

Sobre la difusión de resultados por escuela

Felipe Martínez Rizo*

CUADERNO No. 15



Instituto Nacional para la
Evaluación de la Educación

**COLECCIÓN CUADERNOS
DE INVESTIGACIÓN**

ISSN 1665-9457

*Director General del Instituto Nacional para la Evaluación de la Educación

Este texto puede consultarse en: www.inee.edu.mx

MÉXICO, JULIO, 2005

CONTENIDO

Planteamiento de la cuestión	3
Postura del INEE	5
Consideraciones sustantivas	7
Consideraciones jurídicas	19
Conclusiones	23
Anexo	25

PLANTEAMIENTO DE LA CUESTIÓN

La creación del Instituto Nacional para la Evaluación de la Educación (INEE) ha representado el inicio de una nueva fase en lo que se refiere a las actividades de evaluación educativa desarrolladas en México, desde hace varias décadas, especialmente por parte de la Secretaría de Educación Pública (SEP). En esta nueva etapa se busca mejorar cualitativamente no sólo la calidad técnica de las evaluaciones, sino también la amplitud y oportunidad de la difusión de los resultados, indispensables para que la evaluación cumpla su propósito último: contribuir a la mejora de la calidad educativa.

El documento *Política Nacional de Evaluación Educativa* contiene valiosas orientaciones en lo que se refiere a la difusión de resultados. Este documento fue aprobado por la Junta Directiva del INEE, el 14 de octubre de 2003, con base en las atribuciones que confiere al Instituto su Decreto de Creación, y presentado entonces al Secretario de Educación para su posible adopción con carácter normativo para el conjunto del sistema educativo nacional. Se transcribe a continuación el apartado relativo al tema que ahora interesa.

Lineamientos para la difusión de resultados de la Política Nacional de Evaluación

La información que resulte de las evaluaciones de la educación nacional deberá hacerse pública de manera transparente y oportuna, en beneficio de la sociedad en general y de los sectores con especial interés en ella: maestros, padres de familia, empresarios, autoridades educativas, legisladores y estudiosos, entre otros. Para ello:

- ◆ *Los resultados de la evaluación estarán disponibles en formatos dirigidos a diversos públicos usuarios de la misma, con el propósito de facilitar su aprovechamiento y su adecuada interpretación. La información tendrá diversos niveles de complejidad, agregación y especificidad según el público al que vaya dirigida. Se promoverá el desarrollo de una cultura de la evaluación en todos los sectores de la sociedad.*
- ◆ *Para propiciar una correcta interpretación, la difusión de las evaluaciones deberá incluir información del contexto que corresponda, facilitando la formulación de juicios ponderados y permitiendo un trato justo y equitativo a los diversos sectores sociales, en particular de los grupos vulnerables o en situación de desventaja.*
- ◆ *Las instancias responsables de llevar a cabo las evaluaciones deberán difundirlas en los tiempos más cercanos a su realización que técnicamente sea posible, para asegurar su ade-*

cuando aprovechamiento, tanto en el mejoramiento de la calidad de la educación como en la rendición de cuentas.

- ◆ *La difusión de resultados de las evaluaciones educativas respetará el principio de protección de la privacidad de las personas que establece la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental.*
- ◆ *Se promoverá la realización de estudios de opinión para contrastar los resultados de evaluaciones objetivas de la calidad educativa con la percepción que tengan de ella los diferentes sectores de la sociedad.*
- ◆ *Se difundirá también información suficiente sobre los criterios y procedimientos metodológicos que sustenten las evaluaciones, para que especialistas independientes puedan valorar su consistencia.*
- ◆ *Las bases de datos derivadas de las evaluaciones estarán a disposición de investigadores y analistas, en formatos que faciliten su análisis y manejo, al tiempo que aseguren la privacidad de los evaluados.*

La difusión de resultados desagregados por escuela

Los esfuerzos de difusión de resultados emprendidos por la SEP y el INEE han sido recibidos con interés por las autoridades, la comunidad educativa, y la sociedad en general. Los medios de comunicación les han prestado especial atención. El creciente interés por parte de los diversos actores interesados se ha traducido en una mayor demanda para que los resultados se difundan de la manera más amplia, poniéndose al alcance no sólo de autoridades educativas, sino también de toda persona interesada en ellos.

En lo que se refiere al grado de detalle que deberá tener la difusión de resultados, parece haber consenso en dos puntos: deben difundirse a nivel nacional y estatal; además de que no deben proporcionarse por alumnos individuales. No hay consenso, en cambio, sobre la conveniencia de dar resultados escuela por escuela. El punto de vista de algunos padres de familia y otros sectores sociales suele privilegiar la importancia de hacerlo, para apoyar la rendición de cuentas, mientras el sector magisterial suele verlo con reticencia.

A favor de la difusión de resultados por escuela se aduce el argumento de que, para propósitos de mejora, es insuficiente dar resultados a nivel nacional o estatal, pues, son demasiado generales. En relación con los padres de familia que tienen a sus hijos en escuelas privadas, se señala que conocer los resultados por escuela es necesario para sustentar la decisión de cuál institución seleccionar para enviar a sus hijos. En el caso de quienes tienen a sus hijos en escuelas públicas, en especial cuando no hay posibilidades reales de elección, se señala también que sólo conociendo los resultados de las mismas, podrán los padres de familia exigir de las autoridades el mejoramiento de la calidad y la corrección de las deficiencias.

Desde el otro punto de vista se subrayan los efectos negativos inherentes a la difusión de resultados de evaluación por escuela, si no se consideran todos los aspectos que pueden influir de manera relevante, lo que puede llevar a consecuencias no equitativas, especialmente para las escuelas que atienden a alumnos de un medio menos favorable.

Los partidarios de la difusión de resultados por escuela sospechan que la postura del sector magisterial es defensiva y puede encubrir deficiencias que conviene conocer para corregirlas; quienes comparten la perspectiva magisterial tienden a considerar que la postura opuesta desconoce la complejidad del trabajo escolar y puede tener un impacto negativo en él. Los dos puntos de vista anteriores, presentados en forma simplificada, tienen en común, explicablemente, la consideración de los aspectos técnicos especializados.

En opinión del INEE, para decidir si es adecuado o no ofrecer resultados desagregados en el nivel de escuela deben tenerse en cuenta, en primer lugar, las características técnicas de la evaluación, que se derivan, a su vez, del propósito que persiguen. Estas características técnicas deben considerarse al evaluar la necesidad, implicaciones y consecuencias de cierto tipo de difusión de resultados. Con base en estas consideraciones se presenta la siguiente:

POSTURA DEL INEE

El INEE considera que la decisión de difundir o no resultados de evaluaciones desagregados para cada escuela debe tomarse con base en las características técnicas de las evaluaciones, derivadas a su vez de los propósitos de cada evaluación. Cuando se pretende evaluar escuelas en lo individual, los instrumentos y las aplicaciones deben reunir características precisas que hacen posible el cumplimiento de ese propósito; si se pretende en cambio evaluar al sistema educativo como tal, los instrumentos y la aplicación deberán tener otras características que harían inadecuado el uso de los resultados para un propósito diferente.

