



Instituto Nacional para la  
Evaluación de la Educación

# Jornadas de Evaluación Educativa

México, D.F. 14 y 15 de julio de 2004



## MEMORIA



## CONTENIDO

---

Palabras del Dr. Reyes Tamez Guerra al inaugurar las <i>Jornadas de Evaluación Educativa</i>	5
Presente y Futuro de la Evaluación Educativa en México	11
El Sistema de Información para la Mejora Educativa	45
Los Indicadores del INEE: <i>Panorama Educativo 2003</i>	71
Comparabilidad de los Resultados de las Evaluaciones	93
Las Comparaciones Entre Entidades: Alcances y Límites de los <i>Ranking</i>	127





**Palabras del Dr. Reyes  
Tamez Guerra al inaugurar  
las *Jornadas de Evaluación  
Educativa***





## **Intervención del doctor Reyes Tamez Guerra, Secretario de Educación Pública, al inaugurar las *Jornadas de Evaluación Educativa***

**M**uy buenos días. Me da mucho gusto saludarlos y ver la respuesta que se obtuvo a la convocatoria para la realización de estas Jornadas de Evaluación Educativa. Me produce gran satisfacción que se hayan superado las expectativas, porque demuestra el interés que existe en todo el país por la evaluación, y es muy importante avanzar y consolidar esta cultura, lo cual es fundamental para el desarrollo del sistema educativo del país.

La educación es un factor fundamental para el mejoramiento individual y colectivo, por ello la evaluación está presente en las políticas educativas de casi todo el mundo; una buena evaluación es requisito indispensable para dotar a la población de una educación de calidad, logrando este objetivo será posible fortalecer la calidad de los programas y las escuelas.

La existencia del Instituto Nacional para la Evaluación de la Educación, encarna la realización de un proyecto que comenzó a gestarse en el equipo de transición del Presidente Vicente Fox y quedó de manifiesto en el Programa Nacional de Educación 2001-2006, como una pieza importante en la consolidación del sistema nacional de evaluación.

Para mejorar el nivel educativo de los mexicanos, es necesario contar con evidencias precisas acerca de la situación que guarda el servicio educativo, es decir, detectar las deficiencias que impiden al Sistema Educativo Nacional avanzar hacia el camino de la calidad.

El Instituto tiene la misión, como lo hemos visto ya, de monitorear la calidad de la educación que se imparte en el país y divulgar con transparencia sus diagnósticos. Estas tareas las realiza mediante la evaluación de muestras representativas de estudiantes y escuelas de todo el país y proporcionando además un conjunto amplio y sólido de indicadores educativos.

Para lograr sus propósitos, en el INEE se han integrado una Junta Directiva plural, un Consejo Técnico con expertos mexicanos y extranjeros de reconocido prestigio en el ámbito de la evaluación, quienes garantizan la solidez técnica de los procesos evaluativos, y un Consejo Consultivo con los representantes de las Áreas de Evaluación de todas las entidades federativas, para asegurar la pertinencia y operatividad de la evaluación.

Los diversos actores que confluyen en la actividad educativa: profesores, padres de familia, directivos escolares, supervisores, autoridades locales y autoridad federal, necesitan contar con datos e información válida y confiable para apoyar la toma de decisiones.

Es importante resaltar que para el gobierno federal, en especial para la Secretaría de Educación Pública y para los usuarios y beneficiarios del sistema educativo, es muy importante contar con un Instituto Nacional para la Evaluación de la Educación técnicamente vigoroso, un Instituto del cual, en el mediano plazo, se pueda decir que en México estamos orgullosos



por su calidad de trabajo y por la transparencia de las evaluaciones que realiza.

En noviembre de 2003, el INEE entregó resultados de sus primeras evaluaciones y diagnósticos, y lo seguirá haciendo cada vez con mayor calidad y precisión.

Con la ayuda de un buen número de expertos nacionales y extranjeros, se está creando una nueva generación de instrumentos de medición y diagnóstico los cuales darán mayor solidez al trabajo del Instituto.

El desarrollo de una nueva generación de instrumentos de medición representará un largo proceso de trabajo, el cual permitirá contar con mediciones precisas y abiertas al escrutinio de todos.

Los diferentes actores educativos están involucrados en la acción de evaluar, pero el INEE tiene la misión de mostrarnos una visión general y panorámica de la situación de la educación en el país.

Como lo ha comentado el licenciado Martínez Rizo, el INEE no debe ni puede hacer toda la evaluación en el país; todos los organismos que aquí se han mostrado han tenido históricamente un papel importante en la evaluación, la misma Secretaría y autoridades estatales deben participar en una evaluación que, como se ha comentado, esté articulada: que cada una de las evaluaciones cumpla con el propósito fundamental de proporcionarnos información valiosa, la cual nos permita una mejor toma de decisiones. Lo mismo a nivel internacional, deseo decirles que en la reunión de Ministros de Educación que convocó la OCDE, celebrada hace tres meses en Dublín, Irlanda, la posición transmitida como país es: *el reto de la evaluación es lograr, además, que la evaluación realizada en cualquiera de los niveles, se traduzca en recomendaciones concretas de mejora.*

Si se hace una evaluación en un país como la realiza la OCDE, a través de PISA, y ésta se traduce en recomendaciones

concretas de qué tiene que hacer ese país para mejorar los resultados que obtuvo; el trabajo del INEE debe de ir en ese tenor, efectuar evaluaciones a nivel de país, las cuales se traduzcan en recomendaciones concretas sobre acciones necesarias de realizar; y si la Secretaría de Educación Pública desarrolla evaluaciones por escuela, censales, comparables, también a las escuelas debemos darles recomendaciones concretas para mejorar los resultados que están logrando; asimismo, la evaluación de los estudiantes que llega a los padres de familia, debe de ir acompañada no sólo de números, sino también de recomendaciones concretas para los padres de familia de cómo mejorar y apoyar a su hijo o al estudiante específico para lograr mejores resultados educativos.

Creo que éste es un reto adicional que tenemos y donde todos los organismos encargados de la evaluación debemos de participar. Tal y como en este momento lo vienen llevando a cabo las Instituciones de Educación Superior (IES) en la evaluación de las carreras profesionales del país, donde por cada carrera que evalúan, en el informe hacen recomendaciones concretas de los aspectos a mejorar.

Se debe lograr que toda nuestra evaluación tenga ese propósito y ese enfoque, y que todos participemos en la evaluación. Por ello, me da mucho gusto ver a los diversos actores de los organismos de la sociedad civil, que con su presencia demuestran el interés en la evaluación educativa.

Quiero felicitar al Instituto por la labor desempeñada, felicitar a todos ustedes por su interés hacia la evaluación y agradecerles anticipadamente su colaboración, para que podamos avanzar en este tema, que es tan importante para México y el mundo.

Nuevamente los felicito y deseo el mejor de los éxitos al Instituto en las próximas tareas que se ha marcado. Agradezco tam-

bién la participación de la Junta Directiva y del Consejo Técnico en las tareas del Instituto Nacional para la Evaluación de la Educación.

De esta manera, siendo las 10:30 horas del día 14 de julio, me complace de-

clarar inaugurados los trabajos de estas Jornadas de Evaluación Educativa, deseándoles a todos ustedes el mejor de los éxitos.

Enhorabuena y felicidades

**Intervención  
del doctor  
Reyes Tamez  
Guerra, al  
inaugurar las  
*Jornadas de  
Evaluación  
Educativa***





El Secretario de Educación Pública, doctor Reyes Tamez Guerra, al momento de inaugurar las Jornadas el 14 de julio de 2004.



Doctor Reyes Tamez Guerra y licenciado Felipe Martínez Rizo, director general del INEE.

PRIMERA MESA REDONDA

## **Presente y Futuro de la Evaluación Educativa en México**

**Ponente:**

Lic. Felipe Martínez Rizo

**Comentaristas:**

Lic. Ana Ma. Aceves Estrada

Dr. Carlos Muñoz Izquierdo

Dr. Salvador Malo Álvarez

Dr. Mario Rueda Beltrán

Dr. Carlos Ornelas Navarro

Dr. Pedro Ravela Casamayou

Dr. Francisco Mendoza Trejo

Dr. Arturo de la Orden Hoz

**Moderador:**

Mtro. Rafael Vidal Uribe





## Presente y Futuro de la Evaluación Educativa en México

Por FELIPE MARTÍNEZ RIZO\*

Esta presentación permitirá tener una visión de conjunto del quehacer del INEE, en este momento en que se acerca el Segundo Aniversario de la Publicación del Decreto que lo creó en agosto de 2002.

Para comenzar, me referiré a la evaluación educativa en México hasta el año 2003 señalando que, por una parte, se incluyen acciones valiosas, las cuales constituyen la base del trabajo futuro en todos los tipos educativos; los enumero rápidamente:

En educación básica destaca el trabajo de la Secretaría de Educación Pública, a través de la Dirección General de Evaluación, pero también por conducto de la Dirección General de Planeación, Programación y Presupuesto, donde se manejan estadísticas referentes a la educación; de la Subsecretaría de Educación Básica y Normal, donde hay actividades de evaluación, en especial relacionadas con los cursos de actualización de los maestros y en las entidades federativas donde hay una serie de actividades de evaluación; en parte, en apoyo a las que se hacen a nivel federal, pero también en no pocas entidades ya con acciones propias desarrolladas en la última década.

En educación media superior existe una serie de actividades; no hay actividades con tal alcance nacional como las existentes en educación básica, pero hay un gran número de actividades en los distintos subsistemas de educación media superior, como el de la UNAM, el del Colegio de

Bachilleres, la Dirección General de Bachillerato, la Dirección General de Educación Técnica Industrial Agropecuaria, el Conalep, etcétera; las comisiones de planeación de la educación media superior de las entidades federativas en algunas ocasiones tienen algunas actividades, aunque pocas en este terreno; y el Centro Nacional para la Evaluación de la Educación Superior (Ceneval) también tiene importantes actividades en relación con educación media superior.

En educación superior hay muchas actividades tanto a cargo de la Subsecretaría de Educación Superior e Investigación Científica, y la de Educación e Investigación Tecnológicas de la SEP, como son, entre otras, la evaluación de posgrados de SEP-Conacyt.

Las comisiones estatales de planeación de las entidades federativas tienen actividades de evaluación. El Centro Nacional de Evaluación para la Educación Superior por supuesto también. Los Comités de Pares, agencias acreditadas, Copaes, Fimpes, etcétera.

Aunque estas acciones son muchas y valiosas, también tienen limitaciones, las cuales se sistematizan en tres rubros: el primero lo constituye cierta desarticulación, que puede ser en algunas ocasiones excesiva; heterogeneidad de propósitos y no complementariedad, entre las diferentes acciones.

El segundo, una falta de difusión de resultados con implicaciones negativas de varias clases: los resultados no pueden uti-

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



lizarse para decisiones de mejora, la cultura de la evaluación no puede desarrollarse y, en consecuencia, no se detectan las fallas técnicas posibles.

El tercer rubro consiste en las deficiencias técnicas, tanto de planeación como de desarrollo e implementación de las evaluaciones, las cuales impiden llegar a conclusiones sólidas.

### **Viendo a futuro**

Acercas de las necesidades para el futuro, es necesario que las evaluaciones se utilicen para mejorar, ésta es la premisa fundamental: la evaluación tiene sentido en la medida que sirva para mejorar y esto en dos vertientes, por una parte lo referente a las autoridades educativas, llegando inclusive hasta los niveles de supervisión-dirección, y llegando a las escuelas y los maestros para cumplir sus prácticas pedagógicas y de gestión.

Por otro lado, para utilizar los resultados es necesaria la existencia de una difusión transparente y además, para que haya rendición de cuentas; lo cual es una demanda general de la sociedad, no sólo en lo referente al sistema educativo.

En el caso del sistema educativo hay demandas específicas, como las que se refieren a las necesidades de contar con resultados comparables en el tiempo y de tener resultados por escuela.

Para lograr ese objetivo, se necesita que las evaluaciones tengan una alta calidad técnica que es la condición, tanto para sustentar adecuadamente políticas de mejora, como para que la rendición de cuentas tenga un sustento sólido.

En conclusión, se necesita un sistema de evaluación. El primer problema es la pluralidad de instancias no articuladas, en un sistema tan grande como el de México deben de existir múltiples instancias de evaluación, pues no se puede preten-

der que una sola instancia se encargue de toda la evaluación, sino debe haber muchas con diferentes enfoques para los diversos tipos de educación: básica, media y superior, con distintos propósitos, pero no pueden ser atendidas bajo un mismo esquema.

En particular, las demandas mencionadas son legítimas y no pueden ser resueltas simultáneamente por un solo modelo o tipo de evaluación; la retroalimentación para maestros y padres de familia sobre la situación de sus hijos y alumnos, implica evaluaciones que se aplican a todas las escuelas y a todos los alumnos, a todos los grupos en todos los grados, en todas las materias, en varios momentos para que se puedan proporcionar resultados por alumno en plazos más cortos.

### **¿Son buenas las comparaciones?**

La comparación en el tiempo requiere metodologías apropiadas para tales comparaciones, que en sistemas educativos grandes implica muestrear. Teóricamente sería posible hacerlo universalmente, pero en un sistema de dimensiones menores a las del mexicano. Los sistemas de evaluación de los países con mayor experiencia, combinan enfoques masivos y muestrales o avanzan en esa dirección, el caso paradigmático es el de los Estados Unidos, en donde efectivamente se cuenta con sistemas que dan resultados por escuela, pero no a nivel federal, sino a nivel estatal. A nivel federal el sistema existente, el más sólido con el que cuentan, tiene más de treinta años de funcionar y sólo da resultados agregados; tiene prohibido por ley dar resultados por escuela, y de hecho hasta hace una decena de años no daba resultados ni siquiera por estado; actualmente ya cuenta con ellos, y son obligatorios para los estados.

Australia tiene un sistema similar; Chile, quien empezó de otra manera desde el

año 1988 a través de censos, recientemente ha planteado la necesidad de complementar con evaluaciones muestrales para asegurar la comparabilidad.

El sistema mexicano debe tener varias instancias articuladas, complementarias, con propósitos congruentes y mecanismos consistentes con niveles de calidad similares, lo cual se manifiesta en la expresión de *sistema nacional de evaluación*, en lugar de muchas instancias no articuladas.

### **Avances cualitativos**

El arranque de una nueva etapa, es el trabajo del Instituto en sus dos primeros años de existencia; en los cuales ha comenzado a desarrollar un conjunto de acciones que constituyen avances cualitativos en la historia de la evaluación educativa del país. Los enumero brevemente:

Comienzo refiriéndome al Sistema de Indicadores, en el cual se ha desarrollado conceptualmente un sistema muy completo y dinámico, en el sentido que no se da por terminado, sino tiene que ir creciendo y afinándose; se ha producido el Primer Anuario de Indicadores el cual se distribuyó vía Internet desde principios de este año, ya está circulando en papel, y pretendo tomarlo como ejemplo de lo que deseo decir con este avance cualitativo.

El Anuario de Indicadores del INEE no contiene ningún dato nuevo que no existiera en alguna otra fuente, lo novedoso es que se integran y presentan de una manera conjunta, y se explicita la metodología utilizada para integrar los resultados, también presenta la definición de cada indicador, fórmula de cálculo, criterios utilizados en la interpretación y avances, los cuales, por supuesto, son discutibles.

Pero al hacerlo de esta manera, al explicitar la metodología seguida, se pone al alcance de cualquier especialista y/o

persona interesada la posibilidad de discutir, cuestionar y mejorar lo que se está presentando; algo que no sucede cuando se exponen los datos sin la explicación de metodologías.

En lo referente a pruebas nacionales, se difundieron desde finales del año pasado los resultados de diez estudios de pruebas hechas en el pasado, todos están en Internet, en forma preliminar, desde finales del año pasado y en forma definitiva a principios de este año; ya se inició el desarrollo de una nueva generación de pruebas que pretendemos sea cualitativamente mejor.

En lo referente a pruebas internacionales, el año pasado se hizo la aplicación de las pruebas PISA 2003 a una muestra muy grande de alumnos de 15 años en todo el país, para poder realizar análisis por entidad.

Se calificaron los resultados y se están preparando los análisis, México cumplió todas las exigencias de calidad de la OCDE para que sus resultados puedan incluirse en el informe internacional, el cual será presentado públicamente por la OCDE el 7 de diciembre, pretendemos ese mismo día presentar el análisis nacional.

Por otro lado, hemos comenzado una participación activa en la planeación y desarrollo de nuevas pruebas, en concreto las Pruebas 2006. Será la primera vez que México no sólo participe aplicando una prueba internacional, sino participando en su desarrollo, con construcción de preguntas, con piloteo de preguntas, las cuales serán incluidas en la prueba internacional.

Un siguiente punto es la evaluación de escuelas, la cual desde el año pasado desarrolló un conjunto de instrumentos para evaluar distintos aspectos de la vida escolar; también se difundieron desde principios del año en Internet.

Por otro lado, este año se está desarrollando un programa muy importante de



capacitación para personal de las Secretarías de Educación o equivalentes en las 32 entidades federativas de la República, para así contar este mismo año con informes estatales de los resultados de las evaluaciones del año pasado, los cuales se entregaron desde el mes de diciembre a los estados, tanto las realizadas por el Instituto, como las realizadas por la Secretaría de Educación Pública.

El área de Relaciones Nacionales ha conformado la interacción con las entidades federativas a través del Consejo Consultivo, está realizando un proceso de sistematización y mejora, tanto de las aplicaciones y pruebas, como de las lecturas de resultados. Ya se está constituyendo una biblioteca especializada cuyo catálogo computarizado estará también muy pronto en Internet.

En lo referente a Comunicación y Difusión, publicamos el Primer Informe Anual desde el mes de noviembre del año pasado, con la síntesis de los resultados de las evaluaciones; el 17 de noviembre se presentó al Presidente de la República y en rueda de prensa a los medios de comunicación; también se colocó en la página de Internet. Se han puesto en línea todos los productos editoriales realizados hasta hoy, y en especial quiero referirme a las bases de datos con resultados de las evaluaciones.

La base de datos de 2003 está completa en Internet desde hace un par de meses; todos los microdatos, los datos, las respuestas de cada niño a cada pregunta, obviamente sin identificación del mismo, están en Internet. Los resultados de los cuestionarios de contexto aplicados también los encontrarán, para que cualquier persona interesada o cualquier especialista pueda hacer sus propios análisis y deduzca sus propias conclusiones contrastándolas con las del Instituto.

La mejora de los procedimientos de aplicación y lectura referidos, nos permiti-

tirá que la base de resultados de las Pruebas 2004 aplicadas en abril y mayo, estén disponibles para el público en Internet, esperamos, a partir del mes de octubre próximo.

Se ha realizado una serie de actividades para periodistas y, por otro lado, un programa de difusión a través de libros, cuadernos técnicos, folletos y medios electrónicos.

En lo referente a la difusión de resultados por escuela, el Instituto está convencido de la necesidad de hacerla, somos conscientes de que se deben reunir determinados requisitos, concretamente: sean evaluadas todas las escuelas del sistema o subsistema de que se trate; se tenga información sobre las variables del entorno de las escuelas para poder contextualizar los resultados; y se disponga de varias mediciones y no de una sola.

Como consecuencia, las pruebas de estándares nacionales o denominadas de estándares nacionales que se aplican sólo a muestras, no pueden dar lugar a este tipo de difusión, pero sí pueden hacerlo otros elementos. Por ello, desde hace más de un año empezamos a apoyar a la Subsecretaría de Servicios Educativos del Distrito Federal para desarrollar un sistema de difusión de resultados en línea escuela por escuela, el cual será presentado hoy mismo.

Por otro lado, para dar a conocer la demanda de comparación de resultados en el tiempo, como parte de las *Jornadas*, se presentará una mesa con el análisis de las limitaciones de los resultados existentes, las cuales impiden este tipo de comparaciones; pero no nos quedamos en ello, ya en el mes de noviembre pasado presentamos unas modestas comparaciones temporales con los elementos disponibles, que eran resultados de TIMSS, los cuales permitieron hacer las comparaciones 1995-2000.

En diciembre próximo, presentaremos los resultados de PISA 2003, los cuales se podrán comparar con PISA 2000, ello nos permitirá tener resultados por estado, dado el tamaño de la muestra que utilizamos.

El Instituto efectuará el año próximo una aplicación especial para realizar una comparación 2000-2005 y elaborará una serie de aplicaciones de instrumentos nuevos, lo cual permitirá contar a partir del año próximo con una serie consistente para efectuar este tipo de comparaciones en el futuro.

La evaluación que el Instituto busca promover es la evaluación educativa en gran escala, queremos rescatar los elementos valiosos del pasado ya que son la base sobre la cual se puede construir una mejor evaluación en el futuro.

### **Nueva visión de la evaluación**

El trabajo del Instituto comprende ya avances significativos respecto a lo hecho anteriormente, además estamos promoviendo una nueva visión de la evaluación educativa caracterizada por tres rasgos novedosos que corrigen las deficiencias anteriores.

El primero de ellos es el *sistema nacional de evaluación educativa*, éste consiste en una visión integral, articuladora de los distintos elementos de la evaluación de resultados de aprendizaje, de impacto social, pero también la evaluación de recursos y procesos de elementos de contexto de los planteles escolares. Por otro lado, que incluya acciones del Instituto, pero también de la Secretaría de Educación, de las autoridades estatales y diversas agencias especiales, como el Ceneval, Comités de Pares, etcétera, con mecanismos que aseguren la articulación de las instancias, para que sean congruentes entre sí y realmente constituyan un auténtico sistema nacional de evaluación.

El segundo rasgo de esta visión de la evaluación deseable promovida por nosotros, es una evaluación que se difunda como un medio valioso para la mejora y bajo el concepto de Política Nacional de Evaluación, para que ésta llegue por un lado a legisladores, autoridades educativas, directores, supervisores y maestros, lo cual ayudaría a sustentar acciones de mejora; pero si también llega a padres de familia, empresarios, medios de comunicación y sociedad en general, desarrollaremos una cultura de evaluación, la cual apoyará la rendición de cuentas de acuerdo con una política nacional coherente, sancionada ya por la Junta Directiva del Instituto.

El tercer rasgo, es el referente a calidad técnica de la evaluación, la premisa es lograrla para que los resultados sean confiables y puedan sustentar sólidamente la mejora educativa.

Esta pretensión de tener mejor calidad no se puede dejar en una simple promesa, no puede simplemente ofrecerse y esperar que nos crean por buena voluntad.

Para ello, necesitamos contar con elementos muy precisos, los cuales sistematizo en tres puntos:

- Definir de manera explícita los criterios de calidad que deberán cumplir las evaluaciones.
- Desarrollar con toda precisión los procedimientos a seguir para asegurar que esos criterios se cumplan.
- Precisar un conjunto de instancias internas y externas, integradas por especialistas, maestros, usuarios, etcétera, quienes supervisen la operación de los procedimientos y el cumplimiento de los criterios.

La instancia fundamental para el cuidado de la calidad técnica del INEE es el Consejo Técnico, pero deberán crearse, de hecho ya han empezado a constituirse y están empezando a funcionar, diferentes



grupos especializados para aspectos muy precisos que se deben atender.

Con estos elementos se podrá asegurar que la evaluación será de mejor calidad.

A partir de hoy vamos a subir a Internet una primera definición del Modelo de Calidad del Instituto, el cual ayer y antes de ayer fue revisado en una primera versión por el Consejo Técnico, presentado a la Junta Directiva, y el cual debe ser, obviamente, enriquecido, complementado y estar en desarrollo permanente. Dicho modelo nos muestra la complejidad de la tarea, la existencia de decenas de pasos sumamente precisos, cada uno de los cuales se tiene que desarrollar con apego a criterios de calidad puntuales, con procedimientos perfectamente definidos, vigilados por instancias internas y externas que aseguren se cumplan los procedimientos y satisfagan los criterios de calidad.

Con esto llego al final. Lo que el INEE pretende aportar no es completamente novedoso, se basa en las mejores prácticas de los sistemas de evaluación de mayor experiencia en el mundo. Los estándares de calidad existen, los estamos tomando, adaptando e integrando de los mejores sistemas, del sistema de pruebas PISA de la OCDE, de la Asociación Internacional para la Realización del Logro Educativo que hace TIMSS, de los estándares *Service*, de los estándares de las tres agrupaciones profesionales americanas más importantes en este rubro, y del documento mexicano de estándares que publicó el Ceneval en el año 2000.

Desarrollar un sistema de evaluación en gran escala, de gran calidad técnica, implica años de trabajo consistente, sobre todo cuando no se parte de una tradición al respecto. Un ejemplo de ello es el Sistema Nacional de Evaluación de los Estados Unidos, que es el más consolidado en el mundo y el cual comenzó a prepararse en 1963, y a funcionar sólo hasta 1969, que-

riendo decir que se pasaron seis años de planeación y eso que es un país que tenía una gran tradición. Asimismo, incrementó su calidad a partir de 1983, cuando se le encomendó la educación al organismo más grande del mundo en estas cuestiones. Detectó fallas de comparabilidad en el tiempo todavía en 1986, a partir de entonces sofisticó los estándares de control para poder asegurar la comparabilidad de resultados. Apenas en 2002 se definió su obligatoriedad, para las entidades federativas con margen hasta el año próximo. Ese es el tiempo que llevó desarrollar el sistema.

### Los tres horizontes

A continuación veremos que el proyecto de creación del Instituto Nacional para la Evaluación de la Educación en el que se fundamentó el Decreto de Creación, se instala en tres horizontes: 2002, 2006 y 2012:

- Establecía la fecha de 2002 como horizonte para la creación del Instituto,
- 2006 para su consolidación inicial,
- 2012 para su consolidación plena, porque había conciencia de que cuatro años era muy poco para lograr la solidez plena de un sistema; así es que, además de la complejidad técnica está condicionado por los ritmos del sistema educativo.

No se pueden producir indicadores cada mes, hay un ciclo anual. No se pueden evaluar resultados de ciclos escolares cuando uno quiere, tienen su ritmo anual. Por eso se requieren estos tiempos.

Termino diciendo que cumplida la primera meta, la meta de 2002 con el Decreto Presidencial del 8 de agosto de 2002 que lo estableció, el Instituto Nacional para la Evaluación de la Educación llega a su segundo aniversario, avanzando firmemente hacia su consolidación.

El Plan Maestro de Desarrollo del Instituto, el cual fue aprobado el día de ayer 13 de julio, y define las tareas a realizar de 2004 a 2012, es el instrumento guía del trabajo del Instituto para lograr se alcancen metas de mediano y largo plazos.

El Modelo de Calidad del Instituto, en su primera versión, muestra con todo de-

talle la manera precisa de cómo se realizará el trabajo, para que de esta forma el Instituto cumpla su misión de realizar evaluación confiable, y difundirla de manera oportuna y transparente, apoyando los esfuerzos nacionales de ofrecer una educación de buena calidad a todos los niños de México.

---

## **PARTICIPACIÓN DE LOS COMENTARISTAS**

ANA MARÍA ACEVES ESTRADA\*

---

**E**s muy estimulante ver llenos estos salones para discutir acerca de la evaluación, esto quiere decir que todos los esfuerzos que hemos venido haciendo están permeando en la sociedad, y la evaluación se ha vuelto materia de discusión, análisis y aportaciones.

Me congratulo y quiero transmitirles a todos mi confianza, porque no obstante que nos quedan todavía muchas cosas por hacer, y que todavía faltan algunos consensos, los pasos dados son correctos y además son congruentes con las metas y las estrategias que el mismo Programa Nacional de Educación (Pronae) se ha propuesto en esta visión al año 2025.

Si ustedes recuerdan, el Pronae manejaba tres estrategias muy importantes, la primera, era establecer mecanismos rigurosos y confiables, los cuales serían independientes de las autoridades educativas: el INEE es un hecho ya consumado.

La segunda era establecer criterios para la difusión de los resultados y su utilización para el mejoramiento de la calidad de los mismos, algo que también ya está realizando el INEE; y además nos está orientando a las autoridades educativas y a la propia Dirección General de Evaluación

para hacérselos llegar a los periodistas y a la sociedad para que tengan una idea más aproximada de cómo interpretar resultados; y por último, la construcción de un verdadero Sistema de Evaluación, así dice el Programa Nacional, el cual articule las funciones en un contexto de complementariedad para coordinar y consensar acciones, estrategias y programas, creo que éste es el punto más relevante de todos.

### **Delimitación de funciones**

Uno de los aspectos en el cual hemos venido trabajando muy arduo en la construcción del sistema, es en la delimitación de las funciones y las características de las acciones que cada uno de nosotros tenemos.

En este sentido, el INEE se crea como un organismo de vocación e independencia autónoma, pues va a informar a los ciudadanos sobre el sistema en general.

Pero quiero hacer una reflexión con ustedes, sobre la importancia que tiene la otra parte de la evaluación, si nosotros queremos calidad educativa, entonces el concepto tiene dos acepciones: relativa y dinámica; relativa porque el juicio depende del punto de referencia de dónde se

---

\*Directora General de Evaluación de la SEP y Secretaria Técnica del Consejo Consultivo del INEE.



tome, eso es algo muy importante cuando nosotros damos a conocer un resultado de calidad, qué punto de referencia, qué nivel de desarrollo tiene el estado, qué nivel de calidad en el desarrollo de vida tiene la población; y dinámica, nunca se alcanza la calidad absoluta, siempre será probable proponer metas más elevadas y cuando éstas se alcanzan, siempre habrá otras más ambiciosas.

Así, el punto de referencia más pertinente, será la línea del tiempo y la mejora respecto a uno mismo, conceptualizar así la calidad, entonces diríamos que no es un estado, sino una tendencia, una autoexigencia permanente y razonable de superación, la cual no se puede adquirir del exterior, solamente podría surgir del interior mismo del sistema a mejorar, de esta manera la calidad del sistema nacional se alcanzaría o no se alcanzaría en el aula y en la escuela; para que esto sea posible, el proceso de superación constante tendría que contar con un buen sistema de evaluación de carácter censal y permanente, gracias al cual las autoridades y los actores del sistema nacional, pudieran vigilar el estado en que se encuentra la educación y monitorearla.

