Educación (1997b), *Evaluación de logros en lenguaje, lineamientos teóricos, pruebas de 3, 5, 7, 9. 1992-1994*.

Sistema Nacional de Evaluación de la Educación (1997c), *Evaluación de logros en matemáticas, lineamientos teóricos, pruebas de 3, 5, 7, 9. 1992-1994*.

Sistema Nacional de Evaluación de la Educación (1997d), *Factores asociados al logro cognitivo de los estudiantes, grados 3, 5. 1993-1994*.

Ministerio de Educación (2002), "Evaluación censal de calidad".

Programa Nuevo Sistema Escolar (2002), cuadernillos para pruebas de matemáticas y lenguaje para los grados 5 y 9, aplicadas en 2002.

Sistema Nacional de Evaluación de la Educación (2002), cuadernillos para pruebas de matemáticas y lenguaje para los grados 3, 5, 7 y 9, aplicadas en octubre de 2002.

SABER/ICFES (2003), cuadernillos para pruebas de matemáticas y lenguaje para los grados 5 y 9, aplicadas en 2003.

Barón et al. (2003a), *Una mirada a los fundamentos e instrumentos de matemáticas 2002-2003*.

Barón et al. (2003b), *Matemáticas escolares: aportes para orientar procesos de innovación*.

Pérez, Mauricio (2003a), *Una mirada a los fundamentos e instrumentos de lenguaje 2002-2003*.

Pérez, Mauricio (2003b), *Leer y escribir en la escuela: algunos escenarios pedagógicos y didácticos para la reflexión*.

ICFES (2005), "Evaluación censal: lenguaje, matemática, ciencias naturales, ciencias sociales, competencias ciudadanas".

ICFES (2006), "Pruebas saber 2005: estructuras de prueba".

TIMSS, Tercer Estudio Internacional de Matemáticas y Ciencias

Díaz, Carlos Jairo y Soto, Carlos Alberto (1997), *Diseño, metodología y resultados generales -TIMSS- Colombia*. Bogotá, Ministerio de Educación Nacional.

Díaz, Carlos Jairo y Rivera, Hernando (1997), *Manual de habilidades en ciencias y matemáticas -TIMSS- Colombia*. Bogotá, Ministerio de Educación Nacional.

Díaz, Carlos Jairo et al. (1997a), *Análisis y resultados de las pruebas de ciencias -TIMSS- Colombia*. Bogotá, Ministerio de Educación Nacional.

Díaz, Carlos Jairo et al. (1997b), *Análisis y resultados de las pruebas de matemáticas -TIMSS- Colombia*. Bogotá, Ministerio de Educación Nacional.

Martin, Michael O. et al. (1999), *School Contexts for Learning and Instruction*. TIMSS International Study Center, Boston College, U.S.A.