El cuadro siguiente resume las características de cuatro tipos de evaluaciones, señalando el propósito de cada una, y la consecuencia de lo apropiado o no de la difusión de los resultados de cada una en forma desagregada por escuela. La primera columna del cuadro se refiere a las pruebas del INEE, cuyas características, derivadas de su propósito, hacen inapropiada la difusión de resultados por escuela. En las páginas siguientes se explican con detalle las razones técnicas que sustentan esta postura. El Instituto espera que el conocimiento de estas razones sea el punto de partida para el consenso sobre un tema tan importante.

CARACTERÍSTICAS DE ALGUNAS EVALUACIONES E IMPLICACIONES PARA LA DIFUSIÓN DE RESULTADOS POR ESCUELA

	CASO A	CASO B	CASO C	CASO D
	Pruebas del INEE	EXANI I y II	Pruebas del Factor Aprovechamiento Escolar Carrera Magisterial	Pruebas de algunas entidades de los Estados Unidos
PROPÓSITO	Evaluar el sistema educativo como tal	Escoger alumnos de nuevo ingreso	Asignar estímulos a maestros	Evaluar escuelas
IMPACTO	Bajo tanto sobre individuos como sobre escuelas	Alto sobre individuos	Alto sobre individuos	Alto sobre escuelas.
INSTRUMENTOS	Pruebas alineadas al currículo en áreas selectas, criteriosales y matriciales	Pruebas de habilidades, no alineadas al currículo, normativas, no matriciales	Pruebas alineadas al currículo, en las materias enseñadas por los maestros a evaluar; pueden ser normativas y matriciales o no	Pruebas alineadas al currículo en áreas selectas, criteriosales, de preferencia no matriciales
APLICACIÓN	Muestra de escuelas y de alumnos en cada escuela, en grados clave	A todos los aspirantes a ingresar al programa o institución	Censo o muestra suficiente de alumnos de todos los maestros que se registren	Todos los alumnos de las escuelas a evaluar, todos los grados
PERIODICIDAD	Ciclos multi- anuales	Cada ingreso	Intervalos según estímulos	Preferentemente anuales
COMPARABILIDAD EN EL TIEMPO	Necesaria	No relevante	No relevante	Necesaria
INFORMACIÓN DE CONTEXTO	Indispensable	No relevante, salvo si es criterio de admisión	Indispensable	Indispensable
ORDENAMIENTO	Sólo en grupos amplios	Indispensables individuales	En grupos amplios	En grupos amplios
OTROS ASPECTOS	Medida de ganancia deseable	Medida de ganancia no procedente	Medida de ganancia necesaria	Medida de ganancia necesaria
DIFUSIÓN DE RESULTADOS POR ESCUELA	No apropiada	No apropiada	No apropiada	Apropiada

CONSIDERACIONES SUSTANTIVAS

El fundamento de la postura del INEE radica en el hecho de que las evaluaciones cuyo propósito es evaluar un gran conjunto como el sistema educativo de un país o una entidad federativa, arrojan resultados con márgenes de error considerables en el nivel de desagregación de escuela singular, como consecuencia de las decisiones que deben tomarse al diseñar y aplicar los instrumentos. Tal margen de error en el nivel de escuela no se debe a deficiencias de las pruebas, sino a decisiones técnicas dado el propósito para el que fueron concebidas y desarrolladas las evaluaciones.

Además de las consideraciones técnicas anteriores, el INEE señala que la difusión de resultados de las pruebas en un nivel de desagregación para el cual no fueron diseñadas, puede traer consigo dos consecuencias negativas importantes:

- ◆ Dar lugar a interpretaciones incorrectas sobre escuelas singulares, que pueden ser injustas;
- ◆ Reducir significativamente la confiabilidad de los resultados de las pruebas en el nivel nacional y subnacional para el que fueron desarrolladas.

La complejidad de aspectos, propósitos y acercamientos evaluativos

Las decisiones relativas a evaluación educativa deben tener en cuenta, al menos, tres puntos:

- ◆ Que los sistemas educativos modernos son realidades complejas, conformadas por una multiplicidad de componentes que ninguna mirada singular puede captar de manera integral. Ejemplos de esos componentes son: los millones de alumnos, cada uno de los cuales tiene un contexto familiar y social propio; los miles de maestros y escuelas, cada uno con circunstancias igualmente peculiares; las múltiples áreas del currículo, cada una con contenidos variados, potencialmente inagotables.
- ◆ Que una evaluación puede diseñarse con diversos propósitos. Entre otros: decidir si la tarea realizada en casa la tarde anterior por un niño debe ser calificada como sobresaliente, satisfactoria o insuficiente; si un alumno debe ser promovido o no al grado siguiente al fin del curso; definir al ganador de un concurso de aprovechamiento académico; si un maestro merece o no un estímulo especial por su trabajo; si los resultados globales de un sistema educativo nacional o subnacional son superiores o inferiores a los de otro; si hay o no avances en el nivel educativo de un alumno, de un grupo, de una escuela o de un sistema educativo a lo largo del tiempo; la selección de alumnos de nuevo ingreso mediante la aplicación de un examen de admisión; la acreditación de un programa o institución, etcétera.
- ◆ Que hay una gran diversidad de acercamientos y métodos de evaluación, con alcances y limitaciones. Algunos ejemplos de acercamientos son la calificación de tareas escolares que hacen los maestros, o la valoración que éstos hacen sobre la medida en que sus

alumnos han comprendido o no una explicación; la aplicación de una prueba realizada por el propio maestro al fin de un bimestre; la evaluación del trabajo de los docentes que hacen los directores y supervisores; la aplicación de pruebas estandarizadas a todos los alumnos de un estado, entre otros.

Para opinar sobre la pertinencia de un acercamiento evaluativo es sustancial definir cuál es el aspecto preciso del sistema educativo a evaluar, y qué se quiere hacer con los resultados, o sea cuál es el propósito de la evaluación. Una vez precisados el objeto y el propósito de una evaluación será posible seleccionar el acercamiento más apropiado, precisando sus alcances y limitaciones, con lo que habrá una base adecuada para decidir el tipo, más pertinente, de difusión de los resultados.

Las evaluaciones según su impacto y su escala

Una distinción importante, relacionada con el propósito de las evaluaciones, se efectúa entre evaluación de impacto alto o bajo:

- ◆ Una evaluación de alto impacto es aquella cuyos resultados tienen consecuencias fuertes para los sujetos evaluados, como las que sirven para decidir si un alumno es promovido al siguiente grado o no; las de selección para ingresar a una licenciatura; las que determinan la contratación o la promoción de un maestro; las que sirven para acreditar un programa o institución.
- ◆ Las evaluaciones de bajo impacto, en cambio, no tienen consecuencias fuertes, como sucede con los diagnósticos o las evaluaciones llamadas formativas, que informan al maestro sobre la situación de un alumno en un momento dado para sustentar el trabajo subsecuente.

Otro aspecto de gran importancia es la escala de la aplicación:

- ◆ Las evaluaciones en pequeña escala involucran pocos sujetos, como el caso típico de las que hace un maestro de los alumnos de un grupo, que no suelen ser más de dos o tres decenas.
- ◆ Las evaluaciones en gran escala involucran a un número mucho mayor de sujetos. No es fácil precisar el umbral a partir del cual pueden considerarse tales, ya que esta distinción, más que dicotómica, remite a un continuo. Sin embargo, cuando se trata de la evaluación de decenas o centenas de miles de alumnos, no puede haber duda de que se está ante casos de evaluaciones en gran escala.