Es ahí donde la Dirección de Evaluación de la SEP y los sistemas estatales juegan un papel muy importante, en este sentido quiero compartir con ustedes los tres programas que llevamos a cabo: primero, apoyamos al maestro para que éste pueda familiarizarse con instrumentos de evaluación, acercándoselos al inicio del ciclo para elaborar diagnósticos sobre cómo recibe al alumno y poder, de alguna manera, prever reprobaciones y poder trabajar junto con padres de familia sobre la recuperación del atraso escolar de ese alumno.

Otro de los programas es la aplicación censal del instrumento de aprovechamiento escolar, el cual no solamente mide matemáticas o español, sino que mide todas las materias que se imparten tanto en primaria como en secundaria, a efecto de poder contar con la línea del tiempo, y saber cómo va mejorando una escuela.

Tercero, para poder evaluar realmente el avance de una escuela, no sólo tiene que darse la evaluación cuantitativa de la aplicación de un instrumento, también tendrán que darse algunas otras evaluaciones cualitativas, y en este sentido las evaluaciones que está haciendo el Programa de Escuelas de Calidad, –las cuales han sido de las más extensas a nivel internacional– podrán proporcionar instrumentos muy importantes.

Quiero compartir una reflexión más, la escuela en muchas ocasiones recibe alumnos de escuelas exitosas, escuelas que tienen una sobre demanda de alumnos, ésta se da el lujo de recibir alumnos de los mejores; entonces cómo puede efectivamente saberse si una escuela realmente está avanzando. Se necesita medir la ganancia educativa, lo que en un ciclo escolar aporta un maestro, cómo recibe al alumno y cómo lo deja al final, y en esta ganancia educativa el examen de diagnóstico censal va inmerso, tendremos entonces que revisar a futuro una evaluación a mitad y al final del ciclo escolar, ésta debe ser no controlada, para que el maestro, el director de la escuela y las autoridades educativas puedan saber qué es lo que está sucediendo en el aula; hacia allá se orientará y complementará la Dirección General de Evaluación con el INEE.

Con todas estas acciones, podemos congratularnos y tener confianza de que las cosas, a pesar de que hay mucho por hacer, van por el camino correcto.



La exposición del licenciado Martínez Rizo nos ofreció una caracterización de cómo se encontraba la evaluación de la educación en México, cuando fue creado el Instituto Nacional para la Evaluación de la Educación.

Expuso los principales retos que se desprenden de esa situación, mencionó las acciones que el Instituto está llevando a cabo para enfrentarlos y, por último, describió las cualidades más sobresalientes del modelo conceptual que orienta la actividad de estudios del Instituto.

Desde mi punto de vista, la parte medular de la exposición que escuchamos se ubica en el último capítulo de la misma, es ahí donde encontramos una síntesis de las razones –que se apoyan en la confianza–, de que el INEE está creando un escenario que le permita contribuir realmente al mejoramiento de la educación que se imparte en México.

Por tal razón, me voy a referir a dos de los temas tratados en este capítulo, lo haré desde una perspectiva de futuro, señalando algunas tareas planteadas en la exposición que acabamos de escuchar; considero que tienen una alta probabilidad de ser revisadas por el Instituto dentro de los próximos años.

### Tareas por cumplir

En primer término, mencionaré algunas de las actividades que se desprenden de la visión integral del sistema escolar en que se apoya el Plan Maestro del Instituto.

En segundo lugar, me referiré a otras acciones que están encaminadas a promover el uso efectivo de los resultados

que se obtengan en las actividades del propio INEE.

Una vez que el Instituto haya avanzado suficientemente hacia el cumplimiento de las metas fijadas en su Plan Maestro, es probable que pueda abordar otras actividades señaladas en el mismo documento; entre ellas se encuentra el compromiso de generar información acerca de aquellos resultados, en materia de educación, que se definen como el grado en el cual el sistema educativo satisface las necesidades de la sociedad, así como la manera en que la educación contribuye al desarrollo económico, social y político del país.

La publicación titulada: *Panorama Educativo de México*, que acaba de producir del Instituto, parte efectivamente de una visión global del sistema educativo, por lo que además de ofrecer información vinculada directamente con ese sistema, proporciona un perfil democrático del México en el cual estuvo sustentado dicho sistema.

Por tal razón, aparece una serie de libros con diversos datos acerca de la desigualdad social, marginación, participación de la población activa, etcétera, y aspectos de la calidad de vida de la población. Los fenómenos a que se refieren esos datos, condicionan o favorecen, según el caso, la eficiencia y la eficacia de la educación. Es importante recordar que los productos mismos del sistema educativo también influyen para la retroalimentación del contexto socioeconómico del mismo.

Por tanto, usando microdatos de la información a la que me refiero junto con otra que se define, y la cual tendrá alcances más amplios para transformarse en

\*Presidente del Consejo Técnico del INEE e investigador de la Universidad Iberoamericana.



macrodatos, podrá servir para construir indicadores nuevos que reflejen los diferentes impactos que tiene la educación en la sociedad y en el modelo productivo. Para hacerlo, bastaría estimar si en la sociedad existe el comportamiento de algún fenómeno social, ya sea económico, político o cultural a los que me acabo de referir. Esa es una de las tareas que el Instituto hará próximamente para responder de manera adecuada a los requerimientos señalados.

Es, por otra parte, positivo suponer que si la evaluación educativa satisface los requerimientos especificados en la ponencia de Martínez Rizo, podrá surtir efecto en las escuelas educativas, en la adopción de innovaciones, en las prácticas de los diferentes actores que intervienen en la educación; también influirá en el desarrollo de nuevos paradigmas interactivos, en los programas de formación de profesores e investigadores, en la construcción de una opinión pública más madura, así como en la maduración de valores en los diferentes actores que participan en la educación.

Ahora bien, como sabemos, estos impactos sólo se pueden generar si concomitantemente, se desarrolla un conjunto de dinámicas de diversa índole, porque la transferencia de conocimientos depende de la generación de procesos, los cuales son más complejos de lo que a simple vista es aparente.

### **Tiempo y espacio para la información**

Por lo anterior, el Instituto está atento a cada una de las instituciones intere-

sadas en los estudios de evaluación, es decir, que las diversas autoridades educativas del país y el público en general, dispongan del tiempo y los espacios para poder utilizar la información producida.

Asimismo, promoverá la creación de mecanismos que permitan someter los resultados de evaluación a una discusión abierta. Alentará la generación de situaciones para que participen actores como las universidades, maestros, secretarios, órganos no gubernamentales, etcétera, todo lo cual requiere la creación e integración de canales formales de comunicación para que la información circule adecuadamente. En ese sentido, el INEE deberá avanzar hacia la publicación y difusión de documentos en donde se sinteticen los logros. Asimismo, apoyará a las redes de investigadores y educadores que impulsen intercambios diversos.

Así pues, me parece que el modelo conceptual que está guiando las actividades del Instituto, lo coloca en un escenario que abre cauces a la creación de procesos estratégicos, como ya había mencionado. Esto, por cierto, fue prácticamente imposible en el pasado.

Toca ahora al Instituto aprovechar los lineamientos y las oportunidades que él mismo ha creado durante estos dos años de vida. Al hacerlo, podemos confiar en que este organismo avanzará con firmeza hacia el cumplimiento de las metas fijadas.



Todo aquel que haya estado en contacto con la educación en México se habrá percatado que, como nos acaba de ilustrar esta mañana el director del INEE, en los últimos tres o cuatro lustros se han introducido múltiples procesos e instrumentos de evaluación y creado diversas instancias dedicadas a ello. Dentro de éstas destaca el Instituto Nacional para la Evaluación de la Educación, cuyo segundo aniversario tengo la oportunidad de celebrar anticipadamente con ustedes.

La rapidez con que han crecido las actividades de evaluación en México es muestra, tanto de la necesidad que teníamos de ellas, como de su anterior ausencia. La necesidad explica la multiplicación cualitativa de las acciones evaluatorias y su cobertura creciente de poblaciones y sectores; explica también la proliferación de grupos y la diversidad de aspectos dedicados a los sujetos de la evaluación.

### **Crecimiento de las actividades evaluatorias**

La urgencia por llenar el vacío de evaluación se nota en la aparición de una gran variedad de datos y procedimientos, muchos de ellos aislados e inconexos con los que hoy contamos. Asimismo, en la aguda carencia que tenemos de personal preparado y con experiencia en evaluación educativa, psicometría, técnicas estadísticas y procesamiento de la información. Y, finalmente, en la limitada apreciación que la sociedad tiene acerca de lo que la evaluación aporta al entendimiento y desarrollo de los procesos, instituciones y sistemas educativos.

En consecuencia, a pesar de señalar el vacío, el rápido crecimiento de la evaluación educativa ha traído problemas nuevos a la educación, a la sociedad y a los evaluadores mismos, éstos últimos son frecuentemente acusados de ser culpables de los males que ayudan a detectar; criticados por no evaluar lo que ni siquiera les corresponde evaluar; o cuestionados por no emprender acciones para remediar los problemas que sólo contribuyen a precisar.

Pero no es el momento para lamentar las carencias, menos aún para intentar corregir supuestos agravios; mi intención al mencionar lo anterior es hacerles ver la dinámica que vive la evaluación educativa en México y señalar cuán importante es atender las demandas que la sociedad hace a la evaluación.

Ambos son puntos importantes para anticipar el futuro de la evaluación educativa en México, uno se refiere a lo técnico, el otro a lo social. Sin duda debe de evitarse toda evaluación mal hecha, pero la evaluación que no se usa no es evaluación.

### **¿Hoy qué duda cabe?**

Se usan instrumentos y procesos de evaluación externos a los planteles educativos que se aplican sistemáticamente; también es cierto que se avanza significativamente a fin de diseñarlos y aplicarlos de mejor manera y analizar mejor sus resultados.

Sin embargo, la sociedad mexicana está todavía lejos de cerrar el ciclo de la evaluación, la mayor parte de las acciones de evaluación organizadas se centran hoy

\* Director General del Ceneval.



en día en contenidos educativos, planes de estudio, infraestructura y el profesorado; pocas, en cambio, se ocupan de la congruencia o relevancia de dichos aspectos con los objetivos de la educación, con las características de la población evaluada o con los resultados alcanzados.

Pocos también son los casos en que las evaluaciones tienen consecuencias para los aspectos evaluados, y derivan en cambios de organización, de enfoques, de metodologías, menos aún los que se conocen y sirven de experiencia y ejemplo para otros casos y necesidades. Estoy seguro que en los próximos años la evaluación educativa en México avanzará en esta dirección.

En los entornos de la evaluación también es común que se asuma como algo establecido lo que debe ser motivo de evaluación y se supongan conocidos los instrumentos apropiados para ello, es decir, existe una percepción extendida de que los problemas centrales para avanzar en la evaluación de la educación son tan sólo de voluntad y de recursos.

### **Preguntas para reflexionar**

Sin embargo, los avances en la ciencia de la evaluación, están llevando a cuestionar cada vez más frecuentemente si lo que se está evaluando es lo que debiera ser; si el énfasis en contenidos es más importante para el aprendizaje que aquel en competencias; si se debiera trabajar más en la evaluación del uso que se hace del conocimiento, que la de los conocimientos que se poseen; si convendría prestar más atención a la determinación de los factores que propician o limitan la apropiación del conocimiento, que los que atienden a su memorización o acumulación.

Éstas son algunas de las preguntas que nos hacemos y sobre las cuales estamos

trabajando en el Centro Nacional de Evaluación para la Educación Superior, preguntas que, por otra parte, comparten otros grupos y personas dedicadas a la evaluación educativa y cuya atención trasciende a lo que ocurre en las aulas y los ámbitos escolares.

La urgencia de llevar a cabo actividades en evaluación educativa y la falta de una cultura de evaluación, que caracterizaron la primera etapa de la evaluación educativa en México, llevaron a que, independientemente de la intención para la que fueron diseñadas, todas ellas fuesen de alguna forma diagnósticas, exploratorias, tentativas.

Pero el avance en la comprensión de los procesos evaluatorios, en el entendimiento mismo del sistema educativo y de los factores y problemas asociados a la enseñanza y al aprendizaje, y el más cabal aprecio de la contribución que los primeros pueden y debe hacer en los segundos, están conduciendo a la sociedad mexicana hacia el establecimiento de estándares o referentes para la educación y los aprendizajes.

Son cada vez más numerosos y frecuentes los reclamos que se hacen por la ausencia de indicadores y criterios para determinar cuándo un plantel proporciona un nivel de preparación aceptable o adecuado, si un egresado adquirió o no la instrucción que requiere, y si ésta es acorde al ciclo de estudios.

Finalmente, considero que en los próximos años la evaluación educativa en México avanzará en la formación de cuadros humanos calificados, cuadros que contribuirán a la construcción de los caminos antes descritos y al trazo de otros nuevos, los cuales ayudarán también a que la evaluación educativa se acerque a las demandas sociales sin sacrificio de sus exigencias técnicas.



**M**is comentarios se centrarán en los aspectos, que me parece, el INEE está poniendo en marcha para atender el tema de la evaluación, en particular destacaré los más vinculados con la investigación educativa, ya que ésta representa la preocupación central del Consejo Mexicano de Investigación Educativa.

Destaca, en un primer momento, la reflexión sobre la necesidad de una instancia especializada en evaluación, en particular bajo la figura de Instituto, que sugiere un alto compromiso con el desarrollo de la investigación en evaluación; es decir, con la producción de conocimiento sistemático, además de las otras funciones explicitadas en su Decreto de Creación y sus documentos rectores.

De todos es conocida la presencia creciente de la evaluación en el sector educativo, la cual marca la necesidad y reconocimiento de la profesionalización y especialización en la valoración de las instituciones, programas e individuos.

### **Investigación: una meta**

La existencia de un Instituto dedicado a la evaluación, con un enfoque basado en la investigación, sin duda contribuirá a fortalecer las condiciones para que en nuestro país se avance hacia la producción de conocimiento sistemático en este tema tan relevante.

De acuerdo con el análisis de los documentos que orientan las acciones del INEE, se plantea la realización de investigaciones referidas al desarrollo tecnológico; sobre evaluación, y a la que proporciona explica-

ciones. Todo ello en la perspectiva de contribuir a la mejora del sistema educativo.

Esta perspectiva de investigación seguida por el INEE se constata desde la constitución misma del Consejo Técnico, al que pertenecen investigadores reconocidos en el ámbito de la evaluación de nuestro país y de otros países, lo cual constituye una plataforma de pluralidad académica, diversidad de posturas teóricas y la posibilidad de recuperar experiencias de otros contextos culturales.

Se obtienen también evidencias de este enfoque de investigación prevaleciente en el INEE, al revisar su primer informe y cada uno de los materiales que hasta la fecha se han elaborado, en donde el común denominador es el rescate del conocimiento disponible sobre cada uno de los temas, la explicación de marcos conceptuales, la rigurosidad intelectual en su desarrollo y el sometimiento de los productos a opiniones externas; es decir, características del trabajo intelectual reconocido entre pares.

La forma de interacción del personal del INEE con el Consejo Técnico también apoya la idea de esta postura, al someter al escrutinio minucioso de los consejeros los documentos clave de cada una de las acciones de sus planes de trabajo y al hacerse apoyar por expertos en asuntos especializados cuando la naturaleza de la tarea así lo requiere.

En momentos particulares se ha propiciado el diálogo directo con miembros destacados de la comunidad de investigadores para escuchar sus puntos de vista sobre lo realizado. Tal fue el caso de la presentación del Primer Informe, y de proyec-

---

Presidente del Consejo Mexicano de Investigación Educativa (Comie) y miembro del Consejo Técnico del INEE.



tos de investigación especiales conducidos por académicos externos al Instituto.

En la misma dirección se destacan las convocatorias públicas dirigidas a los investigadores, las cuales ofrecen la oportunidad para elaborar proyectos de investigación sobre los distintos aspectos de la evaluación, los cuales favorecen las condiciones para aumentar el conocimiento disponible sobre el tema.

Una contribución no menos importante está referida a la preocupación en la formación de recursos humanos, como se mencionaba al inicio, cada vez más, debido a la profesionalización y especialización de la evaluación se hace absolutamente necesaria.

Hasta donde tengo conocimiento, el INEE está procurando una formación especializada para el personal permanente del propio Instituto, y para profesionistas externos que quieran dedicarse de forma prioritaria a esta actividad. Por supuesto, se incluyen todos los trabajos de investigación en el marco de posgrados en educación, sobre todo, pero también en otras áreas afines.

### **Difusión del conocimiento**

Sin duda uno de los retos para el INEE será la traducción y difusión del conoci-

miento, derivado de sus múltiples actividades, para los distintos destinatarios o públicos involucrados, de tal manera que se garantice la función primera de la evaluación que es el perfeccionamiento o mejora de la actividad evaluada.

Si pensamos en la diversidad de audiencias involucradas: legisladores, investigadores, inspectores, directivos, planificadores, profesores, alumnos y padres de familia, no tendremos ninguna duda sobre la dificultad de la tarea, y muy probablemente aceptaremos la necesidad de profesionalizar tal actividad.

También habría que reconocer que en cuanto a la perspectiva de contribuir a la cultura en evaluación, se debe partir de la aceptación de un contexto social que en principio la rechaza, en gran medida por el conocimiento de experiencias de malos usos, se evalúa para cerrar un establecimiento, no contratar a un maestro, etcétera. Pero insisto, son malos usos de la evaluación o concepciones que la han identificado con medición, calificaciones injustas o procesos de exclusión.

Por lo tanto, todo esfuerzo que se haga para contribuir a una mejor evaluación será bienvenido, tal es el caso del INEE. Corresponde también a la comunidad de investigadores coadyuvar para que esta experiencia se consolide.



**D**ebo empezar por reconocer que sí hay avances en algunas de las críticas, que no nada más yo, sino otros colegas también, han hecho acerca de la misión del INEE.

En primer lugar, había declaraciones tajantes del director general, de Felipe, de que no se podía hacer evaluación escuela por escuela y que los resultados no eran comparables en el tiempo con los de la Dirección General de Evaluación; ahora nos dice que sí nos encaminamos para allá, va a tomar tiempo pero sí se puede.

Quiero concentrar mis comentarios en la educación básica y en tres de las aportaciones que nos hizo el doctor Reyes Tamez: primero, la misión principal de la evaluación es detectar las deficiencias de escuelas, profesores y alumnos, claro, con el fin de superarlas y de mejorar; segundo, debe ser con diferentes enfoques y, tercero, ser plural, de manera que todos participemos en la evaluación.

Aquí noto cierta vigencia de lo que en otro trabajo denominé el centralismo burocrático a pesar de la descentralización, porque todo sigue emanando del centro hacia los estados, los estados son actores considerados objetos, en donde los miembros de las direcciones de evaluación deben capacitarse porque tienen deficiencias técnicas, como no conocer bien los instrumentos estadísticos; puede ser que haya razones para argumentar eso, pero me parece mal la concepción de que nosotros desde el centro vamos a decir qué y cómo hacerlo, en lugar de ser una invitación a construir juntos este proyecto.

Existe una batería gigantesca de exámenes que aplica año con año la Dirección General de Evaluación de la Secretaría de Educación Pública para efectos de Carrera Magisterial, alrededor de seis millones por año a alumnos de educación básica.

Si ustedes conocen el código o la clave del centro de trabajo de su escuela, pueden entrar a la página de la Dirección General de Evaluación y van a encontrar ahí, escuela por escuela, salón por salón, en todas las materias del currículum, cómo están los resultados de los alumnos de tal maestro o de tal maestra; no cubre educación indígena, no cubre sector privado, por ejemplo, pero el universo considerado es muy amplio.

Y además se hace en una escala que todos podemos entender, cuando se habla de retroalimentar a padres y maestros, es eso, darles la información y que ellos busquen mejorarla, el problema es que se les retroalimenta con escalas desconocidas, dentro de la pluralidad tenemos escalas de 200 a 900, de 300 a 1,100, se dan cantidades infinitas que para padres de familia, es más, no sólo para padres de familia, también estudiantes de maestría y doctorado, son difíciles de interpretar por la dificultad de la escala, cuando la escala del 0 al 100 ya la tenemos familiarizada, nos formamos en las escuelas con el sistema métrico decimal. De tal manera, que aquí hay una sugerencia para mejorar.

### **Interviene la política**

En política los símbolos cuentan. Ustedes recordarán que el 8 de agosto de

\* Investigador de la UAM.



2002, cuando apareció en el Diario Oficial de la Federación el Decreto de Creación del INEE, después de debates previos en la Cámara, mesas redondas, fue el mismo día en que se firmó el así llamado, *Compromiso Social por la Calidad de la Educación*. Se reunió todo el sistema educativo, gobernadores, magisterio, varias secciones sindicales y empresarios en el Museo de Antropología, y también la maestra Elba Esther Gordillo; de tal manera que simbolizaba como si fuera producto de un gran consenso nacional y no es así, creo yo, porque uno de los problemas principales es la cuestión de la dependencia del INEE con respecto a la Secretaría de Educación Pública –el Secretario es el presidente de la Junta Directiva– Rafael Ochoa Guzmán, Secretario General del Sindicato Nacional de Trabajadores de la Educación, y el Presidente de la Fundación SNTE para la Cultura del Maestro, están también dentro de la junta y además otros seis funcionarios de la SEP, de la Secretaría de Hacienda y del Instituto Mexicano del Petróleo. Eso le resta credibilidad en términos de rendición de cuentas.

Pienso que es tan creíble lo que difunde la Dirección General de Evaluación (DGE) como lo que pueda difundir el INEE, y si tenemos dudas respecto a ciertas cuestiones de confiabilidad y representatividad en los exámenes, éstas no se resuelven. Si me permiten, voy a hacer un símil utilizando lo que llevan a cabo las empresas.

### ¿Duplicando esfuerzos?

El INEE, tal y como está concebido en su Decreto de Creación, y por lo que aca-

bamos de escuchar, y leído en otros documentos, parece un órgano de auditoría interna, finalmente al servicio del patrón; lo que mucho de nosotros abogamos, es que desde los estados sea un órgano de auditoría externa, independiente del gobierno e independiente del SNTE. Cuando se fundó muchos esperábamos que la Dirección General de Evaluación Educativa, migrara hacia el INEE, pero no. Entonces estamos duplicando esfuerzos, duplicando gastos y eso va en detrimento de las acciones, ese dinero se puede utilizar para otras cosas.

Habría que explotar la experiencia acumulada en la DGE, alcanzar la evaluación censal, salón por salón, ya no sólo escuela por escuela; debo admitir que estaba equivocado antes, porque pensaba que era escuela por escuela, pero cuando veo lo que ya difunde el INEE, es salón por salón.

Esos resultados deberán hacerse llegar a cada escuela y difundirlos en carteles, así los padres se darán cuenta de cómo están las cosas, sería menos intuitiva la manera como califican los padres de familia en la actualidad a una buena escuela.

Casi todos sabemos, quienes tenemos algo de experiencia por haber trabajado en las escuelas, que las buenas escuelas, las cuales demandan los padres de familia, generalmente son urbanas, pero quieren que sus hijos estén ahí y aunque tengan sobrecupo allí los quieren, y eso es porque algo funciona bien. Y como dijo un maestro de secundaria que entrevisté: “Es que la directora es una dictadora, tenemos que trabajar o no funciona, así son las cosas más o menos”, pero como se podrán dar cuenta, eso sigue siendo intuitivo.



Uruguay es un país muy pequeño, de 3 millones y medio de habitantes, que equivale a un estado mediano o chico aquí en México; de modo que los comentarios y aportes que haga, probablemente aplican más a lo que se puede hacer a nivel de los estados que a nivel federal. Creo que si la experiencia de Uruguay puede tener alguna utilidad para reflexionar, es a ese nivel.

### Evaluación como un instrumento de aprendizaje

Cuando hay tiempos cortos no se puede pretender decir mucho, pero la idea principal que me gustaría transmitir es que en el uso de las evaluaciones hay una alternativa sobre la cual uno tiene que decidir. Ésta consiste en, o bien usar la evaluación para ejercer presión sobre los docentes, para señalar a las escuelas, o bien, usar la evaluación como un instrumento de conocimiento el cual permita aprender a hacer las cosas de otra manera.

La visión que tenemos en Uruguay es esta última, es decir, por más presión externa, por más *ranking* de escuelas que pongamos en la prensa, por más señalamientos que hagamos en las escuelas, diciéndoles “esta escuela es mala, no está logrando lo que tiene que lograr”. Si los educadores, si los docentes no pueden apropiarse de los retos de la evaluación y no hay una inversión en mejorar la capacidad de trabajo de los maestros, no habrá cambios.

Un analista francés de organizaciones, hace un planteamiento muy agudo y muy cierto: “la inversión en recursos humanos

es la más costosa y la más difícil de todas, pero cuando se hace bien es la que da mejores resultados”.

Quiero enfatizar esto porque sé que el tema de los resultados por escuela, de los *ranking* y demás, son de los temas muy discutidos aquí en México.

Déjenme traerles la experiencia de los chilenos; los chilenos llevan veinte años haciendo prácticas semejantes y poniendo los *ranking* en los medios de prensa todos los años, lo cual genera un gran escándalo público, que no le hace demasiado bien a la educación. Tiene un efecto demostrado, al cual yo le llamo *efecto de goma de mascar*, el sistema se va estirando cada vez más en sus diferencias.

¿Por qué? Porque las escuelas que tienen un buen resultado en una evaluación, en el próximo curso tienen fila en la puerta, para acceder a ella. Y obviamente, si tengo en una escuela 600 lugares y la demanda es de mil 200 personas intentando entrar, elijo a los mejores, con lo cual me aseguro que en la próxima evaluación me va a ir todavía mejor.

En el otro extremo pasa exactamente lo opuesto: una escuela que atiende a sectores pobres, a la cual le va relativamente mal en la evaluación, pierde a los hijos de las familias más educadas, porque esos son los que reciben la información, los que se dan cuenta, los que tienen medios para mover a sus hijos a otras escuelas, y por tanto, esa escuela la próxima evaluación va a tener resultados peores. Esto está demostrado estadísticamente.

Orlando Meza, un investigador chileno, ha hecho un trabajo en el cual muestra

\*Miembro del Consejo Técnico y Coordinador Nacional de PISA en Uruguay.



que entre 1996 y el 2001 la distancia entre los colegios particulares pagados, y los colegios municipales en Chile, en los resultados de la prueba SIMSE se incrementó la desviación estándar, se agrandó la brecha entre ellos.

Otro investigador, Martín Carnoy, de la Universidad de Stanford, ha demostrado que lo ocurrido en Chile es que se han ido generando escuelas privadas en el borde de las zonas periféricas, lo que él dice es “descreman a las escuelas públicas”; es decir, de las escuelas públicas se llevan lo mejor, y las escuelas públicas, que son las municipales, pierden a los alumnos de familias un poco más educadas y sus resultados caen.

Esto que comento lo están reconociendo los propios chilenos; hoy día existe una comisión que está revisando el SIMSE y el informe de PISA sobre Chile, que es este libro que acaban de publicar ahora, el cual nos llegó hace 15 días; ellos, a partir de los resultados de PISA, justamente lo que detectan, dice: “en segundo lugar, vemos que en Chile, así como en todos los países, el rendimiento de los estudiantes está fuertemente asociado a su nivel socioeconómico; sin embargo, no en todos los países sucede con la misma fuerza. En Chile ésta es una situación real, o sea, la variable que más incide es el nivel socioeconómico promedio del establecimiento”.

Y luego, en sus análisis de las cosas por hacer, dicen: “es necesario generar medidas que reviertan la actual segregación social y tiendan a una mayor integración de los estudiantes de distinto nivel socioeconómico al interior de las escuelas”; es decir, ellos mismos se dan cuenta de que el efecto de este problema, radica en que el sistema se segmenta, las escuelas se hacen cada vez más homogéneas, hay escuelas para ricos, escuelas para pobres, y cuando menos heterogeneidad interna

hay en una escuela, menos fácil es enseñar y lograr mejores resultados.

## Uruguay y el INEE

En Uruguay hemos tenido un camino completamente distinto, y lo relato muy brevemente, nosotros sí hemos hecho evaluaciones censales, pero en un país muy pequeño, por eso digo que puede ser aplicado a nivel de un estado.

Dudo que el INEE desde el nivel central pudiera evaluar bien censalmente; sin embargo, haciéndolo bien y controlando como debe de controlarse la aplicación en todo el país, sí lo puede hacer en los estados.

En Uruguay se ha trabajado en algunos casos con evaluaciones de tipo censal, en otros casos con evaluaciones muestrales, pero las mismas pruebas se entregan a todas las escuelas, para que éstas las puedan aplicar en forma autónoma y se puedan comparar con los resultados nacionales.

Pero se ha evitado todo tipo de *ranking* en la prensa, se le devuelve a cada escuela sus resultados; de modo que cada escuela se pueda comparar con escuelas de similar composición social, es decir, las escuelas que trabajan en medios marginales no tiene sentido que se comparen con las escuelas de otros medios, de los hijos de los universitarios y demás.

Nosotros no tenemos el caso de México, más complejo todavía, porque en Uruguay no hay población indígena que habla otras lenguas, como ocurre aquí.

Entonces, cada escuela se compara con sus similares, pero es una información que manejan dentro de cada una de ellas. La escuela la recibe, la puede socializar con la familia, pero a la prensa se dan resultados nacionales, no se dan resultados para hacer ranking por escuela. Esto, unido a la evaluación autónoma y a un gran esfuerzo, como decía, de inversión en los maes-

tros que trabajan en las escuelas de las zonas muy pobres, ofrecen una buena perspectiva hacia el futuro.

Estoy hablando de programas de capacitación que se realizan en fines de semana fuera del horario escolar, por los cuales el maestro recibe un sobresueldo, una compensación salarial por dedicación de tiempo fuera de su trabajo, focalizado como digo en las escuelas de los medios más pobres.

Esto ha tenido como efecto en seis años –nosotros tenemos tres mediciones comparables en la educación básica, una en 1996, otra en 1999, otra en 2002– un cierre de la brecha entre las escuelas; es decir, mientras en el año 1996 la distancia –nosotros clasificamos a las escuelas según su alumnado y se llama contexto muy favorable, favorable, desfavorable– entre las escuelas de los medios más pobres y las de los medios más ricos era de cincuenta puntos en términos del porcentaje de alumnos que lograban un nivel satisfac-

torio en la prueba, esa distancia se redujo en 1999 a poco más de cuarenta puntos, y a 33 ó 36, respectivamente en lenguaje y matemáticas, en el año 2002.