- Martin, Michael O. et al. (2000), *Effective Schools in Science and Mathematics*. TIMSS International Study Center, Boston College, U.S.A.
- LLECE, Laboratorio Latinoamericano de Evaluación de Calidad de la Educación**
- Bases de datos originales, facilitadas por Héctor Fernández, representante en Colombia del Laboratorio.
- LLECE (1998), *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado*. Santiago, UNESCO.
- LLECE (2000), *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grados de la educación básica. Segundo informe*. Santiago, UNESCO.
- LLECE (2001), *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grados de la educación básica. Informe técnico*. Santiago, UNESCO.
- LLECE (2002), *Estudio cualitativo de escuelas con resultados destacados en siete países latinoamericanos*. Santiago, UNESCO.
- PISA 2000**
- USA, National Center for Education Statistics (2001), *Outcomes of Learning. Results from the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy*. Washington. [resultados internacionales, con énfasis en los resultados de USA frente a los demás países industrializados]
- (2003a), *Literacy Skills for the World of Tomorrow - Further results from PISA 2000*. OECD/UNESCO. [resultados internacionales, publicación oficial]
- (2003b), *Aptitudes básicas para el mundo de mañana. Otros resultados del proyecto Pisa 2000*. Resumen ejecutivo. OECD/UNESCO.
- PISA 2003**
- (2004a), *Learning for Tomorrow's World - First Results from Pisa 2003*. OECD.
- (2004b), *Problem Solving for Tomorrow's World. First Measures of Cross-Curricular Competencies from PISA 2003*. OECD.
- (2004c), *First Results from Pisa 2003: Executive Summary*. OECD.
- Otras fuentes consultadas**
- Aebli, Hans (1991), *Factores de la Enseñanza que Favorecen el Aprendizaje Autónomo*. Madrid, Narcea.
- Arnold, Karl-Heinz (2001), "Schulleistungenstudien und soziale Gerechtigkeit" [estudios de logro escolar y equidad social], en *Zeitschrift für Pädagogie*, no. 2, pp. 161-177.
- Blatchford, Peter y Martin, Clare (1998), "The effects of class size on classroom processes: 'it's a bit like a treadmill - working hard and getting nowhere fast!'" en *British Journal of Educational Studies*, vol 46, n. 2, June, pp. 118-137.
- Bogotá, Secretaría de Educación (1999), *Hacia una cultura de la evaluación*. Memorias del tercer foro educativo distrital.
- Camargo, Marina (2000), "Estudios de caso sobre la gestión escolar". Informe final de investigación. Bogotá, CORPOEDUCACIÓN.
- Cepeda, Edilberto (s.f.), "Factores asociados al logro cognitivo en matemáticas". Mimeo. Bogotá, Universidad de los Andes, Departamento de matemáticas.
- Fuhr, Thomas (2002), "Das Glück des Kindes" [la felicidad de los niños], en *Zeitschrift für Pädagogie*, no. 48, pp. 514-533.
- Hart, Susan (1998), "A sorry tail: ability, pedagogy and educational reform", en *British Journal of Educational Studies*, vol 46, n. 2, June, pp. 153-168.
- Keeves, John P., ed. (1988), *Educational Research, Methodology and Measurement: an International Handbook*. Oxford, Pergamon Press.
- Lucio, Ricardo (2002), "Estándares académicos en el contexto escolar colombiano, algunas reflexiones" [ponencia leída en el lanzamiento oficial de la primera versión de los estándares escolares]. Santa Marta.
- Melo, Jorge Orlando (1998), "Bibliotecas y educación", en *Revista del Banco de la República* (Bogotá), vol. 71, no. 852 (octubre), pp. 24-37.
- Melo, Jorge Orlando (2001), "Mensaje de error: la educación superior y las bibliotecas", en *El Malpensante* (Bogotá), no. 32 (agosto-septiembre), pp. 25-41.
- Ministerio de Educación Nacional (2002), *Estándares para la excelencia en la educación. Estándares curriculares para las áreas de matemáticas, lengua castellana y ciencias naturales y educación ambiental para la educación preescolar, básica y media*. Bogotá.
- Ministerio de Educación Nacional (2003), *Estándares básicos de calidad en matemáticas y lenguaje*. Bogotá.
- Ministerio de Educación Nacional (2004a), *Estándares básicos de competencias ciudadanas*. Bogotá.

- Ministerio de Educación Nacional (2004b), *Estándares básicos de competencias en ciencias naturales y ciencias sociales*. Bogotá.
- Sarmiento, Alfredo *et al.* (2000), "Evaluación de la calidad de la educación primaria y secundaria en Santafé de Bogotá. Factores asociados al logro académico". Mimeo. Bogotá, Misión Social.
- Stenhouse, L. (1988), "Case study methods", en Keeves (1988), pp. 49-53.
- Tenjo, Jaime *et al.* (1999), "Logros cognitivos y factores asociados al logro: análisis de las pruebas de matemáticas y lenguaje para tercero y quinto de primaria 1997-1998". Informe final de investigación. Bogotá, Universidad Javeriana, Facultad de Economía.
- Uribe, Claudia (2003), "Growing Gains or Growing Pains? The Potential Impact of School Expansion Policies on Student Achievement, in Bogotá, Colombia". Tesis doctoral. Harvard University, Graduate School of Education.
- Vasco, Carlos Eduardo *et al.* (2001), *El saber tiene sentido. Una propuesta de integración curricular*. Bogotá, CINEF.
- Velandia, Wilson (1978), *Structural determinants of verbal and science achievement analysis of a large sample of colombian college applicants*. Tesis doctoral. The University of Connecticut.