La evaluación del aprendizaje de los alumnos en pequeña escala

La tarea de evaluar de manera precisa el grado en que un alumno ha desarrollado los conocimientos y habilidades previstos al final de un ciclo escolar no es sencilla, si se quiere cubrir de manera suficiente las diversas materias o áreas del currículo y los temas de cada

área o materia. Si además se desea conocer el avance del alumno la tarea se complica, ya que se deberá evaluar al inicio y al fin del ciclo escolar. Si se quiere conocer con más detalle el desarrollo de un alumno en intervalos menores (mensual o incluso semanalmente), la tarea se vuelve mucho más compleja.

Si en lugar de un alumno se trata de dos o tres decenas, y además se quiere tener información sobre las circunstancias personales, familiares y sociales de cada uno de ellos, para tenerla en cuenta en el momento de tomar decisiones importantes para el futuro de cada uno de ellos, la tarea se antoja difícil. Eso, nada menos, se espera de los maestros, y es crucial para que el trabajo educativo produzca buenos resultados. Para retroalimentar su propio trabajo docente, así como el esfuerzo de los alumnos, es fundamental que el maestro conozca con precisión el avance de cada uno de ellos. Por lo anterior, la calidad de un sistema educativo se basa en última instancia en el profesionalismo de sus maestros que, además de dominar los contenidos a enseñar y los métodos pedagógicos necesarios para ello, deben también ser capaces de manejar técnicas de evaluación apropiadas para el trabajo en el aula, que les proporcionen la información necesaria para retroalimentar los procesos de conocimiento.

La experiencia muestra que lo anterior no es sencillo, pero sí posible: la evaluación que hace un buen maestro del avance de los alumnos del grupo a su cargo tiene niveles de validez y confiabilidad suficientes para sustentar las decisiones educativas más delicadas.

La evaluación del aprendizaje de los alumnos en gran escala

La pregunta clave para el tema que nos ocupa es si es posible evaluar el aprendizaje con validez y confiabilidad, igualmente suficientes, mediante acercamientos en gran escala.

Si se busca obtener el grado de detalle que permiten las evaluaciones en el aula, la respuesta a la pregunta anterior es clara: esto no es posible con las metodologías de evaluación en gran escala disponibles en la actualidad. Las evaluaciones de aprendizaje en gran escala pueden, en cambio, dar información de calidad suficiente sobre conjuntos grandes de alumnos, en ciertas áreas del currículo y con intervalos de tiempo amplios.

Algunas características de las evaluaciones del aprendizaje en gran escala

La mayoría de las evaluaciones en gran escala que se aplican en la actualidad utilizan pruebas estandarizadas compuestas por preguntas de opción múltiple. Se usan también, aunque con frecuencia mucho menor, instrumentos menos estructurados: preguntas de respuesta construida, ejercicios con problemas o situaciones reales, así como evaluaciones orales y observaciones de las ejecuciones de los evaluados.

El uso de pruebas de opción múltiple en evaluaciones en gran escala no es accidental: aunque tienen limitaciones para la evaluación de niveles cognitivos complejos, su viabi-

lidad para aplicaciones a números grandes de sujetos es mucho mayor que el de acercamientos menos estructurados como los mencionados. El uso de acercamientos alternativos aun con pocos cientos de alumnos (por no hablar de decenas o centenares de miles) implicaría muchas horas de trabajo de un gran número de evaluadores especializados, lo que los hace inviables para su uso en gran escala.

El desarrollo de instrumentos de buena calidad basados en preguntas de opción múltiple es también laborioso, pero una vez desarrolladas, las pruebas pueden aplicarse a miles de sujetos en forma controlada, lo que se traduce en costos unitarios menos elevados. Este tipo de evaluaciones tiene, sin embargo, características que deben tenerse en cuenta para entender para qué tipo de propósitos pueden usarse en forma adecuada.

Aplicaciones censales y muestrales

Un punto importante a considerar es el relativo al número de sujetos a los que puede aplicarse una prueba estandarizada. La disyuntiva es entre una aplicación censal, realizada a todos los sujetos que integran el universo a evaluar, y una aplicación muestral, que se hace a un subconjunto de ese universo. El sentido común sugiere que, si hay recursos suficientes, será preferible una aplicación censal a una muestral: aunque se apliquen correctamente los principios del muestreo moderno, ninguna muestra es perfectamente representativa del universo del que se extrae, y los resultados de aplicaciones muestrales tienen siempre un margen de error (el *error de muestreo*).

Es importante añadir que el error de muestreo puede estimarse con precisión, y que un diseño cuidadoso permite obtener resultados con una precisión considerable a partir de muestras de un tamaño muy inferior al del universo, especialmente cuando se trata de universos con millones de sujetos, como el que constituyen los alumnos del sistema educativo mexicano.

El desconocimiento de las posibilidades del muestreo moderno, hace que el sentido común tienda a preferir las aplicaciones censales. Se piensa también que la diferencia de costos entre éstas y otras muestrales no debe ser excesiva, pues no se entiende que en el caso de universos grandes la diferencia del tamaño de una buena muestra y la del conjunto total es enorme. En el sistema educativo mexicano, por ejemplo, una muestra de menos de 10,000 alumnos puede arrojar resultados bastante precisos y confiables sobre el conjunto de más de dos millones de alumnos de un grado escolar determinado, si no se desean resultados por entidad federativa o modalidad de escuela. La diferencia de costos es, pues, claramente considerable.

A lo anterior se debe añadir el hecho de que *mientras mayor es una aplicación más aumentan los errores que no se derivan del muestreo*, que también afectan de manera importante la calidad de los resultados. Los errores que no se derivan del muestreo: incluyen la no participación en la aplicación de ciertos sujetos por razones no aleatorias, como su

residencia en lugares poco accesibles; menor disponibilidad o interés por participar; variación en las condiciones de aplicación; e incluso sesgos o fraudes propiamente dichos, como ayuda del maestro a los alumnos para que respondan correctamente, copia de unos alumnos a otros, y exclusión de alumnos de bajo rendimiento.

Una muestra cuidadosamente diseñada puede controlar en forma suficiente el error de muestreo y es afectada por otros tipos de error en un grado mucho menor que una aplicación censal. Por ello los resultados de aplicaciones muestrales son finalmente de calidad igual o mejor que los de aplicaciones censales. Añadido a la diferencia de costos lo anterior explica el que los sistemas de evaluación educativa en gran escala utilicen casi siempre muestras y no censos. Las excepciones son pocas, y corresponden a sistemas educativos de dimensiones menores a las del sistema educativo mexicano.

El universo de los contenidos a evaluar y el tamaño de las pruebas

Un segundo aspecto, igualmente mal entendido por el sentido común, se relaciona con el muestreo de los contenidos a evaluar. Además del gran número de alumnos involucrados, la evaluación de aprendizajes tiene que ver, en efecto, con otro universo de grandes dimensiones: el de los contenidos curriculares que los alumnos deben manejar.

El currículo de la educación básica puede verse en principio inagotable, ya que incluye al conjunto de las ciencias naturales y sociales, las humanidades y las artes, la tecnología y los conocimientos prácticos, así como los aspectos actitudinales, valorales y afectivos, además de los cognoscitivos y de habilidades.