Se explica porque las escuelas de los medios más ricos han mejorado muy levemente o se han mantenido igual, y las escuelas de las zonas más pobres son las que más han mejorado por efecto de la inversión en capacitación, por efecto de un programa específico que hay en Uruguay de escuelas de tiempo completo en los estratos más pobres.

Con esto termino, me interesa destacar la idea central de que lo fundamental es tener un esquema de evaluación, el cual esté al servicio de aprender cómo hacer las cosas de otro modo, que esté muy asociado a la capacitación de los maestros. Estoy convencido de que mientras uno no haga ese tipo de inversiones, por más presión y *ranking* que pongamos, no se sabe hacer las cosas de otro modo, seguiremos enseñando igual y no habrá cambio.



## FRANCISCO MENDOZA TREJO\*

**A**gradezco al licenciado Felipe Martínez Rizo su invitación para participar como comentarista en esta reunión, a pesar de que también he sostenido una actitud crítica frente a algunos enfoques.

Hay que ponerle nombre y apellido a la educación. La creación del INEE en agosto del 2002 levantó grandes expectativas en el sentido de que ahora sí la sociedad civil contaba con un Instituto que le informaría acerca de la calidad de la educación que

imparten las escuelas en nuestro país.

Han pasado casi dos años y se observa que existen resistencias a informar de los resultados de las evaluaciones a nivel escuela. La información publicada por el INEE relativa a los resultados de las pruebas aplicadas por este Instituto en el 2003, ha sido a nivel nacional o estatal o bien por modalidad educativa.

Nadie se ha sentido responsable de los resultados presentados en el informe y no se han suscitado reacciones sociales ante

\*Director General del Instituto de Fomento e Investigación Educativa A.C. (IFIE).



los bajos niveles de aprovechamiento escolar.

No se ha querido reconocer que las cifras reportadas con un alto nivel de agregación no interesan a la sociedad, porque éstas no identifican a las escuelas responsables de los buenos y malos resultados.

La realidad es que el INEE no desea asumir responsabilidad en la evaluación de las escuelas. Su función la considera como asesora de las autoridades educativas estatales que estén dispuestas a publicar resultados por escuelas.

El énfasis lo ha puesto en desarrollar marcos de referencia, instrumentos para el diagnóstico y evaluación de la organización y funcionamiento de las escuelas, y en la formulación de un sistema de comunicación de resultados de evaluaciones hasta nivel de escuela.

Solo hasta el año entrante contempla aplicar un modelo de evaluación a una muestra de escuelas, y no queda claro si se difundirán los resultados de estas evaluaciones, ya que en su programa operativo anual indica que los beneficiarios de este proyecto son la SEP y las autoridades estatales.

### **¿Es el INEE un organismo gubernamental?**

La discusión de fondo es la relativa a la misión y naturaleza del INEE, ¿es el INEE un organismo gubernamental orientado a servir principalmente a las autoridades educativas o es un organismo orientado a promover la calidad educativa a través de la evaluación y difusión a la sociedad de los resultados educativos de las escuelas? La evidencia hasta la fecha indica que el INEE es más bien un organismo gubernamental.

Para fundamentar esta afirmación, me referiré a los documentos del propio INEE.

En el *Programa Operativo Anual y Presupuesto 2004*, el punto relativo al *Desarrollo*

*del Subsistema de Evaluación de Escuelas*, establece como meta sustantiva desarrollar estrategias, modelos y herramientas para la evaluación de escuelas, las cuales atiendan a las características específicas de las instituciones escolares de los distintos niveles y modalidades de la educación básica, y sean útiles a las autoridades educativas escolares, locales y nacionales, para la toma de decisiones que contribuyan a elevar la calidad de la educación.

Una primera observación es que esta meta no considera generar información dirigida a los padres y a la sociedad en general para dar a conocer la calidad de la educación que imparten las escuelas, su público lo constituyen las autoridades educativas, locales y nacionales.

El mismo patrón se observa en los siete proyectos que pretende desarrollar la dirección de evaluación de escuelas en este año.

Por otra parte, el presupuesto directo asignado a la Dirección de Evaluación de Escuelas es de 600 mil pesos, lo cual representa menos del 0.5 por ciento del presupuesto anual del Instituto que es de 154 millones de pesos. Esto demuestra el interés marginal del INEE en las evaluaciones de las escuelas.

En un segundo documento del INEE elaborado por la Dirección de Pruebas y Medición, intitulado *Plan General de Evaluación del Aprendizaje* de fecha junio del 2004, al tratar el tema de la confidencialidad de los resultados se menciona, "el INEE tiene la responsabilidad de guardar confidencialidad de la información de los estudiantes, familias y escuelas en lo individual, razón por la cual los resultados que se publican no se desagregan a estos niveles".

En el mismo sentido, en la sección *Cómo se pueden consultar los resultados*, se afirma: "el INEE no conserva registros que permitan la identificación de la ejecución individual de los estudiantes, porque las

pruebas del INEE no están diseñadas para evaluar la ejecución de los estudiantes ni de las escuelas en forma individual, en la evaluación del INEE sólo se provee información de la ejecución de grandes poblaciones de estudiantes, tales como la de los alumnos de sexto año de primaria, a niveles nacional, estatal o la de subgrupos de estas poblaciones como la de género o modalidad escolar”.

### **Convertir al INEE en un motor de la calidad en la educación**

Finalmente deseo hacer algunas propuestas que pudieran contribuir a que el INEE se convierta en un motor de la calidad en la educación:

Primero, dotar de plena autonomía al

Instituto y ciudadanizar completamente su Junta Directiva. Nos hemos olvidado ya del tema de la autonomía y de la ciudadanización, pero si pretendemos que la evaluación de las escuelas quede a cargo de los estados, de las direcciones de evaluación o las áreas de evaluación de los estados, tampoco vamos a tener la autonomía necesaria, no podemos ser juez y parte en los procesos de evaluación.

Segundo, que el INEE publique resultados por escuela de las pruebas nacionales e internacionales en que intervenga. Sólo poniéndole nombre y apellido a la educación podemos identificar oportunidades a nivel escuela y lograr un verdadero rendimiento de cuentas.



### **ARTURO DE LA ORDEN HOZ\***

**E**s para mí un gran honor como educador español, dedicado en parte a la evaluación del sistema educativo en mi país, estar aquí con un grupo de colegas tan distinguidos y ante un público que parece ansioso de mayores conocimientos acerca de la evaluación, o de alguna manera contribuir al desarrollo de la misma.

Al reflexionar sobre qué podría decirles a ustedes en un momento como éste, me viene una serie de ideas un tanto raras y dispersas, por ejemplo, pensaba que una de las tareas más difíciles es precisar la situación actual de cualquier faceta social compleja, esto viene a cuento porque el título de las jornadas es situación actual y futura, por qué, pues simplemente por

las múltiples perspectivas e implicaciones que tiene cualquier faceta social compleja, en consecuencia, a lo más que puedo llegar es a una especie de perspectiva en una línea determinada.

Sin embargo dije: parece que es mucho más fácil especular acerca del futuro de esta misma faceta, yo simplemente me invento un futuro y lo proyecto, pero tengo a cambio un gran riesgo, el riesgo de que me pueda equivocar, pero nadie me va a decir algo hasta que no llegue el futuro.

Y así es como he visto el problema de la evaluación, coincide más o menos con los chistes que hacen los economistas acerca de su función; un economista es aquél que dice cosas erróneas acerca de cómo va la economía de su país, pero conoce las ra-

\*Profesor de la Universidad Complutense de Madrid y miembro del Consejo Técnico del INEE.



zones porque se ha equivocado anteriormente, cuando resulta que se ha equivocado sabe cómo explicar por qué se ha equivocado. Más o menos ésta es la situación en que yo me encuentro, créanme.

### La finalidad de la evaluación

La evaluación es un tipo de realidad, incluso un tipo de idea, un tipo de teoría, que tiene algunos elementos comunes más o menos aceptados por todos, por ejemplo: no cabe duda que la evaluación siempre supone un juicio acerca de algo, y a su vez, éste debe estar apoyado en una información fiable y válida.

A partir de esto todo lo que se añada a la evaluación modifica la perspectiva con la que se le miró. Si nosotros decimos: vamos a pensar en la evaluación educativa, pues obviamente tiene que haber un juicio acerca de algún aspecto de la educación y cuanto mejor sea la información que tenemos, mejor cumpliremos con el cometido.

¿Ahora, para qué hago este juicio, es decir, cuáles son los propósitos, cuál es la finalidad de la evaluación? Puedo responder, y eso he hecho en la Universidad de Madrid con mis alumnos: miren ustedes, yo sólo les puedo decir que la función básica, el propósito de la evaluación educativa y de toda evaluación, es optimizar el objeto evaluado. Pero optimizar el objeto evaluado puede hacerse de múltiples formas.

Y aquí vienen todas las posibles posiciones. Piensen, por ejemplo, ¿qué podemos evaluar? Pues el sistema educativo, ¿cómo va el sistema educativo de un país, de un estado, de una ciudad? o puedo pensar que lo que me importa es: ¿cómo van las escuelas de mi barrio o todas las escuelas de un país?

Tengo que definir la expresión: ¿cómo va? Porque una escuela va de una manera, y dice: va bien. Y otra escuela que tiene otras características, dice: también va

bien, o son iguales: una va bien y otra va mal. Pero por qué, porque estamos pensando que hacemos una evaluación para algo y en la perspectiva que lo miramos nos aparecen esas diferencias.

### ¿Qué ha ocurrido en México?

Me parece que toda la historia de la evaluación en este país ha resucitado con la creación del INEE, éste ha creado multitud de expectativas, temores y posiciones para enfrentar situaciones diversas, entonces aparecen opiniones múltiples.

Lo más importante sería que el INEE defina bien qué quiere evaluar y cómo, y si eso es lo que le encargan lo haga. Así no se le pedirán más responsabilidades de las que tiene, porque sí les quiero decir, que para mí es evaluable todo en el sistema, todo; pero probablemente por evaluar todo, no evaluaría nada.

¿Qué quiero decir? Que el INEE no es un organismo absoluto que vaya a dictar todo lo que se tiene que evaluar. A mí me parece que esa es la idea fundamental, incluso es la que mantengo como consejero técnico: el INEE tiene la obligación de evaluar el sistema educativo mexicano, eso no implica no evaluarse él mismo.

Sin embargo, quien no debe rendir cuentas acerca del estado de la educación es el INEE. El Instituto lo único que dice es: "la educación que yo mido es así". Y al ser así, quien haya sido culpable de esto, que haga rendimiento de cuentas a la sociedad, al parlamento, a los padres, a quien deba hacerlo.

En cuanto tenemos un universo grande, una escuela no es cualquier cosa, es una entidad complejísima; y para hacer un juicio sobre una escuela, se necesita algo más que la media en matemáticas de los alumnos que están en la escuela. Sin embargo, para informar cómo rinden los alumnos de México en matemáticas al final de primaria, sí sirve hacer una buena

prueba con todos los requisitos técnicos, psicométricos, etcétera.

Este tipo de instituciones han surgido en Europa y en España para evaluar el sistema; lo mismo que surgió en Estados Unidos en donde se evalúan también una por una las escuelas, pero no las evalúa la misma gente. Ésta es la cuestión.

No pretendo entrar en ningún tipo de discusión, pero evaluar bien las escuelas es, quizá por su complejidad, muchísimo más exigente porque son muchos más aspectos, no porque tenga que evaluar bien el nivel de rendimiento en matemáticas de todo el país.

Por consiguiente, seamos realistas y exijamos al INEE lo que realmente tiene

que hacer. Y si en algún momento la sociedad le encarga que evalúe escuelas, tendrá que evaluar escuelas, y entonces dirá “si quieren que evalúe el rendimiento en matemáticas y las escuelas, necesito este tipo de cosas”.

Hay un aspecto importante, creo se ha dicho aquí y quiero reconocerlo: el INEE debe de ser independiente y autónomo, pero no por la razón expresada aquí, sino simplemente porque un evaluador no independiente no es un evaluador fiel.

Por tanto, la independencia del evaluador es una exigencia *sine qua non* para evaluar. Eso está claro, pero no como un requisito político, sino como un requisito de la propia tarea de evaluar.



## PREGUNTAS Y RESPUESTAS

### **ANA MARIA ACEVES ESTRADA**

**Usted comentó que la evaluación debe de ser un proceso permanente con uno mismo, entonces ¿dónde caben los resultados de la OCDE, los cuales colocan al país en los últimos lugares?**

Precisamente en la mesa se ha identificado que la Dirección General de Evaluación había venido tratando de realizar la evaluación del Sistema Nacional de Educación. Al dejar esta tarea en manos del INEE, la dirección puede reorientar sus funciones a la identificación de los actores y llevar los resultados de la evaluación a los maestros –los cuales son los motores del cambio– director, jefe de sector, supervisor, técnico pedagógico, entre otros, para retroalimentar su trabajo

**¿Cómo contempla el INEE o la Dirección General de Evaluación coadyuvar**

**para que los maestros utilicen los resultados en forma colegiada?, pues se pretende que estas evaluaciones de diagnóstico censal no controladas, que van a ser instrumentos de ayuda al maestro, puedan ser utilizadas por el propio director de la escuela y las autoridades de la supervisión escolar y los Apoyos Técnicos Pedagógicos (ATP) que las autoridades educativas tienen en las entidades.**

Lo que pretendemos con estos resultados es generalizar, compartir los modelos de las escuelas exitosas y qué problemas se encuentran en los planteles educativos para compartirlos con el resto de las escuelas, y generalizar un método de gestión escolar para que maestros y directores encuentren orientaciones para detectar y resolver sus propios problemas.

El trabajo al que se hace mención habla de un proceso de evaluación, cómo se



ve la descentralización de estas funciones y cuáles son los principales retos de la coordinación entre la dirección y las autoridades educativas. La intención de la dirección general es fortalecer las capacidades y equipamientos de las autoridades educativas, y en esto vamos juntos con el INEE, en el sistema nacional de evaluación.

**¿Cómo poder integrar los datos de las escuelas de calidad, las evaluaciones cualitativas a nivel nacional para que sean representativas, si éstas solamente son un diez por ciento?**

El sistema nacional de evaluación es el lugar donde vamos a llevar a cabo esto, coordinando las iniciativas locales para orientar los criterios de calidad, y por mucho que se den iniciativas locales, éstas no van a poder tener términos de referencia o comparabilidad nacional.

Así que, en este sentido, los programas nacionales van a poder tener aportaciones para que una escuela se compare no solamente consigo misma, sino que también vea cómo van los resultados por escuela, zona, municipio, estado y a nivel nacional.

**SALVADOR MALO ÁLVAREZ**

**¿Cómo ha impactado la evaluación educativa a la selección de aspirantes e instituciones de educación superior?**

La evaluación educativa ha impactado en la selección de aspirantes de muy diversas formas, lo ha realizado en la reducción de los distintos exámenes que los jóvenes hacían ante varias instituciones, y ahora tengan evaluaciones de carácter estatal o concursos, para reducir el número de exámenes a presentar, el costo y el esfuerzo que esto representa.

Además ha contribuido a la mayor equidad, la cual permite comparar a jóvenes en condiciones en que no influyen otras características, excepto aquella que es la re-

sultante de una evaluación, en este caso una evaluación estatal o de la institución.

Creo que en muchos casos, desgraciadamente no tantos como nos gustaría, y los cuales cada vez son más crecientes, hay mayor reflexión de las propias instituciones educativas sobre las características de los estudiantes que están llegando y sobre su comparación con el resto de otras instituciones y espero que eso, en algunos casos lo sabemos, lleve a modificaciones en los procesos educativos y en las tareas que ellos tienen que hacer a largo plazo.

**¿Cómo ha impactado el ingreso al campo laboral?**

Desde el punto de vista del ingreso al campo laboral, el impacto que ha tenido la evaluación todavía es muy pequeño, es poco lo que se ha hecho, pero lo que está sucediendo en el caso del egreso profesional es un incremento de la oferta de profesionistas en las universidades y por lo tanto, incrementa la competencia por los empleos, esto llevará sin duda a que haya cada vez mayor necesidad de presentar evidencias de calidad o de preparación para obtener un empleo. En ese sentido creo que las evaluaciones también influyen.

**¿La evaluación es una parte del proceso educativo o es la meta del proceso educativo?**

La respuesta es la primera. Claro que es parte del proceso educativo. Generalmente digo que en el caso particular del Ceneval el símil más cercano que podemos hacer es el laboratorio médico de un sistema educativo. En el caso del laboratorio médico, éste nos indica los distintos exámenes que debemos aplicar, y nos informa si tenemos PH, cierto nivel de PH o lo que ustedes quieran. El laboratorio médico no hace el diagnóstico y menos aún aplica el tratamiento.

El Ceneval es en buena parte eso, y por lo tanto, los resultados de la evaluación educativa que en el caso particular del Ceneval aplicamos, corresponde a las autoridades, maestros, padres de familia y a toda la sociedad el decidir cómo utilizarlos, para qué utilizarlos. Trataremos de orientar y apoyar con más información, de tal manera que podamos contribuir a que se lean más claramente los resultados. Igual que hacen los laboratorios cuando dicen: el rango es de tal a tal. Eso trataríamos de hacer en casi todas las acciones que desarrollamos.

### **MARIO RUEDA BELTRÁN**

**Algunos profesores manifiestan su rechazo a la evaluación educativa, porque creen que si obtienen malos resultados perderán su empleo. ¿Cómo garantizar a los profesores que no perderán el empleo si dieran malos resultados?**

Esta pregunta me permite hablar de un contexto en donde generalmente la evaluación se plantea con el único propósito de premiar o castigar, y sucede en gran medida por la tradición, y por haber asociado evaluación con calificaciones o con resultados en el caso de los estudiantes de diferentes grados y programas de formación.

Esta situación primero nos habla de la gran dificultad de la delimitación. Por ejemplo, si utilizamos la calificación como un indicador del aprendizaje podríamos concluir muchas cosas, una escuela que en promedio está sobre nueve de calificación y otra sobre seis o siete, diríamos tomando en cuenta ese indicador que, en una se sabe más que en otra.

Sin embargo, este sería un juicio sumamente apresurado porque no sabemos si esa calificación, primero, está reflejando aprendizaje. Todos conocemos las prácti-

cas que confunden en muchos profesores la disciplina con el aprendizaje.

Por ejemplo: tu levantaste la voz, tienes un punto menos en la nota; es decir, una acción disciplinaria que afecta el indicador que está supuestamente queriendo decir qué tanto sabe el estudiante, pero podríamos dar otros ejemplos. En el caso de escuelas en donde hay una política de sobreprotección hacia el estudiantado, a lo mejor porque hay cuotas muy elevadas, podría favorecerse una tendencia a calificar muy bien a los estudiantes.

En el caso de un profesor muy exigente puede haber un seis equivalente a diez o incluso a 11, en la escala de otras escuelas. Es decir, necesitamos trabajar mucho en lo que es la medición; saber cómo se mide y qué estamos midiendo para no confundirnos.

Creo que esta perspectiva nos habla de conceptualizar a la evaluación como una posibilidad de premiar, castigar o de señalar culpables, y nos alerta de que efectivamente es un gran riesgo solamente señalar, porque nos lleva a esos extremos bien ejemplificados hace rato por mi colega uruguayo.

Incluso muchos autores hablan de esta gran disyuntiva ante la evaluación, la evaluación que solamente selecciona, clasifica, señala y tiende a la exclusión; es decir, yo tiendo a hacer la evaluación para quedarme con los mejores estudiantes, con los mejores profesores, con las escuelas que tienen las mejores condiciones; o la otra, en la cual la institución se hace cargo de qué es lo que pasa con sus profesores, con sus estudiantes; la que se responsabiliza del proceso: yo recibo así a los estudiantes, tengo este tipo de profesores con este tipo de comportamiento y qué hago como institución para mejorar el rendimiento de los profesores y de la escuela misma. Creo que en esta segunda perspectiva deberíamos de caminar.



**CARLOS ORNELAS NAVARRO****¿Qué se debe entender por cultura de evaluación y cómo alcanzarla?**

La cuestión de la cultura de la evaluación es de una trayectoria larga, pero en México empezó a calar hondo con el Programa de Modernización Educativa del presidente Salinas, donde la plantea como una de las metas principales; es más, si pudiéramos sintetizar en una palabra la política de ese gobierno para la educación superior, sería evaluación y hubo muchas cosas que se hicieron, se instituyeron nuevos centros y un montón de cosas más.

**¿La rendición de cuentas va dirigida a la certificación laboral?**

No, son dos cosas distintas. Cuando hablaba de la composición de la Junta Directiva, no tengo nada contra las personas, es precisamente la institucionalización, o sea es el Secretario de Educación Pública el presidente, ahorita es Reyes Tamez, pero el día de mañana, en el 2007 va a ser otro.

En lo referente al Consejo Técnico, lo he puesto por escrito, hay personas honorables, son expertos, tienen amplio campo, pero su límite de acción es muy reducido y esto tiene que ver con la siguiente pregunta, que compartimos.

**¿Estamos de acuerdo en que el INEE sea autónomo?**

Por supuesto, abogamos por esa autonomía, Francisco le llama ciudadanización, o sea hay coincidencia en cuanto al juicio que hacemos, yo lo veo también por cuestiones de confianza, de confiabilidad que es la palabra clave para que sea legítimo.

Sobre la cuestión de los *ranking*, en un primer momento, si se hacen los favorecidos son las clases medias, los ricos no, porque van a escuelas privadas; son las clases medias quienes se verían favorecidos y a los pobres no les quedaría más remedio

que ir a la escuela que tienen ahí.

Pero también en una encuesta que se hizo, ésta arrojó que el 58 ó 59 por ciento de los padres de familia, dicen que realmente inscriben a sus hijos en la escuela que querían inscribirlos, o sea que no hay todavía muchos problemas y el resto es porque no les quedaba de otra; en la escuela comunitaria, en el desierto, en la montaña, en las de educación indígena, en las multigrado, no hay de otra y la evaluación servirá para mejorar también aquellas escuelas.

Entonces yo no hablo de *ranking* pero sí defiendo el derecho de los padres a saber el estado de las escuelas, el conocimiento de los maestros y los resultados que obtienen sus alumnos.

**PEDRO RAVELA CASAMAYOU**

Me hacen una pregunta y voy a hacer dos comentarios más.

**¿A qué me refería con lo de las evaluaciones censales?**

La expresión de evaluación censal se usa cuando todas las escuelas y todos los alumnos de un determinado grado son evaluados por oposición a una evaluación en que sólo se evalúa una muestra. La ventaja de una evaluación censal es que todas las escuelas están involucradas y uno puede conocer resultados escuela por escuela, mientras que cuando es una evaluación de tipo muestral, el noventa por ciento de las escuelas mira de afuera y no está involucrada y difícilmente se involucra, el problema son los costos, una evaluación censal es muy costosa.

En Uruguay, lo que hemos hecho son evaluaciones muestrales con una muestra pequeña, pero esas muestras luego se entregan a todas las escuelas, y cada escuela las puede aplicar o corregir con los mismos criterios que se corrige la muestra nacional, y recibe los resultados de la muestra nacional. Entonces tiene un elemento ex-

terno de comparación que nosotros creemos es muy importante, para saber cómo está en relación con otras escuelas.

Me gustaría resaltar dos comentarios, los cuales creo son muy importantes: uno es muy complejo, no tiene solución fácil y la experiencia internacional muestra su complejidad, es la cuestión de la autonomía de un instituto de evaluación.

La mayor parte de los países cuenta con instituciones que de alguna manera u otra están vinculadas al ministerio de educación, obviamente pueden tener distintas integraciones y en eso no me voy a detener. Uno puede decir: una Junta Directiva debe estar integrada con más o menos peso de la sociedad civil, con menos gente del Estado. Lo que a mí sí me interesa resaltar, es que finalmente hay un problema, el problema de fondo se resuelve con una cultura de la independencia de la unidad de evaluación, más que por una solución formal; porque en última instancia siempre hay un tema de recursos, hay países como Honduras, por ejemplo, donde encargaban la evaluación a una universidad completamente autónoma del ministerio, pero ¿los recursos quién los ponía? El ministerio, y cuando el ministerio no quiso o no le gustaron los resultados que daba la Universidad Pedagógica Autónoma, simplemente le cortó los recursos y la evaluación se cayó.

En Estados Unidos es igual, el NAEP lo opera el ETS; pero los fondos los pone el gobierno. Pero ¿qué hay en los Estados Unidos? Hay una cultura instalada y eso es intocable y si lo tocan se arma un gran escándalo. Lo mismo pasa en Chile. El SIMSE forma parte del ministerio, pero hay una cultura de autonomía del SIMSE. Igual sucede en Uruguay. Nosotros somos parte de la Administración Nacional de Educación Pública, pero el Director Nacional de Educación, que es como si fuera el secretario aquí, recibe los resultados tres días antes, cuando está fijada la fecha de publicación de los resultados,

si se entera así ya no tiene modo de intervenir ni de cambiar nada.

También hay experiencias opuestas. Por ejemplo, Perú, donde el ministro recibe los resultados y dice: “Esto no se publica porque nos deja mal parados”.

Entiendo que hay suspicacia en México, porque los resultados fueron durante mucho tiempo cuestión secreta, no se publicaban. Pero salvo la situación de que una ONG, una institución civil, tuvieran los recursos para montar un sistema de evaluación completamente autónomo, no habría autonomía perfecta para evaluar de manera externa y absoluta al sistema.

Creo que la participación en las evaluaciones internacionales es otro modo que garantiza la externalidad de la evaluación, porque sobre todo en la Evaluación PISA, están los controles que garantizan que no puedan manipular los resultados ni el ministerio ni la autoridad local.

El último comentario muy breve, es que efectivamente, y me gustaría resaltar el planteamiento que hacía el doctor Carlos Ornelas, hay un problema muy serio en la comunicación de resultados, y es cierto que la comunicación de resultados a través de promedios, de teorías respuestalistas, que tienen una media de 500 y pueden variar de distintos modos, muchas veces no son comprensibles para la ciudadanía en general.

En México me tocó hacer un estudio a nivel de toda América Latina, de cómo la prensa reporta los resultados. Por ejemplo, aquí en México, el diario Reforma interpretaba los puntajes de teoría respuestalista como si fueran aciertos, y decían: “México quedó a 250 aciertos de los coreanos”. Lo cual es un disparate, porque no son aciertos, no es que cada alumno responde 500 preguntas ó 700 preguntas.

Entonces es un tema muy complejo sobre el cual hay que tener conocimiento para poder opinar, y que se resuelve



por un único camino, y ese es el que toma PISA, el cual transforma los puntajes en lo que se llama niveles de desempeño. PISA transforma los puntajes abstractos en distintos niveles.

Hay un nivel uno, son los alumnos que prácticamente no entienden lo que leen, los cuales sólo hacen las tareas más simples; hay un nivel dos, son alumnos capaces de hacer cosas un poco más complejas, hay un nivel cinco, son los alumnos capaces de realizar las tareas más complejas, etcétera.

Y el mejor modo de reportar, a mi juicio y el más comprensible, es informar qué porcentaje de los alumnos están en cada uno de los niveles. Que es lo que hace finalmente PISA.

Lo que hacemos en Uruguay, nosotros no reportamos con promedio ni nada por el estilo, sino ponemos un nivel de suficiencia, y decimos: de los alumnos uruguayos en matemáticas, el 35 por ciento en 1996 está logrando el nivel satisfactorio, los demás no. Creo que ese es el modo más fácil para que a la opinión pública y al magisterio den a conocer los resultados y éstos sean comprendidos; creo que es un punto muy importante el de la necesidad de traducir del lenguaje técnico a modos comprensibles de informar sobre los resultados.

### **FRANCISCO MENDOZA TREJO**

#### **¿Al ciudadanizar el Consejo Técnico no se caería en la anarquía?**

Bueno, efectivamente, se tiene que tener mucho cuidado en la selección del Consejo Técnico. A mi modo de ver, se podría hacer considerando las organizaciones sociales civiles relacionadas con la educación, por ejemplo, la Asociación de Padres de Familia, empresarios, diferentes sectores sociales deberían tener organizaciones de calidad, no todas exactamente del área de educación.

#### **Considerar si el INEE debería ser descentralizado como el IFE, ¿cómo podría llevarse a cabo esto como institución del gobierno antes del 2006?**

Bueno, no creo que la solución sea replicar 32 *INEES* o 32 *IFES*, o como les queramos llamar. Creo que en cada estado de la República, si queremos respetar la federalización, se deben de crear institutos independientes de evaluación, apoyados o asesorados por el INEE, pero que fueran ciudadanos de esos estados los que estuvieran en la Junta Directiva.

#### **Y respecto a ¿cómo lograr la autonomía?**

Los fondos con los que funcione este Instituto tendrían que ser asignados por el Congreso, no por el gobierno, porque de otra manera le cortamos la independencia.

### **ARTURO DE LA ORDEN HOZ**

No tengo preguntas, eso quiere decir que los he convencido a todos, pero sí quisiera terminar con una precisión.

Cuando he dicho que no se pueden hacer informes escuela por escuela, cuando todo se hace para evaluar el sistema, eso quiero que quede claro, porque en la evaluación del sistema, tal como se hace ordinariamente, se hace con una muestra, la cual se supone es totalmente representativa de la población. Pero no quiere decir que podamos decir algo acerca de una escuela que no esté en la muestra, por lo que digamos de una escuela que está en la muestra.

Por consiguiente, no podemos comparar los datos que tenemos, son datos de rendimiento en las materias matemáticas, comprensión lectora, etcétera, y otros datos de tipo de contexto, o sea, algunos aspectos de cómo funciona pedagógicamente

te el aula o cómo está dirigido tal o cual aspecto de los alumnos, de las familias, que son lo que se llama cuestionarios de contexto. Con estos datos, a nivel muestral sería no sólo ilegítimo, sino verdaderamente injusto, poder ponerse a calificar.

Incluso las escuelas que están dentro de la muestra, qué se puede decir, están ahí elegidas no para ser evaluadas una con otra, sino para aportar datos acerca de cómo funciona el sistema. Por consiguiente, no sería legítimo utilizar esos datos como una calificación de las escuelas y por supuesto pienso yo que esta ilegitimidad llevaría a una injusticia. Se pueden –como he dicho– evaluar las escuelas, pero de otra manera y totalmente al margen de la evaluación del sistema.