El currículo de la primaria y la secundaria, en concreto, comprende de seis a doce áreas, según se agrupen en materias o grandes campos; cada materia o área, a su vez, incluye no menos de tres o cuatro subáreas temáticas, lo cual nos lleva a un total de veinte o treinta áreas temáticas particulares.

Algunas de las áreas del currículo, como la educación artística o la valoral, así como los niveles cognitivos superiores de todos los campos, presentan mayor dificultad para evaluarse adecuadamente con preguntas de opción múltiple, por lo que las pruebas utilizadas en las aplicaciones en gran escala se limitan, por lo general, a las subáreas temáticas de algunas de las áreas curriculares principales, como las de lenguaje, matemáticas, ciencias naturales y, con menor frecuencia, ciencias sociales.

Ahora bien: una pregunta de opción múltiple se refiere necesariamente a un solo tema o aspecto muy particular de cierta área y subárea curricular. Por ello una prueba formada por este tipo de preguntas debe incluir un número relativamente grande de ellas, si quiere cubrir razonablemente el universo de contenidos correspondiente.

Estamos ante otro caso de tensión entre un acercamiento censal y uno muestral: se trata ahora del muestreo de los contenidos a evaluar mediante la prueba de que se trate. Si se

quiere que una prueba valore si los alumnos manejan *todos* los contenidos de cierta área curricular, se necesitaría que tuviera un número grande de preguntas, provocando una aplicación muchas horas, a lo largo de varios días.

La confiabilidad de una prueba tiene que ver, entre otras cosas, con el número de preguntas que incluye. Si se midiera el conocimiento de los alumnos en cierta área curricular mediante una sola pregunta, los resultados tendrían un gran error de medida. Para reducir el error de medida a niveles razonables generalmente se considera que son necesarias unas treinta preguntas para cada tema o subárea. Si se evalúan tres o cuatro áreas (digamos español, matemáticas, ciencias naturales y ciencias sociales) cada una con tres o cuatro subáreas de interés, una prueba deberá tener cuatrocientas o quinientas preguntas (y requerir de ocho a diez horas de aplicación) para que los resultados tuvieran márgenes de error aceptables. Una prueba de esas dimensiones debe aplicarse en varios días, lo que hace crecer de manera no controlable el riesgo de errores que no son de muestreo, como el derivado de que un número considerable de los alumnos que comienzan a contestar la prueba el primer día no acuden el siguiente, especialmente chicos de bajo rendimiento.

Por lo anterior se ha desarrollado una estrategia alternativa, utilizada en los sistemas de evaluación más avanzados: se trata de las pruebas llamadas matriciales, basadas en un diseño de bloques incompletos. Esta modalidad consiste en dividir la prueba total en varias *subpruebas o cuadernillos*, cada uno de los cuales está integrado por un número menor de preguntas, que puede ser respondido por un alumno en una sola sesión. Las preguntas que forman cada cuadernillo se seleccionan y combinan según su contenido y características psicométricas (grado de dificultad, poder de discriminación, entre otras). Luego los cuadernillos se aplican a diferentes alumnos, en forma también cuidadosamente controlada.

Utilizando este tipo de pruebas, los resultados de cada alumno según sus respuestas a las preguntas del cuadernillo correspondiente tienen un margen de error considerable, que impide extraer conclusiones confiables sobre la medida en que cada uno maneja los contenidos de *toda la prueba*. Sin embargo, si estas pruebas se diseñan y aplican correctamente, sus resultados podrán ofrecer información sobre el grado en que *el conjunto de los alumnos* maneja el *conjunto de los contenidos curriculares* correspondientes, con una precisión y márgenes de error controlados.

Características adecuadas para evaluaciones con diferentes propósitos

Para entender mejor las evaluaciones del INEE, es útil comparar su propósito con el de evaluaciones con pretensiones diferentes.

En el caso de evaluaciones cuyo propósito es seleccionar alumnos para ingresar a un programa de licenciatura, es adecuado utilizar un instrumento que permita comparar a los aspirantes entre sí, en aquellos aspectos que puedan predecir mejor el rendimiento futuro

en el programa. Por ello las pruebas de selección son generalmente normativas, y se limitan a la valoración de habilidades generales, como la capacidad de razonamiento verbal o matemático o la de comprensión lectora.

En ese caso no es necesario medir el nivel de competencia de los aspirantes en todas las áreas del currículo de los niveles educativos anteriores. Sí es indispensable, en cambio, evaluar a todos y cada uno de los aspirantes en la forma más confiable posible, pues deberán tomarse decisiones que afectarán a cada uno de manera importante.

Cuando el propósito es valorar la medida en que el conjunto de los alumnos de un sistema educativo han alcanzado efectivamente los objetivos establecidos en el plan de estudios, como es el caso de las evaluaciones del INEE, entonces deben tomarse decisiones diferentes: por una parte, es necesario que los instrumentos incluyan un conjunto de preguntas representativas del conjunto de los contenidos curriculares, en las áreas que se decida evaluar; por otra parte, no será necesario evaluar a todos y cada uno de los alumnos, será suficiente hacerlo con una muestra representativa. Por ello las pruebas del INEE se definen como alineadas al currículo, criterios, y matriciales, y se aplican a muestras nacionales de escuelas y alumnos.

Cuando el propósito sea evaluar escuelas singulares, el diseño de la evaluación deberá tener características diferentes. La doble problemática del muestreo de alumnos y de contenidos que implica evaluar en gran escala el aprendizaje tiene consecuencias importantes para la evaluación de escuelas.

Puede pensarse que si se evalúa el aprendizaje de los alumnos de una escuela, se tendrá todo lo necesario para evaluar al plantel mismo. La realidad es más compleja. Un análisis cuidadoso hace ver que, para evaluar adecuadamente una escuela debería contarse con más elementos de juicio incluyendo, entre otros, los siguientes:

- ◆ Resultados confiables de evaluaciones del aprendizaje alcanzado por los alumnos de todos los grupos y grados, en todas las materias o áreas curriculares, y no sólo en los niveles cognitivos más simples, sino también en los superiores, así como en los aspectos no cognitivos.
- ◆ Información sobre las condiciones de los alumnos al ingresar a la escuela, así como sobre la proporción que permanece en el plantel hasta el final del grado o nivel y la que lo abandona antes de terminar.
- ◆ Resultados de evaluaciones realizadas no sólo al final de un ciclo escolar, sino también al inicio del mismo, para valorar el avance alcanzado.
- ◆ Resultados de varios ciclos, para valorar la consistencia de los resultados y su tendencia, y no sólo de un año, que puede ser excepcional.
- ◆ Información sobre el entorno de la escuela, para distinguir la influencia del contexto en los resultados de lo que deba atribuirse a la escuela misma.

Como se ha dicho antes, el desarrollo y la aplicación de pruebas estandarizadas en gran escala implican procesos laboriosos que no es fácil repetir en un mismo ciclo escolar. Por la naturaleza de estos instrumentos, tampoco es posible aplicarlos en todas las materias o áreas curriculares, ni en todos los aspectos cognitivos y no cognitivos.