En la evaluación del sistema se obtienen datos acerca de cómo son las escuelas que producen esos datos, pero en general pueden decir que las escuelas de tales regiones obtienen mejores rendimientos en uno o en otro rubro, las que tienen tales características de dirección aparece que tienen unas puntuaciones más altas o más bajas, nada más que eso, pero son características que afectan a todo el rendimiento del sistema.

Esto es lo que deseaba precisar, porque no es solamente que no tengamos –por supuesto sería una tarea verdaderamente extraordinaria– un tipo de pruebas para todos los alumnos de todas las escuelas, sino incluso, si quisiéramos hacer algún tipo de reflexión sobre las escuelas, que fuera una evaluación de la escuela un poco más allá de lo que se puede decir, no sería correcto.

## **FELIPE MARTINEZ RIZO**

**Hubo varias preguntas muy detalladas sobre si se evaluarán ciertas materias, como educación física, tecnológica; si se evaluará a adultos, etcétera.**

Mi respuesta va por este lado, coincidiendo con lo que dijo en algún momento el doctor de la Orden, todo se puede evaluar, pero no de la misma manera ni por la misma instancia.

Entonces, el Instituto no pretende evaluar todo, porque no sería posible, pero precisamente por eso creemos que debe haber evaluación por parte de diferentes instancias, así cada cual atenderá lo que le corresponde, de una manera congruente, para que entre todas obtengamos la información más completa.

### **¿Qué opino sobre el acuerdo 200, que es el que define las normas para evaluación en educación básica?**

Me referiré a la cuestión de la interpretación de las evaluaciones, un poco en la dirección señalada por Pedro Ravela, aparte de que sí hay que hacer un esfuerzo por poner en términos inteligibles escalas complejas, también hay que desconfiar de las escalas simples, ese es el otro punto.

Se decía que estamos muy acostumbrados y entendemos muy bien una escala de cero a cien o de cero a diez, cuidado, porque lo entendemos muy clarito, pero mal, o sea, nos engaña; decimos “si sacó diez es muy bueno, si sacó cinco reprobó y si sacó ocho está en medio”. Y efectivamente, a lo mejor el que sacó el ocho está mejor que el que sacó diez en la otra escuela y el que sacó cinco en la otra escuela no hubiera reprobado, y el que sacó seis sí hubiera reprobado en otra escuela, etcétera.

Entonces, estas escalas son muy engañosas y parte de lo que debemos cambiar para acercarnos a una buena cultura de la evaluación es hacer conciencia de la necesidad de evitar interpretaciones simplistas, las cuales parecen muy obvias pero pueden llevar a errores y precisamente el punto fundamental es: la evaluación es algo muy importante, sí, pero es algo complejo y delicado. Si se hace bien puede ser muy



provechosa, puede ser muy útil para la mejora. Si se hace mal puede ser destructiva, negativa.

Como es un proceso tan complejo no lo puede hacer una sola instancia, sino que tienen que colaborar varias y de hecho la instancia fundamental es el maestro. Por más que queramos hacer evaluaciones en gran escala y podamos llegar a hacer evaluaciones que se apliquen a todos los niños y demás, en forma no controlada, siempre será insuficiente. El maestro es el único que puede tener el pulso fino, pero necesita hacerlo bien, estar preparado para ello, mejorar su nivel de profesionalización.

### **¿Cómo influye la educación en la mejora?**

Coincido con lo expresado por Salvador Malo, los sistemas de evaluación son laboratorios clínicos, no son el médico; el médico es el maestro. El laboratorio clínico no le dice al médico: tiene usted un paciente con tanto de azúcar y por lo tanto recétele tal cosa, el médico es el que va a decir qué le da porque ese paciente además puede tener ocho años ó 65, puede estar obeso o flaco, puede tener asma o caries, y sólo el médico profesional que atiende al paciente, quien conoce el caso, puede utilizar adecuadamente el resultado del laboratorio clínico para enriquecer su diagnóstico y aplicar el tratamiento apropiado.

Entonces, coincido con lo que decía Pedro, ¿para qué es un sistema de evaluación? Si pretendemos realizarlo para mejorar todas las escuelas tendrá que ser un sistema enfocado a eso, el cual se conciba, diseñe y planifique de tal manera que le aporte elementos para mejorar la calidad profesional de cada docente, escuela y director, para lograr mejor educación.

El punto al que llego con esto es al de resultados por escuela. Quisiera subrayar que la posición del INEE no ha cambiado desde su nacimiento hasta ahora, seguimos manteniendo la misma postura. El

apoyo a la Subsecretaría del Distrito Federal para desarrollar el sistema de resultados por escuela, que se va a presentar enseguida, no comenzó ayer, eso lo empezamos hace más de un año porque estábamos convencidos que así debíamos hacerlo, pero con base en los resultados que se pueden utilizar para ello, los cuales son de aplicaciones que cubren todas las escuelas censales, como decía Pedro Ravela.

Los sistemas de evaluación internacional en ningún lugar del mundo se usan para dar resultados escuela por escuela, porque son muestrales. Los sistemas americanos estatales que dan resultados por escuela, lo hacen así porque evalúan a todas ellas. El NAEP del que hablé en mi presentación empezó en 1969 y sigue hasta la fecha sin dar resultados por escuela, pero además la ley dice que es delito dar resultados de las evaluaciones del NAEP por escuela, y las personas que hagan esto pueden ser multadas o puestas en la cárcel. Además añade la información que nunca ha ocurrido desde que existe el NAEP; o sea, los resultados de evaluaciones muestrales no deben darse por escuela, sería totalmente incorrecto, injusto, inapropiado.

### **¿No vamos a tener resultados escuela por escuela?**

Sí, nada más que no de las pruebas muestrales, siempre ha sido nuestra postura. Sólo tendremos resultados de las pruebas que se puedan aplicar censalmente y las cuales tengan elementos que nos ayuden a contextualizar y se puedan hacer varias mediciones, qué bueno que las podamos dar por escuela aunque tengan algunas limitaciones, el INEE está totalmente de acuerdo con apoyar a los estados que así lo soliciten.

Ahora haré algunos comentarios: efectivamente, coincidimos con el maestro Mendoza, no consideramos que deba de haber sucursales del Instituto en los estados, sí

creemos que cada uno de ellos deba de tener sus propios sistemas de evaluación y sus propios institutos si los quiere crear.

También coincidimos con el maestro Ravela y con el doctor de la Orden en el sentido de que lo fundamental para que haya independencia no es la estructura jurídica y la composición de un órgano es, por un lado, la solidez técnica del organismo.

Nuestra Junta Directiva de hecho tiene representación de los padres de familia, representación del sector empresarial, representación de organismos independientes como Transparencia Mexicana, Observatorio Ciudadano de la Educación, los cuales hacen que aunque obviamente haya presencia efectivamente de autoridades, el Secretario de Educación o del sindicato de maestros, la representación del gremio magisterial, ninguno de los integrantes de la Junta Directiva pueda tener control de las decisiones. Uno solo de sus integrantes tiene el derecho y la capacidad, en un momento dado de poner en cuestión una decisión, si fuera tomada de manera incorrecta, si no se procediera con rigor, aunque no tenga mayoría numérica.

Por otro lado, un organismo como el Consejo Técnico, el cual cuenta con expertos nacionales e internacionales, obviamente no es susceptible de recibir indicaciones de la autoridad educativa. Efectivamente las personas pueden cambiar, a lo mejor el día de mañana o unos años después, pero creo finalmente que las instituciones no deben de tener fe ciega e incondicional en una persona, se tienen que basar en sistemas, en mecanismos institucionales, y por eso estamos hablando del *Modelo de Calidad del INEE*, el cual ya existe en una versión para difundir que pondremos rápidamente en nuestra página de Internet, para así definir estos estándares de calidad, sus mecanismos y la transparencia de resultados.

La base de datos del año pasado fue la primera que presentamos, en ella se en-

cuentran registrados y señalados perfectamente dato por dato, alumno por alumno y queremos que así sea, todos nuestros sistemas, nuestros productos se han puesto lo más rápidamente posible a la disposición del público, de tal manera que cualquier persona pueda replicar los análisis que hagamos para que corrobore si están bien hechos o no, con la conciencia de que habrá errores, en los elementos que hemos difundido ha habido errores, los hemos señalado nosotros mismos cuando los detectamos porque creemos que eso no nos debe de asustar, si quisiéramos esperar a no tener ningún error, nunca daríamos resultados.

Tenemos que dar resultados rápidamente, de tal suerte que si hay errores, alguien nos ayude a detectarlos, porque si no los presentamos públicamente ni siquiera los vamos a detectar, el hecho de exponerlos al público permite que alguien nos los señale si nosotros mismos no los detectamos, y obviamente esto deberá de agradecerse, porque nos ayudará a hacer las cosas cada vez mejor en un proceso que nos va a llevar tiempo, porque lograr un sistema realmente de buena calidad, no es algo que se pueda lograr de un día para otro.

Por eso reitero el planteamiento que hice al principio: creemos que estamos trabajando con la mayor intensidad, con la mejor voluntad de hacer un buen sistema, creemos en las diferencias de criterio, pues obedecen a la necesidad de escudriñar más en detalle estos elementos tan complicados, no es fácil llegar a acuerdos, pero creemos que ésta es la manera de hacerlo, por ello creemos que en estas jornadas, cuyas siguientes mesas van a centrar la atención en alguno de los puntos más polémicos, precisamente nos ayudarán a todos a irnos acercando a una comprensión común más rica de estas cuestiones. Creo que es lo que necesitamos.



Aspecto general de la sesión en que se abordó el tema del Presente y Futuro de la Evaluación Educativa en México.



Enorme interés generó entre los distintos miembros del sector educativo el desarrollo de las *Jornadas de Evaluación Educativa*.



SEGUNDA MESA REDONDA

## **El Sistema de Información para la Mejora Educativa**

**Ponente:**

Lic. Antonio Ávila Díaz

**Presentación:**

Lic. Sergio Martínez Dunstan

**Comentaristas:**

Mtro. Rafael Vidal Uribe

Dra. Beatriz Calvo Pontón

Dr. Mauricio Fortes Besprovani

Mtra. Sylvia Schmelkes del Valle

Mtra. Giovanna Valenti Negrini

**Moderadora:**

Dra. Guadalupe Ruiz Cuéllar





## El Sistema de Información para la Mejora Educativa

Por ANTONIO ÁVILA DÍAZ\*

Para la Subsecretaría, un propósito prioritario en su periodo de gestión es el producir información para la toma de decisiones en la escuela. En la presentación del *Sistema de Información para la Mejora Escolar* comenzaré por citar algunos antecedentes.

Las administraciones anteriores habían hecho un enorme esfuerzo, muy loable, hay que reconocerlo, por producir y organizar estadísticas alrededor del sector educativo en el Distrito Federal; sin embargo, el énfasis estaba todavía, en aquellos tiempos, dirigido hacia la cobertura y no tanto hacia la calidad.

Esta administración ha tenido como propósito, centrar más su atención en los asuntos de la calidad, por tal motivo nos dimos a la tarea de producir información, la cual permitiera a la escuela obtener un mejor funcionamiento.

Partimos de bases de datos que contenían información dispersa y con pocas posibilidades de ser utilizadas para la toma de decisiones y rendición de cuentas, y las bases de datos a disposición nuestra, estaban más orientadas a los aspectos de administración y no tanto a la mejora de la escuela.

### Construyendo un Sistema de Información

Por lo mismo, se definió como un objetivo de política para el periodo 2001-2006,

construir un sistema de información para la toma de decisiones y la mejora en la gestión escolar en la rendición de cuentas.

Esto se ha venido realizando durante los tres años y medio de nuestra gestión, y al día de hoy hemos logrado algunos avances en lo que llamamos subsistemas de información, los cuales atienden propósitos específicos y de acuerdo con diferentes niveles de mando.

Por ejemplo, contamos con un sistema para las inscripciones en línea de las escuelas normales; un sistema en línea para la inscripción de los docentes a los Centros de Maestros; hemos diseñado y avanzado en el pago electrónico de la nómina a los trabajadores de la educación en el Distrito Federal, que es un número significativo.

Como consecuencia de estos objetivos de política educativa, diseñamos y produjimos el Sistema de Información para la Mejora Escolar (SIME), el cual proporciona información para los mandos escolares, es decir, directores, supervisores y jefes de sector; información útil, comprensible y sistemática, con el propósito de mejorar de manera continua los resultados de los alumnos.

El objetivo es generar información para la toma de decisiones en la escuela, para poder validar y difundir ampliamente los resultados educativos de alumnos y escuelas. Para ello hemos diseñado una estrategia, la cual consta de tres partes.

La primera, ya concluida, consistió en organizar nuestra información, de manera

\*Director General de Planeación, Programación y Presupuesto de la Subsecretaría de Servicios Educativos del Distrito Federal, de la SEP.



que se puedan tener las calificaciones de los alumnos en línea, esto lo hacemos bimestralmente, y en el fin de curso. Así, todos los padres de familia pueden consultar las calificaciones de sus hijos.

Hemos reunido un conjunto de información sobre los planteles, de gran utilidad para la gestión directiva, comenzando por la georreferenciación, esto es, la ubicación geográfica del plantel y todo lo contenido en él: activos físicos y humanos en un plantel, información acerca de sus docentes, alumnos e información sobre infraestructura y equipamiento del cual dispone dicho plantel.

### **Producción del SIME**

Concluida la primera etapa, nos dimos a la tarea de avanzar en la segunda, la cual consistió en producir el SIME, consistente en el desarrollo de instrumentos para la mejora continua, el cual veremos puesto en línea en unos minutos más.

La tercera etapa, consistirá en elaborar un sitio Web para difundir las mejores prácticas derivadas de la reflexión y el análisis de lo que el SIME nos ofrezca. Esto está en construcción, comenzará a partir del próximo ciclo escolar, el SIME lo acabamos de concluir, de hecho no ha sido liberado al público porque estamos haciendo los últimos ajustes.

Junto al interés por generar y ofrecer información útil para la escuela, debo comentarles que se ha hecho un conjunto de encuestas a profesores y usuarios, para contar con una mayor información, la cual nos permita interpretar la vida de nuestras escuelas. Como ejemplo, puedo mencionar los trabajos que se están llevando a cabo con el Fondo Sectorial de la Subsecretaría de Servicios Educativos del Distrito Federal, con Conacyt; y citar una de las investigaciones en curso: *La categorización de las escuelas se-*

*cundarias con base en su desempeño bruto y neto.*

Similar a ésta, hay una decena más de trabajos relacionados con el estudio y reflexión acerca de lo que ocurre en nuestros planteles, lo cual seguramente nos llevará a encontrar muchas pistas de cómo mejorar la vida en los mismos.

El propósito del SIME parte de la intención de conocer cómo avanzamos para poder generar una oportunidad de mejora. Partimos de la idea de avance continuo en nuestros planteles, sabedores de que el conjunto de escuelas y alumnos debe mejorar y toda la comunidad escolar aprender. A partir de ello nos dimos a la tarea de construir el SIME, con la intención de satisfacer tales premisas.

Estas *Jornadas* organizadas por el INEE, son el escenario propicio para agradecer la colaboración que hemos recibido de su parte en la construcción del SIME, ha sido una colaboración entusiasta, generosa, eficaz, y lo que hemos podido realizar ha sido gracias a la participación de ellos; expreso a Felipe Martínez Rizo y a sus colaboradores nuestro agradecimiento por el empeño que han tenido en apoyarnos.

### **Cómo funciona y qué es el sistema**

El SIME utiliza la información de los resultados de los alumnos de cada escuela empleando el *Examen de Diagnóstico para el Ingreso a la Educación Secundaria* (IDANIS). Esta información la ordenamos, analizamos y ponemos a disposición de los mandos medios de nuestra estructura operativa y de la sociedad en general. El modelo trabaja con información censal, multianual y de contexto. Todo esto nos permite identificar la situación de cada escuela y su posibilidad de mejora.

¿Qué es el SIME?, es un instrumento que reporta el nivel de competencia académica de los alumnos de cada escuela

primaria pública, a lo largo del tiempo y en comparación con otras escuelas similares. Eso es el SIME, a partir del IDANIS.

¿Qué es lo que hace el SIME? Los resultados obtenidos por los alumnos en cada escuela se clasifican en tres categorías: mínimo, competente y avanzado. Las escuelas deben tener como mínimo 10 sustentantes para entrar al modelo.

Debo advertir que estos cortes entre mínimo, competente y avanzado no los definimos nosotros, son aquellos que están dentro de la metodología de lo que es el IDANIS. Nosotros adoptamos estos puntos de corte.

Se hace con alumnos que tienen como mínimo tres años de permanencia en la escuela, con la idea de que la permanencia en el plantel es un reflejo de la práctica docente del mismo. Nosotros suponemos que si un alumno permanece tres ó más años en el plantel, está reflejando la práctica docente de éste, y eso es lo que hemos querido identificar y medir.

Para construir grupos de escuelas comparables, se usó como variable de estandarización la escolaridad de los padres, la cual es una aproximación a los indicadores socioeconómicos concentrados en uno solo, pero es un buen predictor de la situación socioeconómica de los hogares de donde provienen los alumnos.

Por último, el SIME presenta de manera gráfica la distribución de los niveles de rendimiento de los estudiantes de cada escuela y se compara con la distribución promedio del rendimiento de los alumnos de las mejores escuelas dentro de su grupo. A esto le llamamos *la oportunidad de mejora*. También se compara con las 10 mejores escuelas del Distrito Federal, con independencia del tipo de población atendida.

La dirección electrónica del SIME es [www.cpar.sep.gob.mx](http://www.cpar.sep.gob.mx), misma que dentro de dos o tres semanas estará liberada al público en general y podrán acceder a ella quienes deseen hacerlo.

Aquí en la página encontramos las posibilidades de acceder a la escuela, la selección por nombre, dirección, zona, delegación, sector o turno, tenemos también la posibilidad de aislar a las matutinas de las vespertinas o de las continuas; también podemos encontrar las escuelas por orden alfabético, tecleando su nombre, o bien con la clave del centro de trabajo del plantel al que queramos llegar.

Asimismo, aparecen las 16 delegaciones del Distrito Federal con el número de escuelas que hay en cada una de ellas, y su representación porcentual dentro del conjunto de las escuelas del Distrito Federal: Azcapotzalco que representa el 6.25; Benito Juárez que es la más pequeña con el 2.69; Iztapalapa que es la más grande, la quinta parte de la matrícula del Distrito Federal, 20.51, por citar algunas. Por la vía de este catálogo también se puede acceder a las escuelas.

Lo tenemos también por nivel de Direcciones; el servicio operativo del Distrito Federal está organizado en seis Direcciones de Educación Primaria, siendo Iztapalapa una Dirección General en particular y las otras corresponden a la Dirección General de Operación y Servicios Educativos.

Un nivel de desagregación de estas direcciones en términos de su estructura operativa, la encontramos por sector y aquí las delegaciones se despliegan en un número mucho mayor de sectores, hasta completar 52 y todavía los podemos abrir más por zona escolar hasta llegar a las aproximadamente 350 existentes en el Distrito Federal. Este grado de desagregación es muy importante para los supervisores, jefes y todos los directores del plantel, así pueden acotar la zona en que trabajan.

### **Tu escuela y las diez mejores comparables o sobresalientes**

El sistema nos da la oportunidad de encontrar, como lo señala su definición, la



muestra de diferencia entre el desempeño de la escuela seleccionada y las diez mejores escuelas comparables sobresalientes, nos da también la posibilidad de conocer las diez mejores escuelas dentro del grupo con que se puede comparar la escuela en cuestión, y nos muestra su desempeño histórico, eso sería el comportamiento de los resultados a lo largo de varios años de *tu escuela y de las diez mejores comparables o sobresalientes*. Las sobresalientes, recuerden, son las diez mejores según el IDANIS de 2003. Esto se puede hacer de manera global o agregar para cada una de las habilidades que mide el IDANIS: aritmética, geometría, lógica, lectura, etcétera.

Algo más que encontrarán en el sitio Web son las cifras con signos negativos, eso nos dará *la oportunidad de mejora* que hay entre una escuela equis y las escuelas comparables de turno matutino, de turno vespertino, en cada una de las habilidades y en el global, etcétera.

El sistema nos da un gran número de opciones de comparación, podemos encontrar en el sitio Web, que la escuela Cantonalli, en el turno vespertino -hay el prejuicio de que las escuelas vespertinas son de muy bajo desempeño en relación con las matutinas-, es sobresaliente, duplica el número de estudiantes que están en el nivel avanzado de la escuela Belice y tiene un porcentaje de padres con baja escolaridad, pues supera también en 28 puntos a la escuela Belice; es decir, la escuela Cantonalli, aparentemente con una población en gran desventaja, está haciendo algo realmente importante.

La utilidad es justamente, que la escuela Belice puede identificar las diez escuelas que son comparables a ella, puede analizarse junto con éstas. Por supuesto, los números por el momento no ofrecen mayor información, lo único que nos muestran son las brechas existentes, esas son *las oportunidades de mejora*.

Y aquí habría que ir a la segunda fase del estudio o la tercera etapa de nuestra estrategia planteada en un principio, la cual consiste en identificar en las escuelas sobresalientes cuáles son los motivos y circunstancias explicadas con los datos existentes.

Por el momento, solamente se nos ocurren especulaciones, una escuela que tiene resultados tan altos, es porque seguramente tiene un buen director, porque la participación de los padres de familia es intensa, porque la comunidad educativa está cohesionada, no sabemos; de ahí que se requiera de una reflexión e investigación profundas, las cuales nos permitan identificar en estas escuelas de alto desempeño, las estrategias que les está dando resultados sobresalientes. Lo mismo vale para las escuelas que encontremos con muy bajo desempeño, qué están haciendo para encontrarse en una situación de resultados deficientes.

Veamos la tercera posibilidad que nos da el SIME: el desempeño histórico.

Nosotros construimos este sistema para tres años, es decir, el IDANIS 2001, 2002 y 2003 es lo que aparece reportado en nuestro sitio, el IDANIS 2004 se aplicó en el mes de junio, hace aproximadamente un mes, la Dirección General de Evaluación se comprometió a procesar la información del Distrito Federal para finales de julio, de manera que podamos actualizar el SIME, y posiblemente para septiembre u octubre tengamos ya un cuarto año de comparación, lo cual, evidentemente nos va a dar mayores pistas sobre lo que queremos investigar.

Esto es, a grandes rasgos, el SIME. Creemos que es una herramienta muy poderosa, de mucha importancia, para el trabajo de supervisores, jefes de sector, direcciones operativas, la cual nos va a permitir ir escuela por escuela, con información para identificar la práctica do-

cente que se está generando, y tras hacer la investigación extraer conclusiones, posiblemente de aplicación universal para el Distrito Federal o parcialmente aplica-

bles al resto de las escuelas y, sobre esa base, avanzar en la mejora continua que es finalmente el propósito que nos hemos trazado.

---

## Examen de Diagnóstico para el Ingreso a la Educación Secundaria (IDANIS)

Por SERGIO MARTÍNEZ DUNSTAN\*

---

La idea de la presentación va en dos sentidos. Tomaré solamente un par de minutos para explicar lo que la Dirección General de Evaluación viene realizando; primeramente diré que la Dirección tiene sus orígenes en 1972 y desde hace varios años trabaja con las áreas estatales de evaluación.

Estas evaluaciones son de medición de los niveles de logro, y algunos otros aspectos. En ese sentido, dado el nuevo contexto, las características de los programas e instrumentos que la Dirección aplica son de carácter interno, diagnóstico-formativo para retroalimentar el desempeño del propio sistema y la evaluación de programas o proyectos, los cuales impliquen lecciones administrativas, académicas y profesionales respecto a personas o instituciones.

La Dirección General de Evaluación, entonces, realiza evaluaciones de alumnos, docentes, escuelas, programas y proyectos de política educativa. Algunos ejemplos son: el IDANIS, ingresos a normales, Aprovechamiento Escolar en Carrera Magisterial, también Preparación Profesional del propio programa Carrera Magisterial, e ingreso a plazas iniciales.

Para el caso de escuelas existen algunos modelos de autoevaluación de centros, proyectos educativos, evaluación cualita-

tiva del programa Escuelas de Calidad, y el Sistema de Evaluación de la Política Educativa. Éste es el ámbito de competencia de la Dirección General de Evaluación.

La evaluación para nosotros es un proceso ordenado, continuo y sistemático, el cual responde a ciertas exigencias válidas, creíbles, dependientes y fiables; se obtiene a través de ciertas técnicas e instrumentos, que tras ser cotejados o comparados con criterios establecidos permite emitir juicios de valor fundamentados que facilitan la toma de decisiones.

### ¿Qué es el IDANIS?

Hablemos concretamente de lo que es el IDANIS: éste mide el nivel de habilidades básicas para el aprendizaje de los alumnos que ingresan a secundaria y proporciona un diagnóstico personalizado por educando. Hasta ahora, esa es la manera como veníamos trabajando, es una buena idea empezar a desglosar por escuelas.

Esto permite un diagnóstico colectivo de la población asignada a un plantel, y un perfil por entidad, municipio, zona escolar y por escuela primaria de procedencia.

El IDANIS no surge conjuntamente con la Dirección General de Evaluación, quien desde los años setenta aplicaba instrumentos para la medición de la actitud ge-

---

\*Director de Lectura, Calificación y Análisis de Resultados, de la Dirección General de Evaluación de la SEP.



neral de los alumnos de nuevo ingreso a secundaria. El examen de ingreso a secundaria, instrumento sicopedagógico para la distribución y asignación de las habilidades es el antecedente primero.

El IDANIS propiamente se diseñó en 1986 y nació en un contexto muy específico, donde la organización de los contenidos en los programas de estudio de educación primaria y secundaria, respondía a un modelo basado en la estructura de las disciplinas, en la parte del conocimiento era donde se centraba la currícula fundamentalmente. Como sabemos, el modelo enfatiza la adquisición de contenidos; desde esta perspectiva, los contenidos tenían mucho que ver con la parte memorística.

### **Las habilidades como concepto central**

La evaluación del aprendizaje se entendía como el acto de recabar evidencias objetivas, sobre lo aprendido por los alumnos de los contenidos definidos en los planes-programas de estudios. El sustento curricular, la tecnología educativa y a su vez, el conductismo, son el contexto pedagógico educativo en que se desenvuelve el IDANIS; además se concibió de manera que la habilidad fuera el concepto central, las habilidades las entendemos como pautas de acción de las cuales se vale el sujeto para enfrentar la resolución de problemas, por lo cual, se expresan como la capacidad para comprender y solucionar una situación.

Las habilidades son elementos subordinados a estrategias cognitivas, elementos que hacen ejecutable al IDANIS. Tiene algunas variables, concretamente habilidad verbal, matemática, razonamiento abstracto que, a su vez se subdividen en algunas otras unidades de diagnóstico; para el caso de la habilidad verbal, comprensión de lectura, complementación de oraciones; para el caso de matemática, aritmética y geometría; y para el razonamiento

abstracto, se construye una serie de figuras con el objeto de medir el nivel de desarrollo y las salidas de aprendizaje.

Un poco más a detalle, podemos concebir la habilidad verbal como el conjunto de recursos vinculados al conocimiento lingüístico, los cuales permiten al alumno procesar información para acceder a significados explícitos e implícitos, superficiales y profundos de textos diversos.

La habilidad matemática se entiende como el conjunto de recursos vinculados también al conocimiento, los cuales permiten al alumno procesar información aritmética y geométrica, para resolver problemas en el ámbito representacional; y la habilidad para el razonamiento abstracto se define como: aquellos recursos que permiten acceder al conocimiento de la representación figurativa bidimensional y procesar información gráfica para la interpretación de representaciones visuales diversas.

Cada una de estas habilidades tiene sus dimensiones, y éstas subdimensiones; para el caso de la habilidad verbal sería la comprensión lectora, compuesta de doce ítems y señala entre otros aspectos: reconocer o construir el significado de vocablos; suprimir información trivial, redundante o ajena al texto; globalizar la información; reconocer la estructura exclusiva; reconstruir ideas de un texto; y realizar inferencias a partir de información explícita.

Otra de las subdimensiones de la habilidad verbal es la complementación de oraciones, la cual también tiene 12 ítems; es la reintegración de un texto utilizando el reconocimiento de la secuencia temporal, identificando la relación lógica, la identificación de relaciones atributivas y reconocimiento de la concordancia gramatical; estas dos subdimensiones corresponden a la habilidad verbal.

Para la habilidad matemática son dos subdimensiones: aritmética, que también

se compone de 12 ítems o preguntas, y pretende medir que el alumno reconozca y relacione los elementos de un problema, el cual implique la traducción de situaciones verbales a su expresión numérica, así como el análisis de datos numéricos y la relación de cálculos básicos.

Para el caso de geometría, otra de las subdimensiones de la habilidad matemática, es la interpretación de relaciones espaciales, figuras, cuerpos representados en dos dimensiones, auxiliado por la expresión aritmética de las mismas y relaciones de igualdad, desigualdad o proporción.

Para el caso del razonamiento formal, podríamos hablar de serie de figuras, también son 12 preguntas; aquí se plantean algunos problemas al alumno referentes a la solución de casos, cuya expresión no siendo verbal y matemática, implica la detección de reglas de un proceso que se tiene en el tiempo y espacio.

En total, la prueba está conformada de sesenta preguntas con diferente grado de dificultad, las preguntas de la prueba son de opción múltiple, de tal modo que el instrumento se clasifica como prueba objetiva y estandarizada.

La lectura óptica, materiales y calificación, se hace en porcentaje de aciertos. La metodología es diseñada en algunos casos por la Dirección General de Evaluación; en otros casos por la entidad. Para efectos de equiparabilidad utilizamos la medida *Rach*.

Aquí hay tres conceptos fundamentales en los que no quiero ahondar ahora, dado que va a ser motivo de la mesa del día de mañana, y es la parte correspondiente a comparabilidad, calibración, y equiparación. Igualmente es necesario ir mejorando el IDANIS, adecuándolo, sobre todo en el nuevo contexto que he señalado, considerando también el modelo de calidad de las evaluaciones planteado por el Instituto

Nacional para la Evaluación de la Educación. Ésta es una acción fundamental.

### **Niveles de logro**

En este sentido, de acuerdo con la calificación que obtiene el alumno, lo ubicamos en alguno de los cuatro niveles de logro. Si lo vemos con el procedimiento clásico que es el porcentaje de aciertos que solamente es para las entidades participantes, esa sería la escala. De acuerdo con lo que tienen los alumnos, se les ubica o se les clasifica en este nivel de logro.