Las pruebas del INEE no se aplican a todos los alumnos de cada escuela seleccionada en la muestra, porque además de elevar los costos de aplicación, no daría resultados valiosos en la escuela, ni mejoraría la calidad a nivel nacional, por las siguientes razones:

- ◆ Las pruebas desarrolladas por el INEE para su aplicación al final del ciclo escolar 2004-2005 tienen un diseño matricial, con 28 y 24 cuadernillos diferentes para primaria y secundaria, respectivamente. Por el diseño matricial de las pruebas, aplicarlas a la totalidad de los alumnos de las escuelas incluidas en la muestra permitiría dar resultados confiables sólo para los planteles más grandes donde se obtendrían respuestas para todos los cuadernillos. Esas escuelas son una pequeña minoría, por lo que su número en la muestra siempre será muy reducido.
- ◆ Aún en escuelas grandes las pruebas no se aplican en todos los grados, y las evaluaciones no informan sobre otros aspectos del funcionamiento de las escuelas. Por lo tanto los resultados de un grado, aún siendo confiables en sí mismos, no son suficientes para formular juicios sólidos sobre la calidad de la escuela en cuestión.
- ◆ Para optimizar el tamaño de la muestra, *en función del propósito de evaluar el sistema educativo*, es mucho mejor aumentar el número de escuelas que el número de alumnos a evaluar en cada escuela.

La experiencia de otros países

Quienes se expresan a favor de la difusión de resultados por escuela, en ocasiones aducen a favor de su postura la experiencia de otros países, como Estados Unidos, Inglaterra y Chile. En los dos últimos países se difunden públicamente ordenamientos de escuelas (*rankings, league tables*) basados en los resultados de evaluaciones en gran escala. Por otro lado, en muchos estados norteamericanos existen sistemas accesibles vía Internet en los que se pueden ver los resultados de cada una de las escuelas del sistema estatal correspondiente. La experiencia internacional, sin embargo, es más compleja, e incluye elementos que llaman la atención sobre el cuidado que debe tenerse al difundir resultados por escuela:

- ◆ Países que han obtenido resultados excelentes en pruebas internacionales, como Finlandia, Holanda o Irlanda, descartan la difusión de ordenamientos de escuelas y la consideran indeseable para propósitos de mejora educativa.
- ◆ En Estados Unidos los sistemas de evaluación que difunden resultados por escuela son siempre de nivel estatal, escala en la que se cuenta con evaluaciones de todas las escuelas correspondientes (*es decir, evaluaciones censales*). El programa *National Assessment*

of Education Progress (NAEP), que evalúa *muestras* nacionales de escuelas, prohíbe dar resultados por escuela, que se consideran confidenciales, al igual que los resultados de alumnos en lo individual, como puede verse en la siguiente cita, tomada de la Guía oficial del NAEP para informar a los padres de familia sobre las características del sistema:

No information on individual students, families, or schools is included in NAEP data releases or reports. No records are kept that connect students' names with the answers they gave on the assessment. In accordance with federal law, all personally identifiable information regarding NAEP participants is confidential. Any individual who violates the confidentiality of NAEP data may be charged with a felony, fined, and/or imprisoned. No such violation of NAEP confidentiality restrictions has ever occurred.

Durante los primeros veinte años del NAEP, que comenzó a funcionar en 1969, ni siquiera se daban resultados por entidad federativa y la participación de cada entidad era voluntaria. Sólo a partir de 2005, con base en la Ley *No Child Left Behind*, de 2002, es obligatoria la participación de todos los estados federales.

- ◆ En Chile, la constatación de que la difusión de ordenamientos de escuelas, junto con otras medidas como la municipalización, la privatización de muchas escuelas o los incrementos salariales al magisterio, no han producido mejoras significativas de la calidad del sistema educativo y, al parecer, han acentuado las desigualdades, ha llevado a proponer modificaciones a las evaluaciones que, sin abandonar el principio de aplicaciones censales en grados clave cada dos años, las complementen con evaluaciones muestrales que permitan apreciar de manera más confiable si hay o no avances en la calidad del conjunto.
- ◆ En Inglaterra, las evaluaciones en que se basan los ordenamientos han comenzado a complementarse recientemente (a partir de 2003) con mediciones del avance de cada escuela (*valor agregado y/o ganancia*), en lugar de la comparación de resultados obtenidos en un solo momento.
- ◆ También en Inglaterra, las evaluaciones externas, como las del programa PISA de la OCDE, están sujetas al párrafo 2a del *National Statistics Code of Practice, Protocol on Data Access and Confidentiality*, que establece: *A guarantee of confidentiality will be given, at the time of collection, to respondents to National Statistics surveys and censuses.* Con base en la misma norma, las escuelas pueden negarse a ser evaluadas, lo que ocurre con frecuencia, dado el rechazo que evaluaciones frecuentes producen entre los maestros; como consecuencia de ello, Inglaterra no consiguió cubrir los requisitos de la muestra mínima estipulada para la aplicación 2003 de PISA, por lo que sus resultados no fueron incluidos en el reporte de la OCDE.

Propósitos y características de las evaluaciones del INEE y difusión de los resultados

El diseño de las pruebas a cargo del INEE se explica a partir de su propósito que consiste en evaluar al sistema educativo nacional como tal, en conjunto, así como los grandes subsistemas de las entidades federativas y los tipos o modalidades de las escuelas consideradas.

Para cumplir este propósito, el INEE ha desarrollado pruebas matriciales, alineadas al currículo, basadas en un muestreo de los contenidos curriculares, criterios, las cuales se aplican a subconjuntos representativos de las escuelas y alumnos de todo el país. Estas características permiten a las evaluaciones del INEE ofrecer resultados confiables sobre el grado en que el conjunto de los alumnos de las escuelas del sistema educativo nacional y sus grandes subsistemas domina el conjunto de los contenidos del currículo.

Aunque las evaluaciones del INEE no ofrezcan resultados confiables de alumnos o escuelas en lo individual, sus resultados pueden ser valiosos para las autoridades educativas, ya que les aportan elementos importantes para sustentar decisiones de política educativa a escala nacional y regional, así como para los maestros, a quienes pueden servir para re-actualizar su práctica docente. Los resultados son también relevantes para la sociedad en general, como muestra el interés que provocan, al proporcionarles elementos para sustentar su legítima exigencia de rendición de cuentas, igualmente en la escala nacional y regional. En particular los resultados de las evaluaciones del INEE pueden aportar información valiosa sobre aspectos como los siguientes:

- ◆ Las entidades o subsistemas que obtienen resultados por encima o por debajo de lo que sus condiciones socioeconómicas harían esperable.
- ◆ Cuáles entidades presentan distribuciones de resultados más o menos equitativas en los logros educativos, teniendo en cuenta el perfil social de su población.
- ◆ En qué modalidades educativas los alumnos obtienen mejores resultados, en igualdad de condiciones socioeconómicas.
- ◆ Cómo evolucionan los logros de los alumnos en las diversas áreas evaluadas a lo largo del tiempo, en las entidades federativas, los sectores sociales y las modalidades de escuelas.
- ◆ Las áreas del currículo en las que la mayoría de los alumnos obtiene resultados bajos, para que se revisen tanto los planes y programas como las estrategias docentes.
- ◆ Qué segmentos de la población tienen mayores dificultades en relación con ciertas partes del currículo, así como en cuáles entidades parece haber mayor dificultad en determinados sectores de la población.
- ◆ Qué aspectos de la formación inicial y permanente de los docentes deberían reforzarse para atender las áreas problemáticas detectadas.