Los niveles de logro significan algo, este significado es el siguiente: primeramente, el nivel A significa que el alumno carece de estrategias para abordar tareas que implican el uso de habilidades medidas, o bien, aplica estrategias inadecuadas, de tal modo que las habilidades incluidas en su plan no son las pertinentes.

A causa de estos problemas tiene limitaciones serias para apropiarse del nuevo aprendizaje y consolidar el que aparentemente posee. En este sentido, es imprescindible desarrollar estrategias para el aprendizaje.

Ese es el nivel más bajo, por consiguiente, en el nivel más alto el alumno aplica las estrategias adecuadas para abordar las tareas, muestra un manejo que tiende a ser sobresaliente de las habilidades medidas, pero esto no significa que sus estrategias hayan llegado a un punto tras el cual no existan posibilidades de desarrollo.

Por lo tanto, es recomendable que realice actividades orientadas a la depuración en sus estrategias, a fin de que su aprendizaje sea más sólido y consistente.

Como decíamos, se proporciona un diagnóstico personalizado por alumno, se le clasifica, y además se plantean recomendaciones al propio profesor o al alumno de cómo superarse o qué ejercicios necesitaría realizar para pasar de un nivel de otro.



## ¿Cómo inicia el IDANIS?

En 1995 empezó en el Distrito Federal y Tabasco, aunque sabemos que desde 1986 se viene aplicando. Éste ha venido evolucionando de tal manera que en 2003 fueron 675 mil 920 alumnos medidos con el IDANIS en alrededor de 20 entidades, hay algunas que tienen ya tres, cuatro o más aplicaciones.

Veremos algunos de los resultados, como las medias que han obtenido las entidades en los promedios por año de evaluación. Ahora no voy a entrar al análisis, pero esto da pauta, insisto, para la mesa del día de mañana, porque es necesario revisar esos detalles, sobre todo para las series históricas y la parte de la comparabilidad.

Es necesario enfatizar que solamente en el Distrito Federal se aplica un cuestionario de contexto, en las demás entidades el IDANIS no va acompañado de cuestionario de contexto.

Si hay algún interés de alguna entidad por la metodología del SIME, debería de considerar también el contexto y conocer cómo el indicador se va construyendo, por ejemplo, los atributos socioculturales de la madre.

Esto es a grandes rasgos el IDANIS. Hay un punto fundamental en el IDANIS a discutir, esperemos que el día de mañana se pueda ahondar más en el procesamiento, calificación, calibración, equiparación y comparabilidad en las series de tiempo, en las habilidades, y por entidades.

---

## PARTICIPACIÓN DE LOS COMENTARISTAS

RAFAEL VIDAL URIBE\*

Siendo miembro del Instituto, he sido de los impulsores del SIME, y lo que quisiera hacer son algunas observaciones y advertencias, para cuando estamos penetrando en lo que es este Sistema de Información para la Mejora Escolar.

Es importante mencionar que este proyecto empezó hace más de un año; efectivamente, nosotros tenemos mucho interés en los resultados escuela por escuela; no creemos sea una labor que el INEE deba de hacer, pero sí consideramos importante de desarrollar.

El Instituto, con su capacidad tecnológica, puede ayudar para que esto ocurra, y desde hace un año impulsamos el proyecto; platicamos con la Subsecretaría de Servicios Educativos para el Distrito Federal

para llevarlo adelante, en principio como experiencia piloto, para que en el futuro se pueda desarrollar, también en algunas otras entidades federativas.

La expectativa del sistema es que vaya más allá de un *ranking* simple, en el que se evaluara a los alumnos con alguna prueba estandarizada, y luego sacáramos los promedios de rendimiento y ordenáramos a las escuelas.

Quisimos evitarlo eso e impulsamos el proyecto del SIME, en el que efectivamente se compara a las escuelas, pero con escuelas, como se dice ahí, comparables, con las que están en la misma categoría. Esto es fundamental, porque el propósito es que las escuelas identifiquen los aspectos a mejorar.

---

\*Director General Adjunto del Instituto Nacional para la Evaluación de la Educación.

El SIME debe pasar a una segunda etapa, es lo más importante, la cual se conoce como *las mejores prácticas*; esto quiere decir que los directores de las escuelas platiquen con los directores de las diez mejores escuelas comparables; lo que se pretende es que el SIME sirva para provocar el diálogo entre autoridades, para ver qué estrategias se siguen para tener éxito en las escuelas que son comparables.

### **El cuestionario de contexto**

El IDANIS es un instrumento que no fue pensado para el SIME, sino para ubicar a los niños que pasan del sexto al primero de secundaria; para tales propósitos el instrumento funciona adecuadamente.

Gracias a la tenacidad de la Subsecretaría de Servicios Educativos del Distrito Federal, de aplicar esto consistentemente, cuentan con resultados de muchos años. Y dado que se ha aplicado el cuestionario de contexto, –en otras entidades se aplica IDANIS y no se aplica cuestionario de contexto– se pudieron realizar los estu-

dios para ver cuál variable era la que hacía operativa esta aplicación. Gracias a eso tenemos un conjunto de datos estudiados y se encontró que la variable con la que era posible analizar la información y observar las diferencias, era justamente *escolaridad de los padres*.

Esta variable resultó pertinente para hacer el análisis. Es una advertencia importante porque cualquiera podría decir eventualmente, que otra variable de contexto podría ser mejor, pero no la tenemos contemplada. Un asunto clave es que, para tener un sistema como el que hoy se presenta, con todas estas advertencias, se requiere desde luego del censo.

Sería pertinente mencionar que el sistema está empezando y por tanto se debe tener mucho cuidado, con los resultados encontrados. No vamos a encontrar cambios espectaculares, cuando estemos hablando de la educación del país que tiene una inercia enorme y no vamos a observar las mejoras de la noche a la mañana. Debemos considerar tales aspectos con muchísimo sigilo.



**BEATRIZ CALVO PONTÓN\***

**M**e ha interesado conocer de manera directa, permaneciendo en escuelas y en las aulas largas temporadas de tiempo, cómo funcionan las escuelas y las aulas, qué sucede a su interior, es decir, el énfasis que pongo en la investigación que realizo, es en los procesos, saber cómo se van construyendo éstos de manera cotidiana.

En este sentido, me interesa saber cómo se producen las formas de gestión escolar

al interior de una escuela, cómo se toman las decisiones, cómo enseñan los maestros, qué aprenden los alumnos, cómo evalúan o califican los maestros, etcétera.

Desde esta óptica haré mis comentarios sobre el SIME, con la idea de reflexionar sobre el alcance y la utilidad de esta forma de conocer las escuelas.

En primer lugar deseo comentar sobre la perspectiva del SIME; se trata de un sistema de información de resultados por es-

\*Académica del Centro de Investigaciones y Estudios Superiores en Antropología Social y miembro del Consejo Técnico del INEE.



cuela y a mi juicio esto implica algo sumamente importante, pues al mismo tiempo se abordan dos dimensiones que no siempre son fáciles de manejar. Por una parte, el ofrecimiento de un panorama macro al abarcar todas las escuelas públicas del Distrito Federal.

Por otro lado, es mirar a la escuela como el centro de atención, es decir, desde una dimensión micro o local, se estudia el impacto en la calidad del aprendizaje en escuelas concretas. Se trata de un instrumento muy importante para coadyuvar al proceso de toma de decisión al interior de las escuelas, y es quizás una de las cuestiones que más pondero en esta forma de hacer diagnósticos de escuelas, de ahí se desprenden varias cuestiones teórico metodológicas que me gustaría comentar.

### **Generalizar desde lo particular**

El estudio generaliza desde lo particular, es decir, el punto de referencia para poder generalizar es lo particular, en este caso la escuela, lugar donde suceden los hechos educativos, es a partir de lo particular, inmediato y local, que se construyen patrones generales presentes en distintas situaciones, los cuales se manifiestan de forma concreta en cada una de ellas, de ahí la importancia de centrarse en la escuela. Esta forma de acercarnos considera a la escuela como punto de partida, es decir, con una visión desde dentro y como punto de llegada hacia dentro.

Por otra parte, se comentaba la contextualización, el entorno socioeconómico de las escuelas, que es un elemento importante a considerar en la evaluación de éstas, y significa que se entienden las escuelas no como entidades cerradas en sí mismas, circunscritas y aisladas, sino articuladas a realidades y problemáticas más amplias.

Otro punto que el estudio muestra es el movimiento de las escuelas; no ofrece una

visión estática al presentar información a lo largo del tiempo, los resultados se refieren a tres generaciones de alumnos de sexto grado: 2001, 2002 y 2003; se trata de un panorama de los últimos tres años, escuela por escuela, a través del cual se pretende mostrar cómo se comporta la población de estudiantes de un año a otro, con respecto al aprendizaje. Finalmente se trata de comparar el nivel de aprendizaje alcanzado por los alumnos de escuelas similares en el Distrito Federal, y esto es una virtud metodológica muy importante.

### **La utilidad del SIME**

Otros comentarios que quisiera hacer son con relación a la utilidad del SIME. En primer lugar, nos permite consultar información de las escuelas públicas del Distrito Federal a nivel de la totalidad y desglosadas por delegación, sector y zona escolar, inclusive por turno matutino y vespertino considerando, al mismo tiempo, la escolaridad de la madre como parte de este contexto, presente como indicador socioeconómico importante.

De esta manera, el SIME permite una diversidad de análisis al relacionar una serie de elementos importantes; los cuales podrían ser un apoyo interesante para directores de escuela y supervisores de zonas escolares de educación básica, en el trabajo de diagnóstico de las mismas, debido a la escasa utilidad de los diagnósticos que se acostumbran realizar.

El SIME representa una herramienta para que directores y supervisores conozcan más y mejor a sus escuelas. El acceso a la información desagregada por delegación, sector y zona escolar, podrá ayudarles a elaborar diagnósticos que verdaderamente coadyuven a mejorar la calidad en la educación de los alumnos de las escuelas que pertenecen a sus zonas escolares.

En este sentido, es una forma de acercarse, de mantener a los supervisores cerca de las escuelas. Pueden identificar el nivel de aprendizaje de los alumnos de cada una de las escuelas en su zona escolar; el grado de avance de sus escuelas durante los tres últimos años; y comparar las escuelas al interior de su zona y con respecto a escuelas de otra zona escolar, considerando indicadores del entorno.

Esta comparación permite a los supervisores detectar en su misma zona escuelas con excelentes o buenos resultados, con resultados regulares, bajos o muy bajos, y de ahí deberían surgir nuevas preguntas, por ejemplo: ¿por qué hay escuelas buenas y escuelas deficientes en la misma zona escolar? ¿Qué determina dentro de la misma zona escolar, que una escuela tenga un buen rendimiento mientras otra no? Con base en estas preguntas, los supervisores podrán seleccionar las escuelas que convenga sean estudiadas a profundidad.

El SIME por sí solo no puede responder a estas preguntas, pues tienen que ver más con información que dé cuenta de procesos al interior de las escuelas, relativa al funcionamiento cotidiano de éstas, así como las culturas escolares que se construyen por los propios sujetos educativos.

Estos procesos se vinculan con las instalaciones, las condiciones materiales de trabajo, el ambiente, las formas de operación y de gestión, el uso del tiempo académico, las condiciones pedagógicas relacionadas con procesos de enseñanza, aprendizaje, planes de estudio, currículo, contenidos, formación, actualización y superación docente, los procesos en las aulas, las prácticas de los maestros en el aula, las formas de relacionarse y de trabajo colectivo e individual de los docentes, el papel del director, la participación de los padres y madres de familia, así como de ciertos grupos de la comunidad, etcétera.

Es decir, tiene que ver con la caja negra de la escuela o con lo que se conoce como *el lado oscuro e invisible de la misma*. El SIME no permite detectar tales cuestiones, pero es una invitación a que se haga este tipo de acercamientos para conocer exactamente ¿cómo funcionan las escuelas, por qué se dan unos resultados y no se dan otros?

### **El impacto del proyecto**

Las posibilidades del impacto de los resultados de este tipo de estudios, en primer lugar se dan en la detección y jerarquización de los problemas educativos más apremiantes en cada zona escolar, en cada escuela, en la realización de ejercicios de contrastación entre zonas escolares y entre escuelas que implican comparar diferentes situaciones.

En segunda instancia, en la toma de decisiones a nivel de autoridades educativas, jefes de sector, supervisiones escolares, direcciones de escuelas y docentes en sus aulas, con objeto de definir acciones y planificar la educación, ofreciendo alternativas de solución reales y factibles a los problemas en los diferentes niveles, considerando las características compartidas, así como las específicas y diferenciadas.

Otra forma de impactar es con la invitación para que otros estados reproduzcan este tipo de estudios, ajustándolos obviamente a sus particularidades, ello tendría al menos dos grandes beneficios: cada estado conocería sus escuelas y, un avance en estudios cooperativos a nivel nacional. Claro, sería un trabajo que no necesariamente haría el INEE, pero es un gran acierto del Instituto apoyar e impulsar iniciativas en todo el país.

Como miembro del Consejo Técnico del INEE, me interesa enfatizar este punto. Finalmente, quisiera expresar una preocupación. Habrá que buscar la manera para



que el acceso de cualquier persona a la información del SIME o a estudios similares que impliquen tener acceso a todas las escuelas, no lleve el riesgo de hacer mal uso de los datos, afectando negativamente y de manera directa a las escuelas.

Sabemos que estas situaciones son reales, se puede utilizar la información para evidenciar escuelas, directores, maestros, pero esto no es parte de la política ni de la ética que una institución como el INEE debe seguir.

---

## MAURICIO FORTES BESPROVANI\*

---

**N**o se puede menos que celebrar poner a la disposición de la sociedad mexicana este nuevo elemento de información, el cual se deriva de un análisis inteligente con relación a cómo está evolucionando la educación básica, en este caso particular, las escuelas primarias del Distrito Federal.

En una primera etapa, cuando surgió la ya muy mentada cultura de la evaluación, se empezaron a introducir elementos cualitativos para tratar de entender un sistema tan complejo como es la educación, y también para percibir la situación de la educación básica en México.

En la medida que hayamos logrado cuantificar diversos indicadores, o índices y parámetros sobre este fenómeno tan complicado como es la educación, podremos determinar con mayor precisión dónde actuar, ya sea para corregir, avanzar o mejorar, de hecho es el lema de varios de nuestros centros de trabajo, y cómo poner en el plano adecuado los cambios que se necesitan.

El SIME representa un nuevo y poderoso sensor o termómetro; en la mañana se hablaba del Ceneval o el INEE, como laboratorios clínicos, en el sentido de que toman medidas, tienen ciertos referentes, pero el diagnóstico no lo hace ni el INEE

ni las agencias de evaluación, sino los tomadores de decisiones, pero sí es bueno tener un termómetro que constantemente esté tomando esas mediciones.

### **Todo es perfectible**

En este caso, tenemos un importante instrumento para las escuelas del Distrito Federal, no es perfecto, pero ha tenido sobre todo el cuidado de estar basado en un principio de justicia, al establecer comparaciones para la mejora en clases o categorías definidas por una variable, en ese caso es la escolaridad de los padres de familia. Es desafortunado, pero desafortunado en el sentido de que el alumno no puede remediar esa situación.

Hay otras variables, por ejemplo los hábitos de lectura, los cuales sí puede mejorar el alumno y de esa forma él se vuelve un diseñador de su propio futuro, pero aquellos que no dependen de él, lo menos que puede hacer tanto la SEP en el Distrito Federal o el INEE, es tomar en cuenta estos elementos de justicia. La escolaridad de los padres ciertamente es una variable muy importante y qué bueno que se toma en cuenta.

El segundo punto tiene que ver con los propósitos de la evaluación, pero más bien

---

\*Director General Adjunto de Investigación y Normas del Centro Nacional de Evaluación para la Educación Superior (Ceneval).

sobre los propósitos de la información. ¿A quién se informa? A las escuelas mismas, a los supervisores y a quienes van a tomar una decisión. Pero una segunda parte de la pregunta es, a quién está dirigida esta información. Y cada respuesta deberá asociarse con una manera diferente de analizar la información.

Podemos hacer un simple ordenamiento por categorías, como ya se dijo, de las escuelas, pero también se puede hacer un análisis más elaborado, con elementos estadísticos que permitan identificar aspectos fuertes y débiles en cada plantel, turno o grupo.

Esto se puede hacer y publicar, pero si el propósito es hacer de la información un poderoso instrumento de mejora, conviene identificar el sector al que se dirigen los resultados de nuestras evaluaciones, para lograr que sirvan de estímulo para la superación.

Es algo difícil, a veces la prensa no nos entiende bien cuando decimos *las informaciones que podemos producir deben tener diferentes destinatarios, no porque se quieran ocultar, sino por el uso que se le pueda dar*. Dicho de otra manera, sería deseable organizar la información de diferentes formas, según el sector al cual la queremos enviar. Ahora, en la era de Internet, esto parece ser un planteamiento muy realista.

Si pensamos, por ejemplo, en padres de familia, para ellos ciertamente es importante conocer la calidad y el nivel de servicios educativos que reciben sus hijos, pero en ocasiones podría ser igual de importante tener una idea de cuál es el programa de trabajo de cada maestro, qué tipo de actividades tiene planeadas a lo largo del ciclo escolar y, muy importante, cómo los padres pueden facilitar el trabajo escolar de sus hijos y apoyar al maestro, es decir, fomentar la participación de los padres de familia en la escuela pero con mayor información; si el destinatario son

los alumnos para ellos sería interesante conocer, además de lo que se espera de su trabajo durante el año escolar, la serie de actividades reforzadoras, muchas de ellas extraescolares como deportes, cultura o concursos.

Otro grupo es el de los supervisores a quienes puede ser muy útil el SIME, o los administradores o directivos de los planteles, creo que este sistema es particularmente adecuado para este sector: permite a los directores ubicar su escuela y los mecanismos que el propio director tiene de gestión en relación con sus colegas pares, pares en el sentido de escuelas comparables en la misma categoría.

Para los inspectores será útil identificar las buenas prácticas que se llevan a cabo en las escuelas de mayor rango en el ordenamiento, y llevar esas experiencias a los demás planteles como ya se mencionó aquí.

Para los tomadores de decisión de políticas públicas, al identificar los planteles en mayor desventaja, será posible coordinar y dirigir acciones tendientes a remediar problemas, subrayo dirigir no una acción general, sino una vez identificado el origen de la desventaja, es más eficaz dirigir esfuerzos concretos.

También resulta válido identificar problemáticas comunes derivadas de planes de estudio, formación de profesores, contenidos, infraestructura de plantel, problemas en la gestión de la dirección y otros más.

### **Mejora para la planeación educativa**

Tenemos la cuestión de los estándares como tercer punto, la sigla M del SIME se refiere a la mejora, es Sistema de Información para la Mejora Educativa y esto requiere de un referente, que sea un referente de consenso, el cual obedece también a la planeación educativa, que en este caso



es la del Distrito Federal o puede ser la del país si es que somos ambiciosos.

El sistema actual establece una cuantificación de la brecha de mejora con respecto a las diez mejores escuelas de su categoría, pero yo creo que podría ir más lejos si también se pudieran cuantificar las brechas de oportunidad de mejora con respecto a criterios estandarizados de rendimiento, así como la situación con respecto a regiones geográficas más grandes, es decir, pensar en el SIME, no solamente en el Distrito Federal, sino en los municipios, entidades y, por qué no, en todo el país.

¿Qué pasa cuando comparamos a la mejor escuela o a las mejores?, su oportunidad de mejora va a ser al revés, porque va el universo de comparación y las que quedan son las que están debajo de ella, entonces va a dar un resultado paradójico. Entonces, quizás conviene revisar también los estándares y poder mejorar el sistema para tener diferentes referentes, incluyendo los internacionales, para comparar a las mejores escuelas en otra categoría, entre las mejores de las mejores.

Queremos pensar en un sistema de información que no sea estático, es decir, que no se limite a una región, aquí es natural porque esto lo produjo la Subsecretaría de Servicios Educativos del Distrito Federal, pero afortunadamente ya hay un análisis longitudinal que es extremadamente valioso, sería interesante recurrir a otros instrumentos de evaluación como podrían ser los estándares, siendo consciente de que los estándares son una prueba muestral y no censal y coincido con lo que se dijo también en la mañana, que no es legítimo hacer ningún juicio, de hecho es ilegal, sobre una prueba censal.

Pero sí es posible tratar de definir una correlación de validez estadística para tratar de ir más lejos, tener más información, ampliar el espectro de oportunidades de mejora y ubicar con toda precisión las es-

trategias de aprendizaje, en comparación con otros estados y países.

### Usando otros instrumentos

Finalmente, me gustaría mencionar cómo tratar de usar otras herramientas, no solamente el IDANIS, esto se tiene que hacer con algún rigor estadístico para poder comparar un tipo de instrumento con otro y quiero mostrar que esto sí se puede hacer y puede llevarnos a situaciones interesantes.

El EXANI I que aplica el Ceneval a otro nivel, al final de la secundaria, es el examen de ingreso a la preparatoria, pero tenemos resultados de que tiene una correlación muy alta, una correlación estadística muy alta con los IDANIS; eso nos permite hacer un estudio, ya se mencionó el que estamos haciendo sobre el valor agregado que puede dar la secundaria, tomando como una medición de inicio los IDANIS y una medición de fin al EXANI I; pero quiero mencionar lo que sucede cuando uno no toma ordenamientos brutos de las escuelas tal y como son.

Tenemos dos tipos de ordenamiento, el bruto, que es muy injusto, son seis niveles que van desde muy bueno, hasta muy malo; pero también el neto que introduce el principio de justicia o de equidad, es decir, no podemos comparar escuelas en las cuales los padres de familia tienen escolaridades muy diferentes; de hecho sabemos que un padre que terminó la secundaria, en promedio logra que sus hijos tengan un nueve por ciento más de aciertos que una madre que no terminó la primaria, o sea, sí es muy importante esa variable y por eso celebramos que el SIME la haya tomado en cuenta.

La columna de netos es cuando se han quitado esas variables, la variable *género* también es importante, sí hay diferencias entre los hombres y las mujeres. Entonces, si quitamos todas esas variables que actúan en desventaja de los estudiantes,

tenemos una visión de la escuela mucho más objetiva, porque lo que estamos realmente midiendo es: qué es lo que le dio la escuela al niño: ¿qué oportunidades de aprendizaje proporcionó?

Y si uno hace ese tipo de ordenamientos de las primeras dos escuelas, la que sale en primer lugar es una escuela que en realidad es muy buena porque selecciona muy bien a sus estudiantes, es una escuela privada, tiene excelentes estudiantes porque tiene un sistema muy riguroso de selección, pero si quitamos ese factor, ya en

el rendimiento neto la escuela que pasa a la posición uno es una escuela que resulta ser pública. Esto es lo que está haciendo el SIME y creo que se puede mejorar mucho.

Creemos que es una experiencia muy buena, le metemos muchos números, pero creemos que los números no solamente dan una visión muy buena de la realidad, al mismo tiempo nos quita muchos prejuicios y nos dan una idea mucho más auténtica de lo que la escuela está haciendo, en cuanto a las oportunidades de aprendizaje en los alumnos.



## SYLVIA SCHMELKES DEL VALLE\*

**D**eseo advertir que en términos generales estoy en contra de la manera como históricamente en el mundo, se han venido aplicando estas devoluciones, escuela por escuela, de los ejercicios de evaluación, y lo quiero decir porque es el punto de partida de lo que expresaré a continuación.

Voy a explicar por qué estoy en contra de la manera como tradicionalmente se han hecho estos ejercicios; voy a señalar cómo siento que el SIME supera algunas de estas cuestiones que hacen sospechar de cualquier intento de devolución escuela por escuela, y voy a señalar finalmente algunas sugerencias acerca de cómo poder mejorar estos aspectos que el SIME toma en cuenta y son aciertos que permiten superar situaciones tradicionalmente presentes.

### **Comparar lo comparable**

La primera razón por la que estoy en contra de esta manera tradicional de de-

volver información escuela por escuela, es porque me parece que es un ejercicio de suyo injusto por dos razones.

La primera, porque lo que hacen estos ejercicios tradicionales es comparar lo incomparable, se trata de comparar escuelas que no son iguales por la población a la que sirven, de manera muy importante por la educabilidad de los hogares de donde vienen los alumnos, por el contexto en el que operan, incluyendo el nivel socioeconómico de los alumnos, y por los recursos con los que cuentan las escuelas: éstas no necesariamente cuentan con los mismos recursos para poder enfrentar el trabajo que tienen asignado o comisionado.

Voy a poner un ejemplo, que no es el caso aquí, pero supongamos que se comparan escuelas privadas y públicas. Además de las diferencias socioeconómicas, que son bien conocidas, las privadas además tienen la característica de que seleccionan a sus alumnos y de que expulsan a aquellos que no les conviene tener.

\*Coordinadora General de Educación Intercultural Bilingüe en la Secretaría de Educación Pública, académica del Departamento de Investigaciones Educativas del Cinvestav, y miembro del Consejo Técnico del INEE.



Entonces, las públicas, al menos en teoría no pueden hacer eso, digo al menos en teoría, porque a veces estos ejercicios, como ya decía Pedro Ravela en la mañana, provocan que también las escuelas públicas seleccionen, a lo mejor no expulsan, pero cuando menos sí seleccionan.

Entonces ocurren los hechos que acaba de mostrar Mauricio Fortes, de cómo cuando uno realmente compara, tratando de buscar comparabilidad, los rangos cambian de manera muy notable.

A mí me parece que el SIME efectivamente trata de comparar las escuelas que son comparables entre sí, para empezar solamente compara las públicas, no tiene una base de datos de escuelas privadas, compara con las diez mejores del estrato en cuestión y en ese sentido el estrato es el que se considera comparable, medido por la escolaridad de la madre, que yo también, en lo personal considero que es una variable muy potente y en sí misma un muy buen *proxy* de otras variables importantes que sabemos afectan el rendimiento educativo.

### **Preocupación por la *oportunidad de mejora***

Al igual que Mauricio Fortes, estoy también preocupada por este uso del término *oportunidad de mejora*, porque efectivamente aquella escuela que está en el lugar diez, bueno, el primer lugar de las diez que se están comparando, pues ya lo único que puede tener es oportunidad de empeorar, ya no oportunidad de mejorar. Entonces, sí creo que sea necesario cambiar este término, creo que es muy importante.

Y, por otro lado, también a mí me preocupa el uso del término de *oportunidad de mejora*, por el mensaje que puede estar enviando el sistema, en el sentido de que debe esperarse menos de las escuelas más pobres o con características más pobres.

Creo que no es cierto, la investigación educativa nos muestra muchísimos casos de escuelas que trabajando en situaciones difíciles, de pobreza o de minoría cultural, los resultados que obtienen los alumnos son incluso mejores que los de las mejores escuelas de zonas ricas.

Entonces, la *oportunidad de mejora* es mucho más que la señalada por el sistema y tal vez habría que indicarlo, quizá por niveles de oportunidad de mejora, pero todas las escuelas pueden mejorar, incluso las que ya están en el nivel más alto de su estrato; pero de todos modos el acierto es que compara lo comparable. Creo que éste es el mayor aporte del sistema.

Considero que, de alguna manera, estamos suponiendo que los recursos con los que cuentan las escuelas para hacer esto son los mismos, porque la variable que estamos tomando para homogeneizar estratos es la escolaridad de la madre, no estamos tomando ninguna variable que dependa del sistema para desigualar escuelas, o sea, con variables de inequidad introducidas por el propio sistema.

Entonces, me parecería muy importante, una de dos, o bien asegurar que las escuelas efectivamente tienen los mismos recursos, que a lo mejor es algo que ya está asegurado, respecto de los materiales, no necesariamente respecto de los recursos humanos, o bien, también dar cuenta de estas diferencias en los recursos que dependen del propio sistema.

La última recomendación al respecto es que el sistema a lo mejor podría hacer algo para llevar al usuario, mucho más a comparar lo comparable, que a comparar con un ranking simple de todo el Distrito Federal.

También considero que la manera como se han venido haciendo estas devoluciones es injusta, porque no comparan en el tiempo. Siempre he dicho que la definición de una buena escuela es aquella

que es capaz de mejorar respecto de ella misma, como dijo Ana María Aceves esta mañana, la mejor escuela es aquella que muestra esta capacidad de tener una pendiente alta respecto de ella misma en el tiempo.

Y cuando una devolución escuela por escuela no refleja eso, es injusto, porque aquellos maestros que parten de una escuela que está en muy mal estado, y que ponen todo el esfuerzo por mejorarla, pero en una determinada evaluación se expone ese mal estado, porque por más que ha mejorado, de todos modos está baja, lo único que puede provocar esta devolución de información, y lo ha hecho, es frustración del cuerpo docente con su director, y eso genera una espiral hacia abajo en la misma calidad educativa, porque es un esfuerzo muy grande que no está siendo reconocido.

Entonces, es injusto cuando no se compara en el tiempo y el SIME tiene justamente el acierto de comparar en el tiempo y presenta hasta la fecha la serie histórica de los últimos tres años.

De repente uno se queda con la duda de si efectivamente son comparables las pruebas año con año, porque encontramos unos saltos muy grandes de algunas áreas que uno no se puede explicar.

Entonces, la condición para poder comparar es tener la certeza de que el instrumento efectivamente es estable en el tiempo y nos permite comparar año con año.

Si esto fuera así, si pudiéramos realmente asegurar que el instrumento es estable, yo sugeriría que tomáramos otro criterio de ordenamiento en las escuelas: la pendiente, el grado de pendiente de la curva de mejora a lo largo del tiempo, no nada más como algo respecto de lo cual comparamos con otras escuelas y otras pendientes, sino como una forma de ordenar las escuelas en el sentido de que la mejor

escuela sería no aquella que tiene mayor porcentaje de aciertos o mayor puntaje en la escala, sino aquella que ha aumentado más respecto de sí misma; y entonces nos pudiéramos comparar respecto de esta pendiente, desde luego también reconociendo que entre más abajo se está, más fuerte puede ser la pendiente; entre más arriba se está, la pendiente ya empieza a nivelarse.

De todos modos, partir de que la escuela buena es aquella capaz de mejorar de manera consistente respecto de ella misma, debe ser considerado, sobre todo tomando en cuenta la injusticia de hacer lo contrario.

### **Efectos de difundir la información**

Por tanto, la primera razón por la que estoy en contra de esa forma tradicional de comparación es que es injusta. La segunda, es que está basada en supuestos que no siempre se cumplen; un supuesto muy claro respecto de esta posibilidad de devolver escuela por escuela, es que las escuelas tienen grados de autonomía y pueden tomar decisiones importantes.

Si bien en México hemos avanzado mucho en este sentido, pues estamos pidiendo a las escuelas lleven a cabo su propio diagnóstico, realicen su proyecto escolar, añadiendo a los márgenes de maniobra que tiene un director de escuela; todavía las posibilidades digamos normadas de tomar decisiones son sumamente estrechas en el caso mexicano. Todavía el sistema educativo mexicano está montado sobre la base de emitir un conjunto de instrucciones, para que los operadores las sigan, no para que quienes están en la escuela tomen decisiones.

Insisto, estamos avanzando muchísimo en este asunto de la autonomía de las escuelas, convencidos de que las escuelas pueden hacer la diferencia, pero nos que-



da mucho de lo anterior, todavía hay sanción sobre todo de parte de supervisores escolares, a aquello que significa tomar decisiones que se salgan de lo que está muy claramente instruido. Esto implica reforzar la autonomía, y mi recomendación es seguir en este camino que México inició, y el Distrito Federal de manera muy importante.

El IDANIS es una prueba que como ya se dijo, fue diseñada con un propósito diferente al de evaluar la calidad de la primaria, fue diseñada con el propósito de distribuir a los alumnos en las secundarias.

Dado que la libertad de elección no se cumple, es muy importante tomar en cuenta que en el momento en que la información de IDANIS se haga pública repercutirá en una mayor exigencia de cierto sector de la sociedad, por tener libertad de elección de escuela y no sé si estamos lo suficientemente preparados para enfrentar ese efecto, porque exigirá el cambio de escuela cuando encuentre que su hijo está en una escuela no muy buena; entonces, aunque la escuela que se juzga es la primaria no la secundaria, de todas maneras, habrá ahí un intento por tener información acerca de las mejores secundarias, para mandar a sus hijos o de las primarias, para enviar a los hermanos. No sé si estamos como país preparados para enfrentar la consecuencia que tendrá hacer pública esta información.

### **Exigencia de los padres de familia por una mayor calidad**

También estoy en contra por una tercera razón, que de alguna manera ya mencioné, aquello que dijo Pedro Ravela en la mañana: este efecto de chicle de goma de mascar que él mencionaba, pues devolver información escuela por escuela de la manera como tradicionalmente se ha hecho polariza, porque los mejores alumnos,

los que pueden elegir, cuyos padres están preocupados, son aquellos que se van a las mejores escuelas y lo contrario sucede con las peores escuelas, el efecto polarizante sabemos es un efecto desmoralizador, frustrante para los cuerpos docentes y finalmente esto es un efecto necesario de prevenir.

Cuando se hacen públicas las informaciones de devolución de escuela por escuela, tenemos que suponer una de dos cosas, o se puede elegir la escuela a la que se va a mandar a los niños, o bien los padres de familia tendrán más elementos para exigir mayor calidad a las escuelas donde están sus hijos.

Esto segundo es muy importante, y tiene que ver con el derecho de los padres de familia de estar informados respecto a la calidad de las escuelas, pero me preguntaría si efectivamente estamos preparados para responder a la creciente exigencia de los padres, que seguramente será consecuencia de hacer pública la información. Cómo nos estamos pertrechando como sistema educativo, para hacer frente a cualquiera de los efectos posibles de hacerlas públicas.

Con todo, a mí me parece que en el Consejo Técnico del INEE, en algún momento dijimos que para proporcionar información escuela por escuela, habría que cumplir tres condiciones al menos: la primera condición es que fuera una evaluación censal; la segunda condición es que hubiera la posibilidad de comparar en el tiempo, es decir, que no se cometiera la injusticia de juzgar por una sola vez a una escuela determinada, y tercera, que hubiera información de contexto que permitiera comparar lo comparable. Estas tres condiciones se cumplen y creo que el SIME permite superar aspectos que quienes estamos en contra de la devolución escuela por escuela, consideramos deben de cuidarse.



**H**aré mis comentarios básicamente desde la perspectiva de política pública donde la evaluación es un paso en el ciclo de las políticas; por tanto, desde la formulación misma de éstas, los cursos de acción a seguir deben contemplar los procesos de evaluación.

¿Esto qué significa? Garantizar que las decisiones y las acciones que se están instrumentando puedan contar con una constante retroalimentación respecto de los avances y límites en la consecución de los objetivos y de los fines estratégicos que orientan a las políticas.

El segundo principio se refiere a que los procesos de evaluación tienen que apoyarse en información objetiva y confiable, de ahí que un paso natural al instalarse la evaluación como un elemento importante en la política pública educativa, sea la creación de sistemas de información sobre diversas dimensiones: escuelas, egresados, docentes, procesos de enseñanza, de aprendizaje, impacto y otras más.

Si no existen estos diseños de información, es difícil que la evaluación pueda realmente alcanzar objetividad y a su vez dar orientaciones para las políticas públicas.

Dicho lo anterior, me referiré concretamente al Sistema de Información para la Mejora Educativa, un poco para polemizar con Sylvia.

### **Coordenadas claras para orientar decisiones**

Si bien es importante todo lo que señala Sylvia en términos de los riesgos de contar con este tipo de sistemas de información creo que bien cuidados, como ella

misma señala, pueden dar elementos o coordenadas más claras para orientar las decisiones y las acciones para la mejora educativa, y en este caso, fundamentalmente, para la escuela.

Una de las ventajas del SIME reside en que ha sido diseñado expresamente para incidir en los esfuerzos educativos de cada una de las escuelas y paralelamente contribuir a la toma de decisiones de las autoridades escolares y de las autoridades educativas.

Al mismo tiempo que proporciona información sobre los logros y debilidades de las escuelas participantes, entiende que la escuela no es el único factor a considerar, y por ello la información producida es consistentemente contextual.

La información sobre el nivel de habilidades adquiridas por los alumnos en sus escuelas, se valora tomando en cuenta variables sociodemográficas y económicas que influyen tanto en la escuela, como en los alumnos.

En el marco de la evaluación y de la creación de los sistemas de información, como progresivamente hemos caminado en el país durante los últimos años, el SIME enfoca a las escuelas y sus alumnos.

Y aquí lo importante es que no procesa la información con base en el desempeño de una escuela o un alumno ideal, sino basándose en una comparación relevante con escuelas, alumnos y factores de aprendizaje reales, así como con realizaciones y debilidades demostradas y en una serie de tiempo, es decir, no es en un momento, y además es darles una comparabilidad de acuerdo con similitudes y no hay importantes diferencias.

\*Académica de la Universidad Autónoma Metropolitana y actualmente Directora General de la Facultad Latinoamericana de Ciencias Sociales, sede México.



## Sistemas de información, instrumentos poderosos

Los sistemas de información ligados a la evaluación, son instrumentos muy poderosos para la orientación de las prácticas educativas de las escuelas y de las decisiones que se toman. En la medida en que se ha cuidado de que no se convierta en una perspectiva abstracta y normativa el diseño y la información que arroja el SIME, esto hace que los maestros y las autoridades educativas no se encuentren ante la necesidad de emular a una escuela perfecta llena de alumnos ideales, como realmente no sucedería.

Esto es muy importante y debe cuidarse mucho, no sólo en el diseño, como se cuidó con el SIME, sino en la forma como se va a manejar y a procesar la información que arroje el sistema.

Información hemos tenido mucha, pruebas hemos generado también muchas y muy diversas, pero es un buen momento para que en este avance de la evaluación y de la política educativa, con el enfoque de política pública, como se ha ido desarrollando, al menos desde el inicio de la década de los noventa, se empiecen a llenar ciertos vacíos para la gestión de la política pública, porque evaluamos mucho, pero no generamos y procesamos esas evaluaciones, están quedándose en los ámbitos de los evaluadores, y no siempre con una adecuada retroalimentación con quienes realmente son el objeto de la evaluación: escuelas y procesos de enseñanza-aprendizaje fundamentalmente, porque estamos tendiendo a la calidad y al aseguramiento de la calidad.

Con la creación del SIME se da un paso importante en términos de la gestión de la política educativa y en ese sentido, visualizar la generación de información para el diseño de políticas educativas es algo muy positivo.

Proveer a las escuelas con información objetiva y relevante sobre sus fortalezas y debilidades, siguiendo o sugiriendo modelos de inspiración realistas y alcanzables, también es positivo por el periodo abarcado; es decir, aportar a las escuelas y a las autoridades índices de avance o de retroceso en sus labores y contribuir a la cultura de transparencia de la información.

Otro elemento a considerar es que uno de los resultados importantes del SIME, si se cuida en términos de la política educativa, es que en el mediano plazo contribuirá a disminuir las resistencias a la aplicación de diversas medidas.

En la década de los noventa teníamos que instituciones educativas, maestros e incluso los mismos alumnos habían generado resistencias a la evaluación, y tener información confiable basada en pruebas validadas, que permitan hacer un seguimiento, le dan a las comunidades académicas también posibilidades, y a los directivos, para interactuar con ellos y orientar la determinación de resultados a alcanzar, así como pensar en incentivos que puedan ser utilizados.

### Algunas sugerencias

Sin embargo, hay sugerencias que se pueden hacer, además de las muy atinadas que señalaron los colegas que me precedieron. Me parece que es imperativo explicar a todos los involucrados que la información sobre las escuelas no debe basarse en un solo resultado, por eso son necesarias mediciones varias, para alcanzar niveles satisfactorios de certeza, esto se ha intentado y se ha promovido en el SIME, pero por supuesto es mejorable y perfectible.

Es importante acompañar las evaluaciones cuantitativas con investigaciones periódicas y representativas de otros componentes del desempeño de las escuelas,

en términos del ambiente y la infraestructura en los que se realizan los esfuerzos de enseñanza y aprendizaje.

Debe tratarse de evitar difundir –y esto es difícil, por lo que señalaba justamente Sylvia– jerarquías de resultados que no tengan en cuenta las diferencias del contexto de cada escuela; es decir, evitar jerarquizar fuera de los marcos de la comparabilidad, para no enviciar algo tan importante para la toma de decisiones y el mejoramiento de la escuela.

Si me permiten una sugerencia que deriva de mi sesgo profesional, es que los diseñadores e implementadores del SIME deben de estar abiertos a la evaluación del sistema de información; en otras palabras, a una constante valoración de los instrumentos que están utilizando para la configuración del SIME.

También es recomendable construir muestras representativas de estudiantes provenientes de distintas escuelas y dar seguimiento a sus logros en etapas posteriores, dados los grupos y las características de cooperabilidad que ustedes han fijado; sobre todo de la trayectoria posterior a sus estudios de primaria, para com-

plementar la evaluación que se realizó a través del IDANIS, con las particularidades que tiene el instrumento para canalizar u orientar hacia los estudios de secundaria y conocer con mayor profundidad los efectos de la educación, las oportunidades que les abrió, los obstáculos que les significó o que encontraron ellos en su desarrollo.

Tuve la oportunidad, junto con varios investigadores de la Flacso y de otras instituciones, de conocer el SIME y decir que desde la óptica de los académicos, el sistema facilita una entrada más organizada e informada a la problemática educativa. Esto permitirá profundizar más en aquellas dimensiones y variables que los investigadores han incorporado en sus protocolos de investigación, y a las cuales les asignan un potencial explicativo. Por tanto, me parece que realmente contribuye a la investigación educativa que ha sido tan importante para, incluso, orientar la política educativa durante al menos la última década.

Por último, quiero felicitar a la comunidad educativa del país, investigadores, educadores sobre todo, y autoridades educativas, por la creación del INEE.



## PREGUNTAS Y RESPUESTAS

### **ANTONIO ÁVILA DÍAZ**

En primer lugar quiero agradecer a Giovanna Valenti, Sylvia Schmelkes, Mauricio Fortes, Beatriz Calvo y a Rafael Vidal sus comentarios y observaciones sobre el instrumento que presentamos, son todos muy pertinentes. Estoy cierto que nos vienen a enriquecer.

Como tiendo a ser optimista, creo que la mayor parte de las observaciones que nos hicieron fueron positivas. Eso nos

alienta a suponer que el instrumento que diseñamos va a tener un buen uso y, por supuesto, también hay comentarios que nos llaman la atención sobre ciertas cosas que debemos cuidar y así lo haremos.

En particular me quedo con la observación que nos hace la doctora Calvo en el sentido de que es deseable que no se vaya a hacer mal uso de ese instrumento en el ánimo de dañar a planteles, maestros y escuelas. Esto es algo que nos preocupa porque de ninguna manera es la intención del



sistema. Lo último que pretendemos es venir a calificar como bueno o malo a un plantel y a sus componentes, sería triste realmente que alguien quisiera extraer este tipo de conclusiones. Sin embargo, es posible que nos enfrentemos, una vez que esto sea público, a reacciones de esta naturaleza.

Sylvia Schmelkes nos alertaba sobre lo mismo, decía: han elaborado un instrumento, pero no sabemos todavía cuál va a ser la reacción sobre éste y la capacidad de respuesta que tendremos. Lo que puedo decir es que no adivinamos las reacciones, podemos adelantar algunas cosas, pero la verdad, va a ser la realidad la que se ocupe de decirnos el impacto generado, y las dudas creadas. Las abordaremos en su momento, pues es el reto de hacer este tipo de instrumentos, plantearlos, proponerlos y una vez con la reacción de la sociedad, tratar de mejorarlos.

### **¿El SIME se articula con el programa Escuelas de Calidad?**

No. Está pensado de manera independiente, sin embargo, las primarias que forman parte del programa Escuelas de Calidad están incluidas en el catálogo que presentamos y es posible para quien quiera hacer alguna indagatoria particular, aislando el universo de las escuelas del PEC, llevarla a cabo, pues encontrará información pertinente para su estudio.

### **¿Dónde está lo novedoso del SIME si esto se ha hecho hace ya más de 10 años?**

Desde hace más de 10 años se levanta el IDANIS, pero el SIME tal y como lo estamos presentando ahora, es decir, tomar los resultados del IDANIS y comparar las escuelas sobre la base de una variable, que es la escolaridad de los padres, no se había hecho, es algo cien por ciento novedoso. Parece simple, pero no fue sencillo elaborarlo y creemos se trata de un instru-

mento importante. Si no fuera novedoso, no estaríamos aquí reunidos.

### **Me comentan también que hay un modelo de excelencia que es la secundaria anexa a la Escuela Normal Superior, y por qué no reproducir en el resto de las secundarias lo que ahí ocurre.**

Quisiera llamar la atención de que el SIME está construido para las escuelas primarias no para secundarias, en el momento en que lleguemos a hacer esto para las secundarias, que esperamos sea una ruta de investigación que podamos emprender pronto junto con el Ceneval posiblemente, el estudio de la práctica docente de la escuela anexa a la normal será de relevancia.

### **¿No sería conveniente desagregar el grado de escolaridad de los padres de familia y no quedarnos en la definición de educación básica o no educación básica?**

Posiblemente se pudiera hacer, esta vez no lo consideramos necesario, es un modelo que por supuesto no es perfecto; está sujeto a modificaciones, a mejoras, posiblemente más adelante pudiésemos pensar en desagregaciones; en esta primera etapa no se nos ocurrió, ciertamente lo haría más complejo. Alguien podrá señalar que es una sola variable de contexto la ofrecida, y se podría hacer una canasta de tres, cuatro, cinco o muchas más, pero eso nos llevaría a un grado de complejidad analítica mucho mayor. Esto posiblemente con el tiempo lo podamos alcanzar, pero por el momento pensamos que con esa única variable de contexto, podemos ofrecer resultados que tienen un grado de validez relevante.

### **SERGIO MARTÍNEZ DUNSTAN**

### **¿Hay riesgo de algún efecto perverso en la utilización del SIME?**

Efectivamente pudiera haber, hay un riesgo de algún efecto perverso, y esa sería la visión social punitiva, y obviamente aquí el problema sería que las escuelas sobrevaloradas o bien evaluadas, tengan una sobre demanda, y las escuelas valoradas un poco más abajo podrían tener una baja demanda. Tendría que cuidarse mucho esa situación. Curiosamente, esto va en contra de lo que se ha señalado en muchos estudios, la masificación de la

enseñanza va en detrimento de la calidad y si las escuelas bien evaluadas van a estar sobre demandadas podría haber algún problema. Como lo comenté también en la presentación, claro que nosotros estamos en la mejor disposición de que las cosas se hagan bien, vale la pena lograrlo, porque por más información que haya, si no está bien procesada y tiene detalles incorrectos, sería contraproducente.

## **El Sistema de Información para la Mejora Educativa**

Preguntas y respuestas





En el marco de las Jornadas fue presentado el Sistema de Información para la Mejora Educativa (SIME).

La dirección electrónica del SIME es [www.cpar.sep.gob.mx](http://www.cpar.sep.gob.mx).



TERCERA MESA REDONDA

**Los Indicadores  
del INEE: *Panorama  
Educativo 2003***

**Ponente:**

Dra. Patricia Muñiz Martelon

**Comentaristas:**

Dra. Ana Barahona Echeverría

Dra. Teresa Bracho González

Mtra. Margarita Zorrilla Fierro

Act. Andrés Ortiz Brizuela

Dr. Germán Treviño González

**Moderador:**

Dr. Héctor Robles Vázquez





## Los Indicadores del INEE: *Panorama Educativo 2003*

Por PATRICIA MUÑIZ MARTELON\*

**P**anorama Educativo de México: *Indicadores del Sistema Educativo Nacional*, es un producto que el Instituto se ha propuesto desarrollar anualmente, una publicación con todos los indicadores para la evaluación del sistema educativo nacional.

Lo que ofrecen los indicadores, o lo que intenta ofrecer este *Panorama* es una visión global del sistema educativo, incluyendo el contexto que lo enmarca y al cual, finalmente, el sistema educativo responde; por ello empezamos a considerar desde las variables económicas, sociodemográficas de México para, dentro de ellas, situar la oferta y la demanda del sistema educativo y la respuesta que éste ofrece a nuestro país.

¿Qué pretendemos? tener un seguimiento del sistema educativo que permita, como ya dije, evaluarlo, diagnosticarlo, y ofrecer recomendaciones sobre las problemáticas que se vislumbran.

El objetivo ha sido en este tiempo la construcción de un sistema de indicadores para evaluar el sistema educativo; esto implica enfrentar una situación en la cual el sistema tiene un dinamismo propio para irse adaptando, así como el sistema educativo tiene el suyo propio, lo cual quiere decir que no podemos llegar a tener un sistema de indicadores fijo y permanente, porque vamos dando respuesta también a la dinámica del sistema, y a los desafíos implícitos entre los que se encuentra

la disposición de información estadística para el desarrollo de indicadores.

Vamos a presentar entonces el resultado: una publicación de aquellos indicadores que por el momento hemos ido desarrollando, no quiere decir que se trate de un sistema ya terminado y que sean todos los indicadores considerados, los cuales deberían ser útiles para evaluar, medir o diagnosticar al sistema educativo, sino que es el primer paso para elaborar nuestro trabajo y perfeccionarlo, haciéndolo cada vez más útil para la sociedad y las autoridades, quienes toman las decisiones sobre el sistema educativo.

El proceso que implica desarrollar un sistema de indicadores se divide en tres apartados: primero, el establecimiento de los objetivos y el modelo de evaluación a usar para evaluar el Sistema Educativo Nacional.

Segundo, las definiciones teóricas y metodológicas de los indicadores. El sistema de indicadores es un conjunto de los mismos, los cuales definen algunos criterios útiles que se integran de manera clara para delimitar todas las funciones del sistema y se complementan para tener todos los datos necesarios y así juzgar el estado del sistema educativo nacional.

El tercero, es obtener información necesaria, conceptual y técnicamente útil y de calidad para elaborar estos indicadores.

\*Directora de Indicadores Educativos del Instituto Nacional para la Evaluación de la Educación.



## ¿Qué es el Panorama Educativo 2003?

El documento *Panorama Educativo 2003*, es una primera aproximación del desarrollo del sistema de indicadores para evaluar el desempeño y la calidad del sistema educativo mexicano.

¿Qué nos entrega éste? una visión general de la realidad educativa y también el esfuerzo por aportar un instrumento útil para la rendición de cuentas y toma de decisiones.

¿Qué contiene el libro? Primero, tiene la parte de la misión institucional, los aspectos generales y después viene descrito un conjunto de indicadores distribuidos de la siguiente manera: siete, del contexto socio educativo; ocho, sobre el contexto socio económico; nueve, sobre el contexto sociodemográfico; siete, sobre recursos humanos en el sistema educativo; seis, sobre recursos materiales en el sistema educativo; cinco, sobre acceso y permanencia; diez, sobre procesos escolares, y diez más, sobre resultados educativos.

La siguiente parte del *Panorama Educativo 2003* presenta unas tablas resumen de indicadores, en esta sección los indicadores están presentados por entidad federativa.

Para finalizar, el documento contiene una propuesta de los indicadores que hasta el momento vislumbramos se pueden ir desarrollando en el futuro.

¿Cómo se presenta cada uno de los indicadores dentro de nuestro libro? Los indicadores están integrados por seis elementos comunes para casi todos ellos. Primero, tienen una ficha de identificación que los define; la fórmula con la cual se calculó el indicador, la fuente de la cual se obtuvieron los datos para calcularlo y la interpretación que damos a ese indicador.

Después, hay una pequeña descripción de la importancia que, consideramos, tiene para nosotros el indicador dentro de

nuestro sistema y la presentación de los resultados.

En muchos de ellos, hasta donde hemos podido hacerlo, y cuando en algunos casos es necesario, encontramos una referencia comparativa internacional. No todos nuestros indicadores, por el momento, han podido ser construidos de manera que podamos compararlos con la información internacional disponible, las fuentes de información presentan variaciones y solamente en algunos de ellos hemos podido comparar, porque sabemos que el indicador contra el cual nos estamos comparando está construido de la misma manera, pero nuestra intención es lograrlo, y podamos construir indicadores que nos permitan medir nuestro sistema educativo con los de otros países, para de esta manera tener una idea más clara de cómo está la situación de México en relación con el resto del mundo.

En otros casos, cuando las fuentes lo han permitido, los indicadores presentan algunas desagregaciones. Cuando es pertinente y necesario, porque se hizo algún corte específico con la información, se presenta una nota técnica desglosando con toda claridad qué tipo de cortes se realizaron a la fuente de información, si se hizo alguna selección de datos, cómo fue hecha, o bien, si el dato puede presentar algún sesgo o limitación por la misma fuente de la que proceden los datos.

Otro elemento es la ficha de identificación, la cual aparece para cada uno de los indicadores propuestos en el *Panorama Educativo*. Contiene el nombre del indicador, fórmula de cálculo, definición, interpretación y fuente de información.

Otro rubro, es la manera como están presentados los resultados. En general, encontramos una primera tabla que dice: años promedio de escolaridad de la población de 15 a 64 años de edad. Normalmente se presenta una primera tabla que

nos muestra el total por entidad federativa y el total nacional con una pretensión de que tengamos, en aquellos casos donde es pertinente, una distinción por género.

La segunda parte nos ejemplifica otra forma de presentar la información, a través de gráficas. En este caso se encuentra el componente de *índice de acceso a los medios de comunicación*, en las viviendas que fueron censadas.

Para finalizar la presentación de cómo están los indicadores, mostramos lo que sería una nota técnica. En este caso corresponde a los medios de comunicación en el hogar; aquí se expresa cuáles son los cortes y qué se está considerando como *medios*, para sacar un índice en el hogar, y además encontraremos explicitadas las fórmulas de cómo están tomados estos datos.

En otra gráfica se presenta una parte de una hoja resumen por entidad federativa, donde encontramos cuatro columnas que presentan, en primer lugar, la cifra del promedio nacional, o en el caso, si es, por ejemplo, la población, el total de población del país, después viene la cifra del estado correspondiente; si ésta fuera la hoja de Aguascalientes, la segunda columna correspondería a los datos de la entidad, y después hay otras dos columnas, una de ellas menciona el máximo que hemos encontrado en la República y la otra el mínimo, para que cada persona pueda consultar la hoja de una entidad, los referentes, el indicador a nivel nacional, y cuente con una aproximación útil para evaluar qué tan cerca o lejos se ubica del máximo posible en algún aspecto.

Ésta es una parte que aún no hemos terminado de trabajar, porque lo deseable es tener ciertos parámetros contra los cuales comparar las cifras que obtenemos en el indicador, y determinar con mayor certeza si es positivo que sea más alto, o es negativo.

## **Alcances, limitaciones y aportes**

Existen indicadores los cuales hemos considerado deberían de estar en nuestro sistema, así como algunas limitaciones por las cuales aún no los podemos tener. En algunos de ellos no existe la información para desarrollarlos, nunca se han preguntado en ninguna encuesta y la SEP tampoco lo ha realizado, aunque en algunos casos, tiene información que podemos utilizar. En otros casos, no hemos llegado a definiciones conceptuales claras de qué tipo de aspectos colocar en el indicador, en otros más, porque técnicamente no se ha logrado precisar cómo desarrollar el indicador.

Un ejemplo de ello, nos fue sugerido por un grupo de expertos, a los cuales presentamos los indicadores para que fueran los jueces de la validez del trabajo, y es el relativo al *índice de bienes culturales en el hogar*. Antes de podernos meter a revisar las fuentes de información, necesitamos hacer una definición muy clara de qué significan los bienes culturales de un hogar en términos de su influencia, de su aporte al sistema educativo, o su vinculación con él, de lo contrario, el indicador no sería útil para nuestro sistema. Es el trabajo que estamos desarrollando, muchos de ellos están considerados pero aún no los tenemos construidos y es lo que muestra el libro.

En la publicación también encontramos algunos alcances y aportes, una descripción comprensiva y amplia de las condiciones contextuales y endógenas del Sistema Educativo Mexicano y su operación. Tenemos en algunos casos la posibilidad de comparar internacionalmente parte importante de la información que estamos construyendo.

Acerca de la exploración sobre características relevantes de la planta docente, deseo expresar que alguna información utilizada es sobre Carrera Magisterial, al



respecto tenemos muchos datos incluidos aquí, por desgracia con algunas limitantes.

Por otra parte, hay una desagregación por género de la mayoría de los indicadores, así como medidas alternativas acerca del tránsito del alumnado en primaria y secundaria; algunos de esos indicadores los hemos podido desagregar un poco más. Y tenemos una presentación sintética del escenario educativo en el año, por entidad federativa.

Ahora bien, todavía existen bastantes limitaciones y algunas salvedades que hay que aclarar: falta un desarrollo conceptual, pues lo que tenemos hasta el momento es un poco incipiente.

Contamos con poco equilibrio en el avance de los componentes del sistema; si ustedes ya revisaron el texto, se podrán dar cuenta de la presencia de muchos indicadores cargados hacia las partes de contexto, o sea, en número son más que los encontrados en la parte de procesos educativos, los cuales nos toman un poco más de tiempo en el desarrollo conceptual. Otro problema, es una limitación un poco difícil de manejar y creo no lo vamos a lograr, aunque por supuesto lo intentaremos: es el asunto de la información, la cual viene en varios cortes de tiempos diferentes, y las fuentes de información generadas, que obviamente no son nuestras, y las cuales se crean con temporalidades diferentes.

Aunque nosotros decimos que es un *Panorama Educativo de 2003*, las fuentes de información que hemos incluido pueden ir desde 1998 ó 2000, una es el censo como fuente importante, con una temporalidad hasta el 2010, y si hacen un conteo, probablemente tengamos información en 2005. Otra fuente de información es la SEP, quien nos proporciona información año con año, lo cual permite obtener una fotografía no instantánea, sino de un periodo amplio, en el que están considerados

posiblemente cinco o diez años, en algunos indicadores de manera más centrada, en otros no. Pero no es una foto estática; pues debemos jugar con esos tiempos y esa visión.

### Fuentes de información

Otro punto a tratar son las fuentes de información utilizadas hasta el momento y mostrar precisa y exactamente la periodicidad y dificultad que tienen, dejar claro que dependemos completamente de la información, sobre todo, la que no podemos levantar por cuenta propia, como toda aquella que viene del INEGI, las encuestas nacionales, el censo, etcétera; todas las fuentes de información de la SEP, de la DGPPP. Por ello caminamos con paso un poco más lento, pero ésta es nuestra circunstancia y con ella vamos a trabajar lo mejor que se pueda.

En otra parte del libro se muestran las diversas fuentes y las temporalidades diferentes, las cuales presentamos en los mismos bloques; la primera, el porcentaje de madres que completaron su educación básica por entidad federativa, corresponde y está situada en el 2000 porque se toma del censo del INEGI.

El siguiente porcentaje de escuelas primarias multigrado y alumnos atendidos en ellas, por entidad federativa, está tomada de la DGPPP del periodo 2002-2003. Luego tenemos eficiencia terminal en primaria y secundaria por entidad federativa también de la misma fuente, pero de 2001-2002.

En otros casos, la misma fuente es parcial, como ejemplo tenemos información de la edad promedio de los docentes, por entidad federativa, éste es uno de los datos que hace unos días alguno de los medios de comunicación difundió, tomado exactamente de nuestro anuario, pero el periodista no hizo la lectura completa; evi-

dentamente hablaba de lo docentes del país y nosotros hemos marcado una limitante: los datos son de los docentes del país, pero sólo los inscritos en Carrera Magisterial.

Quiero aclarar, antes de terminar: estamos intentando no sólo medir al sistema educativo con la información de todos sus integrantes. Nuestro interés también es, diagnosticar al sistema con todos aquellos no incluidos, saber quiénes son los niños que no han llegado y entonces ver desde

esa perspectiva cuál es la respuesta o por qué el sistema no ha logrado acercar a todos, atraerlos, contenerlos y enseñarles, lograr así la cobertura del sistema educativo.

Ahora pido a mis compañeros de la mesa sus aportes y luego a todos ustedes, estamos abiertos a todas las críticas, comentarios y retroalimentaciones, porque ésta es una tarea dura, la cual va empezando, es el primer paso, pero se intenta que sea útil para todos nosotros.

---

## **PARTICIPACIÓN DE LOS COMENTARISTAS**

**ANA BARAHONA ECHEVERRÍA\***

---

**E**n el Decreto de Creación del INEE, quedó establecido que su misión es la de contribuir a que el Sistema Educativo Mexicano ofrezca una educación de buena calidad con equidad a todos sus alumnos.