Con base en los alcances y límites de las evaluaciones que lleva a cabo, se derivan de su diseño y, en última instancia, del propósito para el cual se desarrollaron, el INEE considera

que sus resultados no deben usarse para sustentar decisiones sobre alumnos o escuelas en lo individual, y no deben difundirse en forma desagregada a esos niveles.

Esta opinión no se debe a una postura de oposición a la transparencia de la información pública. Por el contrario, el INEE considera la difusión de los resultados parte fundamental de su misión, indispensable para que la evaluación contribuya a la mejora de la calidad educativa. La gran cantidad de acciones de difusión que el Instituto emprendió desde sus inicios es la mejor prueba de la seriedad de esta postura.

Una evaluación válida debe arrojar resultados confiables para alcanzar sus propósitos. En ese caso la información será un sustento sólido para las decisiones que se basen en ella y deberá difundirse amplia y oportunamente. Pero si por su diseño una evaluación no arroja información confiable sobre ciertos niveles, ésta no deberá difundirse con ese grado de desagregación porque, lejos de ser sustento sólido para las decisiones respectivas, inducirá a error a los usuarios.

No se trata de temer que una información precisa lleve a decisiones que no parezcan adecuadas a algunos actores del sistema educativo: cuando se cuente con información confiable, ésta deberá difundirse, y la responsabilidad de las decisiones corresponderá a la autoridad respectiva. Se trata en cambio, como se ha explicado, de un caso en que la información no es confiable al nivel de las escuelas porque se diseñó para proporcionar información en otro nivel de agregación. Por ello, el INEE considera inadecuado el uso de los resultados de sus evaluaciones para sustentar decisiones o juicios sobre escuelas individuales. Esto no debe propiciarse, ya que las evaluaciones no reúnen las condiciones necesarias para cumplir adecuadamente con un propósito distinto de aquel para el cual fueron diseñadas.

Algunos ejemplos permitirán precisar esta idea y mostrar que el riesgo de interpretaciones erróneas no es imaginario. El margen de error de los puntajes obtenidos en las pruebas nacionales frecuentemente es mayor que la distancia existente entre los resultados promedio de una entidad federativa de los de otra, sobre todo cuando se trata de escuelas de una misma modalidad. Por ello un ordenamiento simple, que defina 32 lugares para otras tantas entidades, es engañoso, la distancia entre muchas entidades es tan pequeña que hay una probabilidad considerable de que el orden real sea diferente al que definen los promedios. El INEE recomienda por ello utilizar agrupaciones de rendimiento más o menos alto. En el caso de las entidades federativas, en lugar de 32 posiciones es más razonable hablar de tres a cinco grupos de estados, definidos según se encuentren más o menos por encima o debajo de la media nacional.

Evidentemente, en evaluaciones que involucran a decenas de miles de escuelas y centenares de miles de maestros, los márgenes de error no pueden ni remotamente definir ordenamientos a nivel de escuela o maestro individual. Es seguro que habrá varios miles de escuelas y muchos miles de maestros cuyo desempeño no difiere estadísticamente uno

del otro. Por ello hablar del mejor maestro o la mejor escuela no tiene sustento en la información disponible, aun sin tener en cuenta la influencia de factores externos a la escuela, y la posibilidad de errores no derivados del muestreo.

Un caso ejemplar de esta problemática son las evaluaciones del Factor Aprovechamiento Escolar del programa de Carrera Magisterial. Como la participación en ellas es voluntaria, el conjunto de los participantes en un año determinado no es ni un censo ni una muestra aleatoria. Además, la exclusión de algunos maestros que reciben estímulos con base en otros criterios, como los de escuelas indígenas, hace que la proporción de los maestros de cada entidad que participan en el programa no sea comparable.

Por todo ello los resultados que, junto con otros factores, pueden ser adecuados para sustentar decisiones sobre la asignación de estímulos a maestros en lo individual, no lo son para comparar resultados agregados al nivel de las entidades federativas. Esto generará confusión, sobre todo cuando los resultados no coinciden con los de las pruebas de estándares nacionales, lo cual es de esperarse dadas las diferencias de propósitos y características.

Las diversas evaluaciones y el Sistema Nacional de Evaluación Educativa

Que los resultados relativos al nivel nacional no sustenten decisiones sobre alumnos o escuelas en lo individual no implica que éstas sean irrelevantes para esos niveles, ni que no se puedan diseñar evaluaciones apropiadas para la toma de decisiones sobre escuelas o alumnos. La evaluación del avance de cada alumno es fundamental; cada maestro la hace habitualmente, y así deberá seguir ocurriendo, si bien es deseable que su calidad mejore.

La evaluación de las escuelas individuales es también importante, como argumentan quienes desean que se difundan resultados de las evaluaciones del INEE escuela por escuela. El INEE considera factible y necesario diseñar e implementar evaluaciones con características técnicas adecuadas para tener resultados confiables sobre diversos aspectos del desempeño de las escuelas en lo individual. Por el tamaño del sistema educativo nacional, el Instituto opina también que el nivel adecuado para desarrollar esas evaluaciones es el de las entidades federativas, sus regiones, y las zonas escolares.

Dada la complejidad que enfrentan las evaluaciones educativas, no es posible que una sola instancia se encargue de toda la gama de acercamientos necesarios para tener elementos suficientes para sustentar las diversas decisiones que deben tomarse. Por ello lo deseable es construir un verdadero Sistema Nacional de Evaluación Educativa, integrado por la SEP, las áreas de evaluación de las entidades federativas, el INEE y otras instancias, que trabajen de manera articulada, en estrecha colaboración.

Aunque la evaluación en gran escala del aprendizaje de los alumnos que hace el INEE no puede sustentar decisiones sobre escuelas o alumnos en lo individual, sus resultados

pueden y deben aprovecharse para dar referentes a las evaluaciones de otros niveles, validarlas globalmente, y enriquecer el trabajo docente y el de gestión en todos los niveles del sistema educativo. La misión del INEE incluye también contribuir a desarrollar la capacidad técnica de evaluación en todas las entidades federativas y, en general, colaborar con cualquier instancia que tenga propósitos similares a los suyos. Ese es el papel que el Instituto considera que debe jugar en el marco del auténtico Sistema Nacional de Evaluación Educativa que México necesita.

CONSIDERACIONES JURÍDICAS

Las disposiciones de la Ley de Transparencia y Acceso a la Información

Dado lo reciente de la utilización de sistemas masivos de evaluación educativa en la mayor parte de los países del mundo, las legislaciones no suelen contemplar la difusión de resultados por escuela de manera precisa y es difícil establecer cuáles son las mejores prácticas en la materia. México no es excepción en este sentido, por ello parece necesario que se precisen directrices claras al respecto. El documento antes mencionado de Política Nacional de Evaluación podría enriquecerse en este sentido antes de elevar su carácter normativo.

La normatividad vigente sobre evaluación educativa (Ley General de Educación, Acuerdo N° 200) no contiene disposiciones que contemplen de manera precisa la cuestión, por lo que es necesario remitirse a otros ordenamientos.