Para lograr esto, el Instituto básicamente ha venido realizando tres acciones: primera, diagnósticos permanentes los cuales reflejan la situación del Sistema Educativo Nacional con los subsistemas con los cuales cuenta y ayudan, en la medida de lo posible, a explicar los factores causales que la determinan.

Segunda, difundir sus resultados de la manera más transparente, para desarrollar una verdadera cultura de la evaluación y colaborar con la rendición de cuentas; y

Tercera, promover el uso de la evaluación para apoyar la mejora de la calidad, obviamente por razones de tiempo no voy a poder hablar de todas estas cuestiones, me gustaría ahondar sobre el punto de los diagnósticos.

### **El Sistema de Indicadores**

En el Decreto, como les decía yo, se establece que es función del Instituto, desarrollar y mantener un Sistema de Indicadores, con el cual podamos ser capaces de observar el estado del sistema en sí mismo y su movimiento a lo largo de los años. ¿Para qué queremos hacer eso?, con el propósito de elaborar diagnósticos de la situación educativa, y saber en qué medida se está cumpliendo o no con la equidad, eficiencia, relevancia e impacto social de la educación.

Esta función de la institución es de largo aliento, pues será sólo a través de varios años que se contará primero, con un sólido Sistema de Indicadores en virtud de lo acabado de exponer por Patricia, y entonces se podrá observar el movimiento de alguno de ellos y evaluarse verdaderamente al sistema.

Desarrollar indicadores complejos, los cuales combinen información de estadís-

---

\*Profesora e investigadora de la UNAM en la Facultad de Ciencias. Miembro del Sistema Nacional de Investigadores y miembro del subgrupo de indicadores educativos del Comité Técnico del INEE.



ticas educativas con resultados de pruebas de aprendizaje y evaluación de escuelas, además con elementos del contexto socioeconómico de las escuelas, es un reto para el Instituto; y entre las primeras acciones, la inmediata fue validar indicadores diseñados con anterioridad para otros propósitos; monitorear la confiabilidad de la información en la cual se basaban éstos; hacer análisis rigurosos de la información del Sistema de Indicadores Nacionales y también estatales, indicadores con los que ya se contaba, de tal forma que se cubrieran los estratos rural, urbano, público, privado, etcétera.

Como ustedes saben, el Instituto cuenta con tres áreas básicas: Indicadores Educativos, Pruebas y Medición, y Evaluación de Escuelas. El área de indicadores es el componente que le da unidad al conjunto, la construcción de un sistema complejo de indicadores implica la aplicación habitual de pruebas de aprendizaje, a muestras representativas de alumnos, cuyos resultados complementan información administrativa existente, de igual forma los indicadores sobre procesos complejos y delicados como son la enseñanza y el aprendizaje, tanto los que tienen lugar dentro del aula como al interior de las escuelas.

Entonces, el área de indicadores proporciona unidad al conjunto del sistema de evaluación, en la medida que se precisan los aspectos particulares referidos hace un momento: relevancia, eficiencia, eficacia, impacto y equidad.

Para tal efecto, el Sistema de Indicadores que nos ha presentado Patricia, se concibe como un conjunto bien derivado lógicamente de un concepto de calidad, el cual se caracteriza por su elevada técnica, tanto en la recopilación, integración, análisis y presentación de la información.

Un Sistema de Indicadores para evaluar al Sistema Educativo Mexicano, debe tener como objetivo proporcionar información sobre el estado y evolución del mismo,

en forma de un conjunto de indicadores como se presentan en este volumen, los cuales cubran todas las dimensiones del concepto de calidad, y todos los niveles y componentes del sistema, pero además, los debe presentar de una manera sintética y continuada para realizar análisis longitudinales y en forma congruente con los sistemas internacionales, de manera que puedan hacerse también comparaciones transversales con otros países y otros sistemas educativos.

### **¿En dónde estamos?**

Bueno, ¿qué se ha hecho? Cuando se creó el INEE, una primera aproximación al campo de los diagnósticos, fue que se contaba con mucha información, ésta era muy heterogénea, en muchas ocasiones con deficiencias técnicas y un nivel muy elemental en la explotación de los datos, en suma, se podría decir que no existía una visión integral de un verdadero Sistema de Indicadores Educativos.

Así, una de las primeras tareas de la Dirección de Indicadores Educativos, fue investigar lo existente con el propósito de delimitar la validez de la información con la que se contaba, buscando en diferentes fuentes la relevante para la construcción de indicadores propios.

Empezaron a desarrollarse varias líneas, pero en particular aquella que trataba de la evaluación de factores de contexto, referidos a la elaboración de indicadores que sinteticen información más importante del entorno de alumnos, maestros, directores y escuelas.

Este tipo de estudio es lo que se concibe como un estudio a nivel macro, ya que involucra a un sistema complejo y los resultados, su impacto social, económico, cultural; y por ser un estudio macro, es la parte que articula, como decía yo hace un momento, todas las labores del Instituto.

Una vez que se empezó la labor de sistematización proveniente de las fuentes como el INEGI, la SEP, etcétera, se elaboraron los primeros indicadores, los cuales fueron sometidos a un escrutinio severo. Esta tarea empezó básicamente después de la creación, hace un año y medio; se formaron grupos con expertos para discutir la relevancia y las limitaciones técnicas de los indicadores, luego se formaron grupos de jueceo en donde se involucraron expertos y usuarios, con el propósito de elevar la calidad técnica en la valoración de estos indicadores.

Posteriormente se presentó el trabajo en diferentes ámbitos: personal de la SEP, delegaciones estatales, representantes de los padres de familia, etcétera; con la finalidad de ver la reacción ante los indicadores y saber un poco más acerca de su relevancia y su impacto.

Hubo retroalimentación de docentes, directores, supervisores escolares, etcétera, quienes colaboraron en la calibración y ajuste de algunos de ellos, con el fin de garantizar un uso masivo de los datos producidos por el INEE. El resultado del trabajo a lo largo de un año y medio, es el libro que hoy nos acaban de presentar.

Hace seis meses, debo decir, se distribuyó en forma electrónica el ejemplar que ahora está en papel, cumpliendo así con un aspecto fundamental del Instituto: la difusión de sus resultados.

Con la publicación del tomo de Indicadores del Sistema Educativo Mexicano se da, yo creo, un paso sin precedentes en la evaluación de la educación en nuestro país. Creo que vamos por muy buen camino y los felicito a todos.



## TERESA BRACHO GONZÁLEZ\*

Lo que voy a comentar gira no tanto en funciones específicas del documento, que ya ustedes habrán tenido oportunidad de ver o tendrán oportunidad de leer en su versión impresa, sino tratar de ver entre líneas cómo la función del INEE se expresa en este documento; cómo lo hecho por el INEE en estos dos años parece deberá reforzarse.

### ¿Qué es evaluar?

La primera tarea del INEE fue decir: ¿qué es evaluar? Estamos acostumbrados en México a que evaluación significa sanción, calificación, clasificación, con miras a

un proceso de promoción de grado o de certificación, y ese no es todo el significado de evaluar.

Evaluar, en la corta existencia que tiene el INEE, ha implicado reconocer y dar a conocer, creo que es un trabajo muy importante de comunicación hecho por el INEE, que la evaluación tiene un sentido más amplio, el cual poco tiene que ver con esa función punitiva.

Si uno mira el Diccionario de la Lengua Española, evaluar significa todo lo contrario, significa reconocer el valor de algo. En este caso, es reconocer el valor de la educación, es decir, registrar las cualidades del proceso educativo y qué es lo que éste

\*Profesora e investigadora del Centro de Investigación y Docencia Económicas (CIDE). Es miembro del Sistema Nacional de Investigadores y miembro, también, del subgrupo de indicadores educativos del Comité Técnico del INEE.



agrega o deja de agregar a sus alumnos y a la sociedad en donde se han de desarrollar más adelante. Es decir, se trata de identificar el valor agregado del proceso educativo y de los medios por los cuales alcanza ese valor.

Se trata, por tanto, de emitir un juicio sobre el Sistema Educativo Nacional. Reconocer el valor de algo, de ninguna manera significa ser complacientes con lo que sucede en el sistema, se trata más bien de hacer explícitos los valores que se esperan alcanzar de la educación y el grado hasta el cual se está logrando su cumplimiento.

Como señalaba la doctora Muñiz, estos valores además son dinámicos, cambiantes, cada vez más exigentes y el sistema de evaluación tendrá que ir registrando esa dinámica en la demanda al sistema educativo.

En este primer sentido, ¿qué es evaluar? el documento del *Panorama Educativo de México* muestra que el centro de atención del INEE es el Sistema Educativo Nacional y, por tanto, el trabajo de indicadores presentados aquí permite señalar y puntualizar rezagos y logros del sistema, pero éstos son muy distintos a los que habíamos entendido en el pasado.

Quien haya trabajado con estadísticas nacionales, yo lo he hecho a lo largo de varias décadas, se puede dar cuenta que este trabajo es totalmente distinto, cualitativamente distinto: son los mismos datos pero explicados, son calculados con una orientación conceptual muy clara, con una explicitación de cuáles son las formas de cálculo y no se trata, como se hacía en otras ocasiones, de buscar cuál es el indicador más bonito para demostrar que sí hemos alcanzado todas las metas del Sistema Educativo Nacional, sino de encontrar cuál es el panorama complejo de este sistema.

Por tanto, el libro intenta reflejar esa complejidad en el sistema, reconocer que el Sistema Educativo Mexicano es muy complejo, y tratar de reflejar esa comple-

jididad para abrir la puerta a la contextualización de resultados, logros y rezagos del sistema de evaluación en pruebas, y evaluación de escuelas.

Es decir, se trata de un juicio responsable. Esta idea de responsabilidad, la cual está detrás de la minuciosa explicación de la construcción de cada indicador, sus fuentes, sus limitantes, sus sesgos potenciales, creo que es muy novedosa y permitirá a la larga identificar los espacios para la mejora del sistema.

### ¿Por qué y cuándo evaluar?

La idea de evaluar cuando no se confía en el objeto evaluado, o cuando no se cree en el sujeto sometido a evaluación, es algo que ocurre por la ideología nacional. Y no es el caso, creo que tanto la sociedad política y la sociedad completa, tiene alguna confianza en el sistema educativo. Si no, no me imagino que el Congreso dedicara esa cantidad enorme de recursos al sistema educativo.

Entonces, se debe partir de ese reconocimiento que la sociedad hace, de que algo hace bien el sistema y lo puede hacer mucho mejor. En este sentido, la función de la evaluación es acompañar el proceso educativo y enriquecerlo; no castigarlo, sino enriquecerlo, mejorarlo.

La evaluación busca identificar las capacidades del sistema y señalar de manera responsable sus deficiencias, para poder contribuir a su solución.

No sobra reiterar que ahora el *Panorama Educativo*, por ejemplo, desagrega en muchos indicadores y de manera conjunta, cosa que no existía antes, el tema del rezago.

No se trata de decir *la cobertura ya no es un problema, ahora hablemos sólo de calidad*, sino reconocer que no están todos los niños en la escuela, y saber ¿por qué no están?, ¿quiénes no están? y ¿cómo llegar a ellos?

¿Y por qué? Puede ser que sea un porcentaje mínimo en términos relativos, pero el mandato constitucional se mantiene en la norma del cien por ciento, todos los mexicanos deben de tener educación básica al menos.

En cuanto a qué hacer con la evaluación. No se trata de encontrar culpables, no se trata de esconder resultados; al contrario, se trata de generar mejores sistemas de comunicación de esos resultados. Usualmente, este tipo de información se manejaba sólo para expertos y estaba contenida en un CD o en una base de datos, la cual uno tenía que encontrar, saber cómo rastrear y tratar de interpretar todo el censo, todas las encuestas de ingreso-gasto.

Lo que hace este volumen de indicadores 2003 es comunicar la información existente y disponible en este momento, comunicándola de manera significativa a la sociedad y a los expertos, incluso la información disponible en distintas fuentes. El rastreo de información es realmente impresionante.

### **Información responsable**

Esta información sistematizada y organizada conceptualmente, yo hubiera gozado tenerla antes, pero creo que es importante y va a ser una referencia importante para los académicos; también busca que lo sea para toda la sociedad, porque aquí sí se puede entender qué significa cada indicador, y uno no tiene que ir a rastrear a la burocracia generadora del indicador para saber qué entendía por rezago o qué entendía por cobertura, o si tomo el inicio de este ciclo o el fin de ciclo. Digámoslo claramente, aquí está perfectamente explicado qué se toma y cómo se toma, y responsablemente se informa sobre ello.

Se debe reconocer que la evaluación tiene muchas funciones, hay información

diagnóstica, información de procesos, información de muchos tipos, y parte de este volumen es un primer paso a un diagnóstico mucho más rico de nuestro sistema educativo.

Lo que tiene este sistema es que los indicadores se convierten en interpretables y comunicables. Los indicadores no son mejores cuando la ecuación es extensa, los indicadores son mejores cuando la gente puede entender qué se quiere decir, y se comunican los límites de interpretación de cada uno de ellos.

Por último, ¿qué evaluar? Esto ha sido parte de las complejidades en el Consejo Técnico, en el propio Instituto y parte de la discusión del INEE.

La parte de indicadores, que es donde colaboro en el Consejo Técnico, se refiere a cómo definir claramente el objeto de evaluación, y cómo expresarlo de manera rigurosa y sustentable en indicadores de evaluación.

Como muestra el trabajo sobre indicadores del sistema, es claro que la función del INEE no es sólo aplicar pruebas de rendimiento, evaluar niños, maestros y escuelas; se trata más bien de entender qué permite que niños con cierto tipo de maestros y en cierto tipo de escuelas alcancen los logros de aprendizaje, habilidades, actitudes, etcétera; a qué se compromete el sistema educativo, y reconocer que los niños viven en muchos *Méxicos*; México no es homogéneo, hay una enorme diversidad social, cultural y económica, la cual también es parte de esta evaluación, para entender los alcances del sistema.

Se trata de tener una visión del sistema educativo completo. ¿Podría alguien argumentar que eso es evadir el tema de, por qué alguna vez se planteó en la prensa el compromiso de evaluar niño por niño y escuela por escuela? Es todo lo contrario, es reconocer que la responsabilidad del desarrollo de niño por niño, escuela por es-



cuela, es una responsabilidad del sistema educativo; y que el tratar de identificarlo de manera conjunta, sistemática, seria y

responsable, es parte de esta tarea para contribuir a que todos los niños y escuelas logren mejorar su rendimiento.

---

## MARGARITA ZORRILLA FIERRO\*

---

### ¿Para qué sirven los indicadores en educación?

Quiero empezar contándoles una experiencia con esto que llamamos indicadores educativos. Justamente hace 10 años tuve la oportunidad de realizar un trabajo con el cuerpo de supervisores de preescolar, primaria y secundaria en el estado de Aguascalientes.

En ese estado, se quiso aprovechar el impulso que significó la puesta en marcha de la política de descentralizar el sistema de educación básica, también conocida como federalización de la educación, y se tomó la decisión de llevar a cabo una reforma de la manera de funcionar de la supervisión escolar.

No voy a narrar aquí toda la experiencia, sólo me voy a centrar en un punto muy particular, el cual tiene que ver con el tema que nos ocupa en esta mesa: los indicadores educativos.

La puesta en marcha de una serie de medidas que permitieron reorganizar el servicio de supervisión, exigió contar con una buena dosis de capacitación. Así, el lema de las actividades de capacitación de los supervisores fue *dotar de nuevos contenidos y formas de trabajo a la supervisión escolar*.

La estructura del primer programa de capacitación se organizó en tres módulos:

- El diagnóstico de indicadores de eficiencia;

- Análisis de la política educativa y,
- Manejar una propuesta metodológica para el diseño e implantación de un proyecto educativo desde la perspectiva de la supervisión.

Pensamos este orden apelando a los viejos principios pedagógicos de lo cercano a lo lejano, de lo familiar a lo desconocido, con el tiempo afirmamos la decisión de empezar con el tema de los indicadores de eficiencia.

¿En qué consistió el trabajo de diagnóstico de indicadores de eficiencia de las escuelas? Fundamentalmente, en hacer una traducción de una metodología de investigación cuantitativa y estadística y colocarla en manos de los supervisores.

Se trató de desarrollar habilidades para calcular indicadores de eficiencia; como niños en primero de primaria que cursaron preescolar; reprobación; deserción; eficiencia terminal de la primaria o secundaria; índice de absorción de los egresados de la primaria por la secundaria, y otros más.

Estos indicadores son los más convencionales y cuyo uso es tradicional en el sistema educativo. ¿De dónde debían los supervisores obtener la información básica para elaborarlos? De la Forma 911 que llenan los directores de la escuela al inicio y fin de cada ciclo escolar y la cual representa la base de la estadística continua de la SEP.

---

\*Académica en la Universidad Autónoma de Aguascalientes y miembro del Consejo Técnico del INEE.

Los supervisores dan el visto bueno a la información, la cual entregan a los directores y hacen un concentrado de sus zonas escolares; eso significa que es un material que conocen; sin embargo, a pesar de que los directores y supervisores son los productores de tal información elemental del sistema educativo, ni les interesaba ni confiaban en ella.

Los números son fríos, se decía durante los talleres, la misión del equipo de instructores era muy clara, aprender a calcular los indicadores y con la reflexión hacer hablar a los números y, construir un diagnóstico de la educación básica a partir de la eficiencia de sus escuelas, poca cosa.

A los instructores les decíamos, ustedes sí pueden interpretar los resultados de sus indicadores, pues estamos suponiendo que conocen bien a sus escuelas; por ejemplo, ¿por qué esta escuela que tiene una alta reprobación al término de un ciclo escolar equis, en el ciclo siguiente no tiene repetición, no les parece raro?

Lo que se fue develando en estos casos, es que existen escuelas expulsoras, no admiten a chicos repetidores o resulta que en este municipio terminan la primaria una buena cantidad de niños, pero luego no ingresan a la secundaria, ¿qué pasa? ¿por qué en tales o cuales escuelas o zonas abandonan la escuela una cantidad impresionante de alumnos en el primero de primaria?

Las respuestas se fueron aportando con la experiencia y conocimiento que sí tienen los supervisores, los ejemplos son muchísimos.

El eje articulador del trabajo de diagnóstico de la eficiencia de la educación básica fue una idea muy sencilla, suelo decir que es una definición *ranchera* de la calidad de la educación: *un sistema educativo en general o una escuela es de calidad, si los chicos y chicas que tienen la edad de ir a la escuela acuden a los centros escolares, tie-*

*nen acceso, y cubren la asistencia. Además se quedan en ella y aprenden cuestiones relevantes para su vida presente y futura, permanencia y aprendizaje relevantes, y consiguen concluir con éxito cada grado escolar y nivel educativo como corresponde: egreso oportuno.*

*Igualmente, un elemento importante es el de la equidad, comprendida como la distribución de las oportunidades para aprender todos, porque calidad y equidad son un binomio inseparable.*

### Recuento de un año de trabajo

Identificar los datos correspondientes de cada indicador, calcular el indicador para cada escuela, hacerlo para escuelas rurales y urbanas por grado escolar donde esto es pertinente, integrar una interpretación de los resultados, es un diagnóstico de la situación de un conjunto de escuelas, y a partir de éste, plantearse objetivos y metas para el mejoramiento de la calidad, fue la tarea que nos ocupó un año de trabajo directo.

El aprendizaje colectivo ha sido muy importante, realmente significó lo que podemos llamar una ruptura epistemológica, que colocó a los supervisores en una tesitura distinta frente a la realidad educativa de las escuelas, y a las posibilidades para comprenderla desde un punto de vista muy específico. Podemos decir que empezaron a aprender a mirar a la escuela con otros ojos, la historia ha continuado desde entonces.

Con esto que comparto hoy con ustedes, espero estén imaginando que soy una convencida y enamorada de este tipo de trabajo en la educación.

En el sistema educativo se necesita hablar de todos los elementos que lo conforman de otra manera, necesitamos saber con la mayor precisión posible, si el sistema educativo tiene la capacidad de tener



en la escuela básica a todos los niños y niñas que deben estar en ella. A esto le llamamos cobertura y es un punto de partida fundamental, si los niños no asisten a la escuela, vanos serán los esfuerzos del sistema y nos podemos preguntar ¿quiénes van a la escuela? ¿quiénes no lo hacen? y ¿cuáles son las razones de ello? ¿qué puede hacer el sistema educativo para asegurarle a todos los niños el acceso a la escuela y que asistan, se queden y aprendan? Y puedo seguir con más y más preguntas, que espero ustedes se estén planteando ahora mismo.

El INEE asume el compromiso de poner en marcha un sistema de indicadores educativos que den cuenta de las distintas dimensiones de la calidad de la educación, y los cuales nos permitan observar a lo largo del tiempo su evolución o su involución, porque no debemos olvidar que el sistema educativo es un organismo vivo por lo que sus procesos son dinámicos y cambiantes.

Este año de trabajo en el INEE que me consta ha sido arduo, por lo que significa aterrizar y concretar propósitos, aspiraciones, ideas o convicciones, empieza a dar frutos.

### Un panorama de indicadores

En materia de indicadores tenemos hoy en nuestras manos el primer producto de este trabajo, *Panorama Educativo de México 2003*, es el título del libro que sintetiza cientos de horas de trabajo y nos ofrece eso, un panorama amplio y diverso de nuestro sistema educativo a través de distintos tipos de indicadores.

Quiero destacar que la tarea de diseñar e implementar un sistema de indicadores educativos que nos permita valorar el estado que guarda la educación básica mexicana, supone un trabajo de definición de qué necesitamos saber, qué que-

remos saber y qué podemos saber en un momento dado, de recabar la información básica, la cual nos permita construir el indicador, sistematizar la información y realizar los cálculos de identificar parámetros para compararlos, en fin, es una cadena de decisiones complejas y por su naturaleza son dinámicas.

Fíjense ustedes en un ejemplo de los censos nacionales de población. Ahora nos parece normal que digamos: somos tal porcentaje de mujeres y tal otro de hombres, y esto lo podemos desagregar hasta el nivel más pequeño de la localidad. Pues hace unos pocos años era imposible hacerlo porque no interesaba, no era una necesidad saber cuántos éramos atendiendo al sexo, ni siquiera era una pregunta del cuestionario del censo.

Pero esto no se ha quedado ahí, sólo en la mera contabilidad, sino que ahora se calcula lo que se llama: *el índice de masculinidad*, que se define como la proporción de varones respecto a las mujeres en un sitio determinado, así, ahora podemos saber en qué lugares hay una mayor concentración de mujeres o de varones; para definir políticas de atención a la población femenina esto es fundamental.

Creo que en estas pinceladas nos podemos dar cuenta de por qué son importantes los indicadores educativos, y de por qué es importante realizar un trabajo cuidadoso de definición y construcción de la información básica requerida. Esto significa que hay cosas que necesitamos saber pero de las cuales aún no tenemos la información requerida, se me ocurre, por ejemplo: ¿cuánto cuesta el rezago educativo, cuánto cuesta en Baja California o en Chiapas o en el Distrito Federal? La no calidad también cuesta y suelo pensar que es más cara que la propia calidad.

Puede parecer una verdad de *Perogrullo*, pero no podremos mejorar la educación ni el sistema educativo dando palos

de ciego, necesitamos conocer y comprender lo que sucede, tener información confiable, creíble, válida, útil y a lo largo del tiempo, para que las decisiones que se tomen para fomentar el desarrollo de la educación realmente sean pertinentes al quehacer de las personas en lo individual y relevantes en lo colectivo.

Espero que éste sea el primero de una serie de encuentros a los que el INEE nos convoque, para analizar y reflexionar acerca de la educación básica en México y nos vaya compartiendo sus hallazgos y dificultades en la construcción de un sistema integral de indicadores educativos.

---

## ANDRÉS ORTIZ BRIZUELA\*

---

El hecho de que haya muchos indicadores no significa que se esté trabajando bien, debe hacerse un buen ejercicio de análisis, un ejercicio de articulación que es parte, creo, de la tarea a la que se está enfrentando el Instituto: articular lo que está suelto, mejorar aquellos elementos susceptibles de evolucionar.

Al ver este documento, reflexiono en torno a los indicadores en dos aspectos: uno se refiere a la construcción y el otro es el del uso. En principio, un indicador ideal debe derivarse de un modelo bien diseñado, el cual nos va a llevar a un indicador adecuado.

Otro aspecto es la factibilidad. Si idealizamos mucho y queremos obtener el color de los zapatos de los niños para ver cómo están estos niños en la escuela y pensamos en esa información a la hora de construir los indicadores, por supuesto no podremos concretarlos. Estos son elementos muy importantes, los cuales debemos tener en cuenta quienes estamos involucrados en la tarea.

Un aspecto más, es considerar la información que existe, pienso es lo primero que debemos hacer, y este esfuerzo de publicar el libro representa un paso impor-

tante en ese sentido. Quienes desarrollamos proyectos, somos dados a crear nuestros propios indicadores, no queremos retomar lo existente por considerar que no nos da personalidad a nosotros, ni a nuestros proyectos, ni a nuestra institución.

### **Analizando lo existente**

El primer paso que se ha dado es el más adecuado, debemos analizar lo existente y ver si responde a los objetivos planteados, qué de todo se debe utilizar para evaluar la calidad de la educación.

El concepto de calidad, como todos sabemos y se ha dicho muy reiteradamente, es un concepto muy complejo, y parte del reto radica en ver qué de todo lo que existe, efectivamente va a poder decirle a la sociedad: *estamos evolucionando, estamos mejorando o no el proceso educativo.*

Quisiera referirme ahora al aspecto del uso. Manejar un indicador puede ser una bomba, sobre todo en estos tiempos, donde hay mayor difusión, apertura y conciencia de la sociedad por saber qué está pasando. No en balde se ha hecho énfasis en el sector público en hablar sobre una política de apertura, revisión de cuentas y

---

\*Coordinador Sectorial de Planeación e Información de la Dirección General de Planeación, Programación y Presupuesto de la SEP.



transparencia; eso nos ha obligado a abrir las puertas de la información a la sociedad.

A nosotros, prácticamente toda la información que tenemos ha sido demandada por la sociedad a través del IFAI, y muchas veces la gente no sabe interpretar esos indicadores, y si no los sabe interpretar y los usa para tomar decisiones o para difundir ciertas acciones, vienen las implicaciones. Entonces, tenemos una responsabilidad muy importante quienes los construimos y difundimos. Por ello, es necesario que sean claros, concretos, y ubiquen bien a quienes no están inmersos en el tema.

Hablar de un indicador de cobertura de la educación primaria, no significa hablar de cómo está la calidad de la educación simplemente, el indicador de cobertura es un elemento para considerar cuántos recursos se necesitan asignar a la educación primaria, y ésta cumpla con el objetivo de proporcionar educación a todos los mexicanos.

El interpretar va asociado también a cruzar información de diferentes sistemas, debemos tener mucho cuidado cuando cruzamos información de un sector o de un sistema con otro. Un ejemplo muy claro es el gasto de educación con respecto al PIB.

### **El ejemplo de la OCDE**

La OCDE utiliza criterios diferentes a los que ustedes se van a encontrar y aparecen en el Informe de Gobierno. Éstos son detalles que pueden ayudar a interpretar mejor para tomar mejores decisiones.

Afortunadamente, conforme transcurre el tiempo, se proporciona mayor atención al uso de esta información para tomar decisiones en varios aspectos de la administración.

El primero, es el punto de vista de planeación. En el Programa Nacional de Educación se plantean metas muy concretas en función de indicadores específicos. Se está empezando a asignar recursos a partir de fórmulas de financiamientos, en las cuales se incluyen aspectos de cobertura y eficiencia del sistema educativo, basados precisamente en una serie de indicadores.

Y el otro aspecto es el de espacios. Nosotros vemos en la mayor parte de los documentos y en la prensa, indicadores que son de orden nacional; sin embargo, vivimos en un país de fuertes contrastes en algunos aspectos, un ejemplo es Chiapas y el DF, donde los indicadores de analfabetismo son de un contraste significativo.

Los aspectos descritos deben ser considerados por quienes hacemos uso de los indicadores. Y creo que, en este sentido, el Instituto enfrenta un reto muy grande, y seguramente va a dar respuesta a uno de los reclamos más importantes de la sociedad, el cual es, ¿con qué calidad estamos trabajando? Y es una pregunta a la cual difícilmente se le podría dar respuesta, pero con los esfuerzos que se están realizando, seguramente lo vamos a lograr.

La gran ventaja de informar a la sociedad es lograr un círculo virtuoso, es decir, en la medida que la sociedad entienda mejor cómo estamos trabajando, nos ayudará con la tarea, la cual no tan fácilmente el sector educativo o el Estado puede afrontar.



La publicación de *Panorama Educativo de México 2003* me parece realmente un trabajo muy fino, el cual no tiene precedentes en el Sistema Educativo Nacional.

Hay por ahí documentos parecidos, creo que el INEGI produjo en algún momento algo así como el ABC de las estadísticas educativas, pero conceptualmente me parece que este documento es superior, y cuenta con la información técnica que mucha gente necesita para analizar la información proveniente de los indicadores y así poder realizar investigación.

Es un muy buen prontuario de la información que se puede obtener de las fuentes con las cuales cuenta ahora el país en materia educativa.

Se me ocurre que habría muchos más indicadores, creo que el libro es muy bueno en su propia autocrítica, al presentar algunos indicadores que por supuesto son importantes y los cuales han sido propuestos por investigadores diversos, me imagino que el INEE trabajará eventualmente en ellos.

Por supuesto, también faltan indicadores de corte cualitativo, pero el mismo libro explica las razones por las cuales no aparecen, me imagino que se irá resolviendo esto a corto o mediano plazos.

### El estado actual

La primera: creo que el estado en el cual se encuentran los indicadores en este momento es muy propio de los sistemas evaluativos de países desarrollados y en vías de desarrollo en sus fases intermedias; es

en las fases originales donde resulta difícil saber en qué información confiar y en cuál no. No debemos nunca olvidar lo difícil que es saber cómo se construyen los indicadores y muchísimo más difícil todavía es conocer cómo interpretarlos, y así poder desarrollar investigación a partir de ellos.