La Ley Federal de Acceso a la Información Pública Gubernamental (LFAIPG), define como pública toda la información que no sea confidencial, o que se clasifique expresa y justificadamente como reservada.

La LFAIPG define como confidencial la información relativa a personas, pero se refiere expresamente sólo a las personas físicas y no a las morales, como son las escuelas u otras instituciones:

ARTÍCULO 3. *Para los efectos de esta Ley se entenderá por:*

II. Datos personales: La información concerniente a una persona física, identificada o identificable, entre otras la relativa a su origen étnico...

Los artículos 4, 6 y 7 de la LFAIPG definen claramente la orientación de la nueva normatividad, en el sentido de privilegiar la apertura, de suerte que todo lo que no se defina expresamente como confidencial o reservado se entenderá que debe ser público:

ARTÍCULO 4. *Son objetivos de esta Ley:*

- I. Proveer lo necesario para que toda persona pueda tener acceso a la información mediante procedimientos sencillos y expeditos;
- II. Transparentar la gestión pública mediante la difusión de la información que generan los sujetos obligados;
- III. Garantizar la protección de los datos personales en posesión de los sujetos obligados;
- IV. Favorecer la rendición de cuentas a los ciudadanos, de manera que puedan valorar el desempeño de los sujetos obligados...

ARTÍCULO 6. *En la interpretación de esta Ley se deberá favorecer el principio de publicidad de la información en posesión de los sujetos obligados.*

ARTÍCULO 7. *Con excepción de la información reservada o confidencial prevista en esta Ley, los sujetos obligados deberán poner a disposición del público y actualizar... entre otra, la información siguiente:*

XVI. *Cualquier otra información que sea de utilidad o se considere relevante, además de la que con base a la información estadística responda a las preguntas hechas con más frecuencia por el público.*

El último párrafo de este artículo contiene un punto de especial relevancia para el tema de la difusión de resultados desagregados al nivel de las escuelas individuales.

La información a que se refiere este Artículo deberá publicarse de tal forma que facilite su uso y comprensión por las personas y que permita asegurar su calidad, veracidad, oportunidad y confiabilidad.

La postura del INEE sobre la difusión de resultados por escuela se relaciona en forma directa con el párrafo anterior. Debido a que las características técnicas de las evaluaciones en gran escala que aplica el Instituto no permiten tener resultados confiables en el nivel de las escuelas y los alumnos en lo individual, éstas no deben usarse para sustentar decisiones en esos niveles ni deben difundirse con ese nivel de desagregación.

Los artículos 13 y 14 definen cuál es la información reservada. De interés para el INEE es la fracción V del artículo 13:

ARTÍCULO 13. Como información reservada podrá clasificarse aquella cuya difusión pueda:

V. *Causar un serio perjuicio a las actividades de verificación del cumplimiento de las leyes...*

Con base en la fracción anterior, el INEE ha clasificado como reservada la información de sus bancos de reactivos así como sus pruebas, pues su difusión haría imposible la tarea de evaluar el rendimiento de los alumnos del sistema educativo en forma comparable a lo largo del tiempo.

Para poder contar con pruebas equiparadas es necesario que no se difundan los reactivos comunes a utilizar en varias ocasiones. Por ello el INEE ha reservado la información mencionada y liberará reactivos (y, en su caso, pruebas) que hayan cumplido su vida útil en la medida en que no se afecte la posibilidad de hacer comparaciones en el tiempo.

Los artículos 13 y 14 de la LFAIPG no incluyen otros elementos que puedan aplicarse a la información que maneja el INEE.

El artículo 18 de la LFAIPG retoma la noción de información confidencial, en dos fracciones:

ARTÍCULO 18. *Como información confidencial se considerará:*

- I. *La entregada con tal carácter por los particulares a los sujetos obligados...*
- II. *Los datos personales que requieran el consentimiento de los individuos para su difusión...*

El Capítulo IV de la Ley se dedica a la protección de esos datos personales, para cuya difusión se requiere la autorización del sujeto involucrado, salvo excepciones que precisa el artículo 22:

ARTÍCULO 22. No se requerirá el consentimiento de los individuos para proporcionar los datos personales en los siguientes casos:

- I. *Los necesarios para la prevención o el diagnóstico médico...*
- II. *Los necesarios por razones estadísticas, científicas o de interés general previstas en la Ley, previo procedimiento por el cual no puedan asociarse los datos personales con el individuo a quien se refieran;*
- III. *Cuando se transmitan entre sujetos obligados o dependencias y entidades, siempre y cuando los datos se utilicen para el ejercicio de facultades propias de los mismos;*
- IV. *Cuando exista una orden judicial;*
- V. *A terceros, cuando se contrate la prestación de un servicio que requiera el tratamiento de datos personales. Dichos terceros no podrán utilizar los datos personales para propósitos distintos a aquellos para los cuales se les hubieren transmitido, y*
- VI. *En los demás casos que establezcan las leyes.*

El reglamento de la LFAIPG precisa algunos de los puntos anteriores, en especial:

ARTÍCULO 37. La información confidencial no estará sujeta a plazos de vencimiento y tendrá ese carácter de manera indefinida, salvo que medie el consentimiento expreso del titular de la información o mandamiento escrito emitido por autoridad competente.

ARTÍCULO 40. Para que las dependencias o entidades puedan permitir el acceso a información confidencial requieren obtener el consentimiento expreso de los particulares titulares de la información, por escrito o medio de autenticación equivalente.

La normatividad relativa a la información estadística

Las prácticas de países en los que la evaluación educativa, al igual que los principios democráticos, tienen una larga tradición, coinciden en considerar confidencial la información derivada de estudios estadísticos que contienen datos sobre personas no sólo físicas, sino también morales: en el caso que nos ocupa, las escuelas. Es el caso de los Estados Unidos e Inglaterra, como se ha visto antes. El mismo criterio aplica la OCDE en lo que se refiere a las pruebas del programa PISA.

Este criterio es congruente con los *Principios Fundamentales de la Organización de las Naciones Unidas sobre Estadísticas Oficiales*, cuyo texto completo puede verse en el Anexo 2, y de los que se destaca el Principio 6, que establece inequívocamente que:

Los datos individuales colectados por agencias estadísticas para compilaciones estadísticas, sea que se refieran a personas físicas o morales, deberán ser estrictamente confidenciales y se usarán exclusivamente para propósitos estadísticos.

El instrumento jurídico mexicano que parece más pertinente para aplicar en el caso que nos ocupa es la Ley de Información Estadística y Geográfica (LIEG), vigente desde 1980, que contiene disposiciones precisas, congruentes con el principio anterior de la ONU y similares a las normas inglesas a que se ha hecho referencia.

A continuación se transcriben pasajes relevantes de la LEIG:

ARTÍCULO 8. *A efecto de que la Secretaría establezca y opere un sistema de identificación nacional para fines estadísticos, las unidades económicas, empresas y establecimientos industriales, comerciales, agropecuarios, forestales y pesqueros; los dedicados a la producción o venta de bienes y servicios; las sociedades y asociaciones civiles, así como las demás instituciones públicas, sociales y privadas con fines no lucrativos y **las docentes** y culturales estarán obligadas a inscribirse en los registros que para tales fines lleve la propia Secretaría...*

ARTÍCULO 36. *Serán considerados informantes de los sistemas nacionales:*

- I. Las personas físicas y morales, cuando les sean solicitados datos estadísticos y geográficos por las autoridades competentes.
- II. *Las unidades económicas, empresas y establecimientos...; las sociedades, asociaciones civiles y las instituciones sociales o privadas con fines no lucrativos, **las docentes** y culturales que estén obligadas a inscribirse o a proporcionar datos en los registros administrativos a que se refiere el artículo 8° de este ordenamiento.*

ARTÍCULO 38. *Los datos e informes que los particulares proporcionen para fines estadísticos o provengan de registros administrativos o civiles, serán manejados, para efectos de esta Ley, bajo la observancia de los principios de confidencialidad y reserva y no podrán comuni-*

carce, en ningún caso, en forma nominativa o individualizada, ni harán prueba ante autoridad administrativa o fiscal, ni en juicio o fuera de él. Cuando se deba divulgar la información estadística, ésta no podrá referirse, en ningún caso, a datos relacionados con menos de tres unidades de observación y deberá estar integrada de tal manera que se preserve el anonimato de los informantes.

En el caso de informantes a los que se refiere la fracción II del artículo 36, sólo podrá difundirse información respecto de tres o más unidades de observación localizadas dentro de una misma rama o actividad económica, entidad federativa, municipio, nivel de ingreso, o de cualquier otro indicador estratificado.

Según la LIEG, pues, las escuelas son *unidades de información*, a las que debe aplicarse la disposición de que no deberá divulgarse información sobre alguna en lo particular, ni tampoco en la forma de desagregaciones que hagan que la información corresponda a menos de tres *unidades de información*.

En forma análoga a las disposiciones sobre confidencialidad del Código Inglés antes mencionado, el Art. 39 de la LIEG establece también la obligación de que las personas a quienes se les requieran datos estadísticos o geográficos sean informadas previamente del carácter obligatorio o potestativo de sus respuestas; de las consecuencias de la falsedad de sus respuestas; de la posibilidad del ejercicio del derecho de rectificación; la confidencialidad de la información estadística que proporcionen; la forma en que será divulgada o suministrada...

CONCLUSIONES

- ◆ Las características de las evaluaciones del aprendizaje de alumnos que desarrolla y aplica el INEE, en forma congruente con su propósito de evaluar al sistema educativo en su conjunto, y no a las escuelas ni los alumnos en lo individual, hacen que los resultados no deban usarse para sustentar decisiones en esos niveles y que no sea adecuado difundirlos en ellos.
- ◆ Los elementos del marco jurídico nacional, junto con los derivados de principios internacionales y de la práctica de los organismos más reconocidos, llevan a concluir que los resultados de escuelas individuales en las evaluaciones son confidenciales, al igual que los de personas físicas. Por ello esa información sólo podría difundirse públicamente con el consentimiento de los involucrados, a quienes debería informárseles sobre sus derechos al momento de recabar la información. En el caso de personas físicas, se requeriría el consentimiento del evaluado o de quien tenga la patria potestad. Por lo que se refiere a escuelas singulares, se requeriría la autorización de la autoridad educativa.
- ◆ Aun si se obtuviera el consentimiento de los involucrados no sería adecuado difundir los

resultados, por las razones técnicas presentadas en este documento.

- ◆ La difusión de resultados por escuela puede afectar la confiabilidad de las evaluaciones en gran escala del INEE. Actualmente, las evaluaciones del INEE son de bajo impacto para las escuelas y los alumnos, ya que no tienen consecuencias directas para unas y otros. Esa misma característica hace que no sean vistas con temor y favorece una actitud positiva frente a ellas, lo que contribuye a su confiabilidad en el nivel en que se pretende la tengan: el del sistema educativo como tal y sus grandes subsistemas.

En su forma actual las evaluaciones del INEE pueden ser de gran valor para el sistema educativo nacional. Los avances conseguidos por el Instituto en menos de tres años de operación así lo indican. Salvaguardar la calidad de estas evaluaciones importa también para fortalecer la capacidad de la sociedad para exigir de sus gobernantes el cumplimiento de compromisos, lo cual se vería impedido por interpretaciones erróneas y distorsionadas por las propias autoridades o los medios de comunicación. Si se garantiza su confiabilidad, tanto las decisiones de mejora y búsqueda de la equidad, como las exigencias de rendición de cuentas por parte de la sociedad, tendrán mejor sustento.

El riesgo de distorsión se entiende pensando que, si las evaluaciones del INEE se difundieran en forma desagregada a nivel de la escuela individual se transformarían en evaluaciones de alto impacto para las escuelas. Esto podría afectar seriamente la calidad de las evaluaciones del INEE en los niveles que pretenden cubrir y para los que han sido diseñadas, pues se propiciaría la tendencia entre las escuelas de enseñar para la prueba o incluso prácticas fraudulentas que buscarían elevar artificialmente los resultados. Las evaluaciones se volverían menos confiables para su propósito original, y atenderían mal otro para el que no fueron diseñadas.

Por todo lo anterior, la postura del INEE es que, con base en las características de sus pruebas, en forma congruente con las mejores prácticas del campo en el nivel internacional, y en forma compatible con una lectura cuidadosa del marco jurídico nacional, no debe darse difusión de los resultados en forma desagregada en los niveles de escuela y alumno.

ANEXO

THE UNITED NATIONS FUNDAMENTAL PRINCIPLES OF OFFICIAL STATISTICS

The Statistics Commission,

- ◆ Bearing in mind that official statistical information is an essential basis for development in the economic, demographic, social and environmental fields of mutual knowledge and trade among States and peoples of the world,
- ◆ Bearing in mind that the essential trust of the public in official statistics, and thus the quality of the information available to the Government, the economy, and the public depends largely on the cooperation of citizens, enterprises and other respondents in providing appropriate and reliable data needed for necessary statistical compilations and on the cooperation between users and producers of statistics in order to meet users' need,
- ◆ Recalling the efforts of governmental and non-governmental organizations active in statistics to establish standards and concepts to allow comparisons among countries,
- ◆ Recalling also the Internat. Statistical Inst. Declaration of Professional Ethics,
- ◆ Having expressed the opinion that resolution C (47) adopted by the Economic Commission for Europe on 15 April 1992 is of universal significance,
- ◆ Noting that at its 8th session held in Bangkok in Nov. 1993 the Working Group of Statistical Experts, assigned by the Committee on Statistics of the Economic and Social Commission for Asia and the Pacific to examine the Fundamental Principles, has agreed in principle to the ECE version, and had emphasized that those principles are applicable to all nations,
- ◆ Noting also that, at its eight session, held at Addis Ababa in March 1994, the Joint Conference of African Planners, Statisticians and Demographers, considered that the Fundamental Principles of Official Statistics are of universal significance,

Adopts the present principles of official statistics:

1. Official statistics provide an indispensable element in the information system of a society, serving the government, the economy and the public with data about the economic, demographic, social and environment situation. To this end, official statistics that meet the test of practical utility are to be compiled on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.
2. To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and pre-

sentation of statistical data.

3. To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.
4. The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.
5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on the respondents.
6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.
7. The laws, regulations and measures under which the statistical system operates are to be made public.
8. Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.
9. The use by statistical agencies in each country of the international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.
10. Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.