En ese sentido el INEE ha dado un paso gigantesco en apenas dos años de vida. Sin embargo, viendo hacia el futuro, el Instituto debería buscar otras vías de desarrollo de indicadores. Efectivamente, es muy difícil pensar que en algún momento pueda tener influencia en la elaboración del censo, pero a lo mejor algunas preguntas sí podrían proponerse en el cuestionario ampliado, probablemente lo mismo pueda suceder con muchas otras fuentes.

Llama la atención que pocas de las fuentes encontradas en el documento de indicadores sean netamente educativas. Esto es muy importante, porque a fin de cuentas es la información objeto directo de la política pública.

Es cierto que toda la información contextual es muy importante, se vislumbran temas como el PIB *per cápita*, niveles de pobreza, marginación, etcétera, los cuales pueden tener una influencia en la educación. Por supuesto, como país debemos de trabajar en mejorar el contexto en el cual educamos a nuestros niños; sin embargo, hay otros indicadores también muy importantes, de corte netamente educativo, los cuales sí caen dentro del trabajo que realiza la SEP, y por tanto, pueden ser cambiados con política pública pertinente. Así funcionan países como Inglaterra, Taiwán o Estados Unidos, donde se comi-

\*Candidato a doctor del Programa Políticas Educativas y Sociales de la Escuela de Educación de la Universidad de Harvard.



sionan estudios anuales o bianuales sobre temas diversos.

Interesan temas para México como: satisfacción respecto a la educación de padres y alumnos, encuestas sobre aceptación de valores interculturales en las escuelas, inmigración, niños que van y regresan de Estados Unidos constantemente y sobre los que tenemos prácticamente nula información, discriminación y violencia en las escuelas, participación de padres, etcétera.

Considero que si el INEE tomara algunos de estos estudios en sus manos, y comisionara otros, los investigadores podrían desarrollar indicadores con probabilidades de ser incorporados al acervo de indicadores permanentes del Instituto. Así lo hacen muchos organismos, los cuales producen informes anuales. Por ejemplo, el Informe sobre Desarrollo Humano, cada año presenta el estado de sus indicadores, e introduce indicadores nuevos que provienen de estudios monográficos sobre temas diversos. Algunos de estos indicadores permanecen en el documento, otros no, pero de cualquier manera sirvió, pues se introdujo un tema importante de manera monográfica en el sistema educativo y algunos indicadores pueden incorporarse en un documento definitivo. Así funciona la OCDE con su documento anual sobre indicadores y muchos otros organismos.

### **Indicadores de corte transversal**

Una segunda línea de comentarios va en el sentido de que la gran mayoría de indicadores encontrados aquí son de corte transversal; esto es entendible por el género y fuentes de información de donde provienen. Habría que considerar un buen sistema de información longitudinal; esa es probablemente la cereza del pastel de muchos sistemas de evaluación. Lo es, por ejemplo, en Estados Unidos; ellos tienen varios es-

tudios de carácter longitudinal y siguen a estudiantes por largas temporadas. Por supuesto, supone muchos retos técnicos y es muy caro, es obvio que no podría ser de carácter censal pero con una buena muestra sería suficiente. La información obtenida es de una riqueza técnica sin parangón, de la cual se carece actualmente.

No estoy pensando que el INEE pueda llevar esto a cabo en el corto plazo, pero probablemente para los siguientes dos o tres años, pudiera elaborar algún proyecto en esa línea; sería un salto cuántico con respecto a la información que tenemos.

Este tipo de investigaciones longitudinales tiene la ventaja de que, como dije, sigue a los alumnos. Por ejemplo, en el estudio del *National Educational* de Estados Unidos se sigue a los estudiantes por 14 años, egresen o no de la escuela. Esa información ayuda a entender las causas de la deserción, la influencia del mercado laboral, la influencia de la situación familiar y otros muchos aspectos.

Pasemos a los indicadores que cambian en el tiempo y los que no lo hacen. Por ejemplo, un indicador que no cambia en el tiempo es el género; uno que sí lo hace es el apoyo de la familia a los alumnos, esto permite entender los momentos claves y críticos del sistema educativo.

En fin, estos estudios hacen entrevistas a padres, a alumnos, éste que mencioné particularmente me gusta mucho porque sigue a los alumnos incluso hasta el momento en que van al campo laboral, pueden medir de manera muy interesante algunos de los resultados finales que esperamos del sistema educativo, se evalúan características de las comunidades, aspiraciones educativas y laborales, experiencias escolares, trabajo, experiencias familiares, recursos educativos de las escuelas, de las familias, incluso consumo de alcohol, droga, primeros encuentros sexuales y otras tantas actividades extracurriculares.

Esto es en general una fuente de información muy rica, por supuesto significaría tener un seguimiento por alumno, una clave única de alumno. Poder obtener los recursos, esa es la parte más difícil de los estudios longitudinales, para dar seguimiento a los alumnos por tantos años es un reto, pero creo que al menos el INEE podría pensar con anticipación en un estudio de este tipo.

### **Cómo evaluar a los indicadores**

Una tercera línea de mis comentarios trata de cómo evaluar a los mismos indicadores, por supuesto éstos forman parte del sistema de evaluación, pero también debemos de estar planteándonos constantemente cómo evaluarlos, la UNESCO sugiere una serie de criterios, me imagino que muchos de ellos fueron tomados en cuenta en el proceso de jueceo que nos señala el libro, los cuales deben de ser medidos directamente: objetivo adecuado, desagregado, práctico, confiable, replicable, etcétera.

Me quiero detener probablemente en el tema de la validez. Los instrumentos

educativos, de información y por supuesto los indicadores, no son en estricto sentido válidos o inválidos. Lo válido o inválido es la inferencia que nosotros realizamos de la información que ellos producen y desde esa perspectiva, es factible obtener algunas inferencias inválidas de ciertos indicadores que son válidos para otras cosas.

Me preocupa por ejemplo, la invalidez que pueden tener inferencias de la información que aquí se presenta con respecto a poblaciones, por ejemplo muy diversas; tenemos un examen que se aplica de español y matemáticas, el cual está en el programa de estándares aplicados por igual a la población nacional y a la población indígena. No lo he estudiado directamente, pero la investigación educativa iría en ese sentido, pensar que el instrumento está sesgado, es difícil evaluar a un indígena sobre todo a temprana edad, por el estado de nuestro sistema educativo en este momento, y obtener información válida y confiable sobre muchos temas, especialmente sobre su desempeño en matemáticas, en español; por tanto, deberíamos pensar en crear indicadores culturalmente sensibles.



## **PREGUNTAS Y RESPUESTAS**

### **PATRICIA MUÑIZ MARTELON**

#### **¿Por qué no considerar y calificar las habilidades docentes como un posible indicador?**

¿Por qué no? Podría ser algo a considerarse, tanto como calificarlas en el sentido de calificación, no, pero sí podría en algún momento hacerse una especie de catálogo de habilidades a desarrollar y poderlas trabajar así.

#### **¿La publicación y la transparencia de resultados no promoverá mayor inequidad en el propio sistema?**

No pensaría que así fuera, de alguna manera lo que hemos estado repitiendo es que ésta no es una evaluación para sancionar a alguien, sino para diagnosticar y evaluar en términos globales, al menos en lo que respecta a la parte de indicadores, estamos trabajando con estadísticas, no casos en particular, la estadística no per-



mite sancionar porque no vamos a hablar ni de individuos en particular ni de escuelas, sino de un sistema educativo en general, con sus altas y sus bajas.

### **¿Quién evalúa al INEE en particular?**

Dentro del INEE nos evaluamos todos en nuestra tarea cotidiana, y yo pensaría que la sociedad puede evaluar al INEE; en general los académicos, la Junta Directiva, el Consejo Técnico; somos sujetos a ser evaluados nosotros y nuestras tareas por quien lo quiera hacer.

### **¿Y al sistema educativo en general?**

Eso es lo que estamos intentando hacer desde diversas perspectivas, evaluar al sistema educativo en general.

### **¿Por qué en las zonas indígenas y en las zonas marginales falta calidad en la educación? ¿Qué proponen?**

Nosotros no vamos a mejorar la educación como tal, vamos a evaluarla. Creo que ésta es una misión mucho más amplia y en realidad requiere una reflexión y un conocimiento mucho más profundo del que yo tengo, como para hacer una propuesta así a quemarropa de la calidad de la educación, de la falta de calidad en la educación en las zonas indígenas y marginales.

## **TERESA BRACHO GONZÁLEZ**

### **¿Cuáles serían algunas estrategias para lograr que los alumnos terminaran la educación básica, sobre todo en las zonas rurales?**

Creo que las preguntas están más orientadas a la preocupación que todos tenemos de cómo mejorar al sistema educativo, pero nos queda muy claro que el INEE no es la instancia responsable de la mejora del sistema educativo, no es el INEE quien lleva a cabo la tarea educativa. El INEE lo que hace es evaluarla, eva-

luar los resultados, evaluar el desempeño del sistema educativo y por esta vía tratar de identificar problemas nuevos, tratar incluso de identificar aspectos que se dan por sentados. De repente hay escuelas indígenas que son muy buenas, entonces hay que saber qué pasa en las escuelas indígenas que son buenas y replicarlo, para quitar esas imágenes de que las cosas son blancas o negras o estamos reprobados o aprobados.

Creo que el problema es mucho más complicado, mucho más complejo, y sí diría que no es función del INEE decir lo que debe hacerse. El INEE mostrará lo que revela la evaluación, cuál es el estado de un objeto y señala, cuando mucho, posibles alternativas o caminos que se derivan directamente de la evaluación, pero no es la instancia para tomar decisiones sobre el sistema educativo.

### **Por sus comentarios, ¿no se habría hecho nada antes o lo que se hizo era mentira?**

No, al contrario, yo no digo que todo lo que se haya dicho y hecho fuera mentira; lo que digo es que la información no era abierta, no era clara y no era siempre explícita, incluso para expertos, sobre cómo se llegaba a esa información. Muchas veces uno encontraba un cuadro publicado y no podía distinguir de dónde había salido esa información, cómo se había generado, cómo se había transformado. Como investigador era muy difícil trabajar esa información y hacerla compatible con otras informaciones. Y lo que señalé es lo que hace este libro, nunca dije que lo otro era mentira, es simplemente explicitar cuáles son las condiciones de producción de cada uno de los indicadores, lo que permite: primero, su conocimiento; segundo, su entendimiento; y tercero, ser sometido a crítica directa. Es decir: a mí no me parece esta composición del indicador por tal

y tal y tal. Si no tengo esa composición no puedo hacer una crítica del indicador.

**¿Por qué no se pregunta directamente a los maestros cómo le hacen, en vez de adivinar cómo obtienen alumnos exitosos?**

Claro, hay mucha investigación por hacer, lo que está diciendo el INEE es no sólo contar con la información ya generada, sino generar mejor y más información, pero esto no es un proceso de un día para otro, hay que identificar claramente lo que se quiere buscar y cómo se va a buscar.

**GERMAN TREVIÑO GONZÁLEZ**

**¿Cuál es la diferencia entre un sistema de indicadores y simplemente un conjunto o grupo de indicadores?**

El libro que tenemos entre manos es un sistema de indicadores; es decir, la información está conceptualmente ordenada, es replicable, explica claramente cómo se puede interpretar, las inferencias que de ella se pueden sacar, hay una coherencia interna y me parece que tenemos un sistema de indicadores, el cual se puede ir mejorando, por supuesto, a partir de muchas de las sugerencias de ustedes mismos y de metas que se ha impuesto el INEE para el futuro.

**¿Los indicadores que pudieran ser producidos a partir de estudios realizados por distintos investigadores, podrían tener un buen empleo, o si no, tendrían de alguna manera un sesgo?**

Esta pregunta es importante, porque debemos mantener en la mente que no es solamente un estudio el cual produce indicadores, sino un conjunto de estudios; es decir, todos estos indicadores se validan de manera diferente, tienen que ser coherentes entre sí. Entonces, de alguna manera se protege uno de que una visión particular prevalezca en la interpretación de los indicadores.

**¿Los docentes conocen estos indicadores?**

Me parece que la disponibilidad de esta obra está al alcance de todos, pero creo que sí hace falta y será probablemente trabajo del INEE para los futuros años crear indicadores, los cuales incidan directamente con la práctica educativa impartida en el aula.

**ANDRÉS ORTIZ BRIZUELA**

El fondo de las preguntas que me llegaron están dirigidas a fomentar la cultura de la evaluación:

**¿Cómo detectar si hay sinergia entre las diferentes instituciones para cruzar información?**

**¿Cómo les va a llegar esto a los directores para que también compartan lo que ellos mismos generan?**

Un punto importante es reflejo de la inquietud que en este salón hay: parte de los buenos resultados va a depender de la cultura de la evaluación en la que todos estemos inmersos. Es algo que no ha existido y ha hecho mucha falta, eso va a implicar acudir a algunos seminarios para aprender a construir indicadores. La pregunta que se hace al respecto es:

**¿Dónde puedo aprender a hacer indicadores y a evaluar?**

El Instituto tiene su página Web, a ellos se pueden acercar y seguramente les orientarán sobre ese tipo de temas. La inquietud aquí planteada es muestra de que se está empezando a adquirir una cultura de la evaluación. Antes no evaluábamos, simplemente medíamos, monitoreábamos, que es muy diferente al concepto de evaluación. Creo, por tanto, que ese tipo de eventos y actividades van a ayudar a este gran proceso.





Aspecto general de los panelistas en la mesa sobre Indicadores del INEE: *Panorama Educativo 2003*.



Es función del INEE desarrollar y mantener un sistema nacional de indicadores educativos.

CUARTA MESA REDONDA

## **Comparabilidad de los Resultados de las Evaluaciones**

**Ponente:**

Lic. Felipe Martínez Rizo

**Comentaristas:**

Dr. José Manuel Álvarez Manilla

Dr. Ángel Díaz Barriga

Dr. José Luis Gaviria Soto

Dr. Manuel Gil Antón

Lic. Jorge Hernández Uralde

Dr. Arturo de la Orden Hoz

**Moderador:**

Mtra. Margarita Peón Zapata





## Comparabilidad de los Resultados de las Evaluaciones

Por FELIPE MARTÍNEZ RIZO \*

La comparabilidad de los resultados se encuentra entre los temas que han suscitado mayor polémica en épocas recientes; esto es perfectamente entendible, pues la sociedad espera que la evaluación informe sobre la evolución en el tiempo de la calidad educativa, saber si ésta retrocede o avanza; si es buena, regular o mala en un momento determinado.

Es claro que necesitamos comparar en el tiempo, de lo contrario no tendría mucho sentido hacer una sola evaluación y no darle seguimiento. Saber si la calidad avanza o retrocede, demanda evaluaciones estrictamente comparables y esto se dice fácil pero no es sencillo, es muy complejo.

Desde su Primer Informe Anual, el Instituto advirtió que los resultados del año 2003 no son estrictamente comparables con los de años anteriores; resulta lamentable, pero es importante reconocerlo y obviamente necesitamos sustentar de una manera sólida tal opinión, ese es el propósito de este trabajo.

La forma más sencilla de obtener resultados comparables, sería aplicar la misma prueba en ocasiones sucesivas, pero entonces es fácil que algunos alumnos tengan información sobre las preguntas y sus resultados se vean favorecidos; por lo tanto, se perdería la confiabilidad del proceso.

Para evitar ese riesgo, los sistemas de evaluación deben construir instrumentos que por una parte sean estrictamente

comparables, pero diferentes y esto como dije antes, es fácil de decir pero no de lograr.

### Inicio de las pruebas comparables

El desarrollo de la psicometría a lo largo del siglo XX es un largo esfuerzo por desarrollar instrumentos, cuyos resultados además de válidos y confiables, sean comparables entre sí, aunque se obtengan en distinto lugar y momento.

A ello apuntó en la primera mitad del siglo XX la invención, por ejemplo, de las preguntas de opción múltiple, las preguntas que se hacían normalmente para saber si un alumno había aprendido algo, eran lo que decimos los maestros preguntas de ensayo; el inventar preguntas cerradas de opción múltiple apareció en 1926.

Las técnicas de escalamiento, de estandarización, la teoría clásica de las pruebas, los desarrollos de la teoría de respuesta al reactivo, de *respuesta al ítem tri* –no es el *tri* que nos gusta–; en la segunda mitad del siglo XX fueron esfuerzos en esa dirección.

El *Collage Board*, desde 1941 comenzó a buscar técnicas específicas para asegurar la comparabilidad de los resultados de la prueba más antigua de este tipo, más famosa: el SAT, el *Scholastic Aptitud Test*, y desde 1959, el *American Collage Testing*, hizo también lo mismo.

Las técnicas para contar con resultados estrictamente comparables se designan

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



con la palabra *equating* que en español se suele traducir como equiparación, pero también se dice igualación, y el hecho mismo de que no tengamos una palabra estandarizada entendida por todos, aceptada por todos para designar esto, nos muestra la novedad del asunto. En nuestro medio todavía no es usual que nos preocupemos por ello.

Doy otro dato curioso, Inglaterra aplica pruebas estandarizadas desde 1940, pero todavía hoy están discutiendo el tema, y se discute de manera persistente la pregunta: ¿en qué medida las tendencias del rendimiento de los asuntos que registran las estadísticas son tendencias genuinas o en qué medida reflejan, por ejemplo, el impacto de cambio en los contenidos curriculares, los métodos de evaluación, o que los maestros enfoquen su enseñanza en preparar a los alumnos para las pruebas?

Se han desarrollado varios estudios de los estándares a lo largo del tiempo, pero ninguno ha sido concluyente para acabar con el debate. Todavía en este momento, en el año 2004, no está resuelta la discusión en Inglaterra de si las tendencias que muestran los resultados de las pruebas que vienen aplicando desde hace más de medio siglo, realmente sean comparables, de tal manera que se pueda estar seguro de la medida en que ha habido o no avances, o eventualmente retrocesos en el sistema.

En el tema de la equiparación, por otro lado, tomamos otro referente. En los textos que contienen estándares de calidad para los instrumentos de evaluación y demás que se vienen publicando en los Estados Unidos, quien es el país que desarrolló esto originalmente, no aparecen estándares sobre equiparación en los documentos de 1954, 1955, 1966 ni 1974, aparecen por primera vez en 1985; antes ni siquiera se ponían estándares de calidad

para equiparaciones, y en 1999 vuelven a aparecer.

Por otro lado, el sistema de evaluación educativa más reconocido de América Latina, el sistema chileno, comenzó a desarrollarse en 1988, pero apenas en 1998 empezó a introducir las técnicas de teorías de respuesta al ítem y con base en ellas, técnicas modernas de equiparación.

Quitando a Chile y Brasil, al parecer ningún país de Hispanoamérica cuenta con series de resultados comparables en escala nacional. Uruguay que es pequeño, ya lo está haciendo; Argentina empezó a hacerlo y luego se detuvo. En fin, muchos países europeos tampoco las tienen porque no es algo ni sencillo ni habitual, es algo muy complicado y novedoso.

De manera general, puede decirse que hay distintas metodologías donde se distinguen lo que son diseños para equiparación y métodos estadísticos para equiparación. Por un lado, hay diseños con un grupo o grupos equivalentes sin pruebas ancla, o diseños con grupos no equivalentes con pruebas ancla o ítems ancla.

También hay métodos estadísticos que permiten hacer la equiparación combinándose de distintas maneras con los diversos tipos de diseño, métodos lineales, métodos basados en la *teoría de respuesta al ítem*; todos ellos tienen múltiples variantes, pueden utilizarse en combinación con los distintos diseños, usarse con técnicas de suavizado previo o posterior, en fin, hay gran cantidad de variantes. No hay un método absolutamente superior a los otros, cada uno puede ser más o menos adecuado en determinadas circunstancias.

### La lección que ofrece el NAEP

Veamos lo que llamo "la lección de la anomalía de lectura de las Pruebas NAEP de 1986". Creo que el sistema americano que se llama *National Assessment of Edu-*

*ational Progress*, NAEP, tiene la serie de resultados comparables más importante que hay en cualquier sistema educativo del mundo a partir de 1969.

Sin embargo, los resultados de las pruebas de lectura del NAEP de 1986, mostraban un decremento sorprendente respecto a los resultados de 1984 en el rendimiento de los alumnos de 17 años, y en menor grado en el rendimiento de alumnos de nueve años.

Se consideró extremadamente improbable que cambios de tal magnitud en la competencia lectora hubieran ocurrido en sólo dos años, sin que la comunidad educativa tuviera conciencia de ello.

Es más fácil que uno se equivoque en algunos de los múltiples pasos para construir una prueba adecuada y aplicarla, a que el sistema educativo de un país grande se mueva drásticamente en un plazo corto, sea para arriba, sea para abajo.

El *Educational Testing Service* decidió revisar los resultados verificando en particular la calidad de las equiparaciones, el resultado detectó diferencias en las aplicaciones de 1986, respecto a las de 1984 en el manejo de los ítems comunes a las dos formas de la prueba; en 1986, además de secciones para competencia lectora, la prueba incluía matemáticas o ciencias para alumnos de nueve años y computación, historia o literatura para los de 17.

Los ítems ancla aparecían en distinto orden y el tiempo disponible para responder fue diferente. Con ello, concluyeron que los cambios observados en la construcción de los instrumentos y las condiciones de aplicación, explicaban la diferencia de los resultados más que el nivel de competencia de los alumnos.

La lección de la anomalía de lectura del NAEP de 1986, lleva a la conclusión que ya adelanté: la fuerte inercia de un sistema de educación grande, hace poco probable que en lapsos cortos ocurran cambios fuertes,

sean ascendentes o descendentes; tales cambios se deberán probablemente a defectos de las equiparaciones que deberán revisarse.

### **Criterios generales para efectuar buenas equiparaciones**

Paso en seguida a mencionar los criterios que deberían de seguirse para realizar buenas equiparaciones, pues no basta utilizar diseños de investigación y métodos estadísticos adecuados, se tienen que cumplir otros criterios.

Éstos se sistematizan fundamentalmente en cuatro puntos, que han sido manejados desde 1970 por uno de los autores más importantes en este campo y luego se han retomado habitualmente. El primero se refiere al constructo que se mide, es decir, las pruebas deben ser diseñadas para medir la misma característica: destreza, rasgo latente, lo cual es muy claro.

Un segundo punto, es que debe haber una invarianza en la población: la elección de la subpoblación que se utilice para calcular la función de equiparación no debe afectar los resultados.

Un tercer rasgo es la simetría: el proceso debe ser reversible, si equiparamos de una forma hacia otra o al revés no debe de cambiar la situación.

Y el cuarto, la condición de equidad: las puntuaciones deben ser totalmente intercambiables, sea cual sea el nivel de habilidad de los sujetos.

Esto suena muy técnico, y es muy técnico; si dicen "no lo entendí" no me sorprende, me sorprendería que lo hubieran entendido, pues es muy complicado, son cuestiones sumamente especializadas.

Como decía antes, también podemos tener equiparaciones para seguir a los alumnos, sea para comparar alumnos de grados distintos en un mismo año o para seguir longitudinalmente una cohorte a lo largo del tiempo.



## Requisitos específicos para desarrollar una buena equiparación

Podemos hablar de un conjunto de requisitos para hacer una buena equiparación, los cuatro anteriores son requisitos generales o cuatro condiciones básicas, pero para cumplir esas cuatro condiciones hay que cubrir otros requisitos puntuales:

1. Es indispensable que los contenidos y las especificaciones se encuentren bien definidos.
2. No debe haber cambios de los referentes curriculares en los tiempos que se están evaluando.
3. Definir previamente el modelo a utilizarse, si se trata de una prueba con referencia a una norma estadística, o un criterio.
4. Elegir el diseño metodológico y el método estadístico para las equiparaciones, que como se planteó son muchos, debiendo escogerse el más apropiado.
5. La extensión de la prueba no puede ser demasiado chica, por lo cual se requiere un número suficiente de reactivos en cada dominio a evaluar, para lograr un grado adecuado de confiabilidad.
6. Las formas a equiparar deben ser de la misma extensión aunque tengan un grado de confiabilidad similar; si tenemos una prueba corta y una prueba larga nos va a afectar porque ya no son las mismas condiciones para que los alumnos respondan.
7. Además de tener índices de confiabilidad similares, los ítems deben haberse piloteado previamente y contar con estadísticos para armar las pruebas, fruto de aplicaciones piloto previas.
8. Definir ítems comunes a las formas que se van a equiparar, representativos de la prueba. Debe haber un número suficiente de ítems ancla, pues

aunque teóricamente con uno se podría comparar, el margen de error sería enorme.

9. Los ítems ancla no pueden cambiar, deben ser iguales en las diferentes formas de la prueba.
10. Cada ítem ancla tiene que estar aproximadamente en el mismo lugar. Si tengo un conjunto de ítems ancla pero en una prueba están al principio y en otra al final, se puede afectar la comparabilidad.
11. Las características formales deben ser iguales; por ejemplo, la letra no puede ser chica en una forma y grande en la otra.
12. Los procedimientos de muestreo deben ser adecuados para asegurar la representatividad de los alumnos a los que se aplican las pruebas.
13. Hay procedimientos estandarizados de aplicación y medidas de seguridad, los cuales evitan que los sustentantes conozcan previamente las pruebas para asegurar igualdad de condiciones y ausencia de sesgos.
14. Los procedimientos se deben mantener iguales a lo largo del tiempo.
15. La aplicación debe tener lugar en momentos similares del año.
16. Los procedimientos de lectura de los resultados deben asegurar la calidad de las bases de datos.
17. Y, por último, las técnicas de calibración, escalamiento, estimación de estadísticos, y cálculo de resultados, deben ser adecuadas según los propósitos y las características de la evaluación, según los instrumentos utilizados y las muestras de alumnos.

## Los problemas de equiparación que presentan las pruebas mexicanas

Si aplicamos estos 17 criterios a nuestras pruebas, encontramos que en muchas

ocasiones no se han satisfecho de una manera adecuada. El primero sí, el primero no tuvo cambios, en parte, en los referentes de las pruebas de 1998 a 2003, pero la definición de los dominios de contenido, el mapeo curricular, no fue suficientemente preciso.

Las pruebas de *estándares* teóricamente se conciben como criterios, pero ni en su planeación, ni en el análisis de sus resultados se siguieron consistentemente los principios aplicables a este tipo de pruebas.

La elección del diseño de equiparación es adecuado, en principio, pero el método estadístico escogido no fue el más conveniente porque en este caso se trata de pruebas muy grandes, y nos interesa comparar resultados a todo lo largo de la escala de puntuaciones y los métodos de Rach distorsionan en los extremos.

### **Suficiencia de reactivos**

Hasta el año 2002, las secciones de lectura y matemáticas de las pruebas se trataron como una sola dimensión, cuando en realidad la estructura factorial de las pruebas es compleja y por lo tanto la hipótesis de que es una sola dimensión, no es plausible.

Pese a ello, aun considerando cada sesión como una prueba, en general sí se tuvo un número suficiente de reactivos con excepción de las pruebas de 1998.

Los datos del número de reactivos que hubo en cada prueba, cada año, muestran que en 1998 las pruebas de lectura tuvieron un número muy reducido de reactivos, lo cual no es adecuado; pero en los demás años sí fue un número importante y en matemáticas desde el primer año lo tuvieron.

Hay diferencias muy considerables en número; tenemos desde unas pruebas de 15, 17 reactivos, hasta pruebas de 63,

77 reactivos, por ejemplo, lo cual provoca que no se puedan comparar sencillamente unas con otras.

### **Confiabilidad de los instrumentos**

Los datos sobre grados de confiabilidad, índice de confiabilidad medido con el coeficiente alfa de Cronbach de las pruebas, muestran diferencias muy importantes que tienen que ver con el tamaño de las pruebas y otros puntos. Normalmente se considera deseable que haya confiabilidades del orden de .8, .75 y vemos que en muchos casos sí las hay, pero en otros no.

En el caso de secundaria es muy claro que tenemos confiabilidades en las pruebas de lectura, quitando las de 1998 que eran muy pequeñas, bastante buenas, de .85, .90 inclusive, pero en matemáticas, de .53, .38, .48, que son niveles inaceptables de confiabilidad para una buena prueba.

Por otro lado, los ítems no se pilotearon antes de su aplicación por lo que no se pudo contar con los estadísticos correspondientes antes de armar las pruebas; entonces podemos ver que se falló en el criterio de especificaciones claras y de pilotaje previo. Obviamente los ítems ancla no se pudieron seleccionar de manera que fueran representativos de la prueba y de las características estadísticas de ítems.

En algunos casos, hay inconsistencia en el número de ítems ancla. En 1998 no había ítems ancla, era el primer año en que se aplicaban las pruebas y todos los ítems eran nuevos. Pero en los siguientes años podemos ver, por ejemplo, en el año 1999, hay una prueba totalmente nueva, cero ítems ancla y en cambio en el año 2001 es exactamente la misma prueba de antes, cero ítems nuevos, todos los ítems comunes. Esas situaciones en las que en unos casos hay pocos, en otros muchos, en otros nada, en otros todos, dan como resultado que la comparabilidad sea muy desigual.



Los ítems ancla que se usaron en el número apropiado o inapropiado sí se manejaron idénticos; dados los cambios en el número de ítems, la ubicación de los ítems ancla varió, no siempre fue igual.

### **Características de las pruebas y procedimientos de aplicación**

Las características formales de las pruebas fueron casi iguales, hubo diferencias mínimas por algunos errores de impresión; en general fueron de buena calidad e iguales siempre. Las muestras utilizadas fueron representativas, por tanto, en principio se puede considerar que son buenas y su población es comparable, pero luego vamos a tener una salvedad con esto.

Los procedimientos de aplicación fueron consistentes, pues se aseguraron las condiciones de igualdad de circunstancias; las medidas de seguridad fueron adecuadas, obviamente con sus márgenes de error, entonces este criterio se cumple, pero no el relativo a que la aplicación tenga lugar en momentos similares en todos los años.

En unos años fue al inicio del ciclo escolar y en otros al final del ciclo; además, cuando fue al inicio del ciclo escolar, se aplicaron al grado siguiente considerándolo comparable al grado anterior, lo cual no se sostiene por la pérdida que hay en el paso de un grado a otro, especialmente de sexto de primaria a primero de secundaria, en donde la pérdida no es aleatoria, sino que precisamente los chicos que no continúan son normalmente los de bajo rendimiento y por lo tanto, no son poblaciones comparables.

La precisión de los procedimientos ópticos es buena, y este criterio se cumple, pero en lo referente a las técnicas de análisis utilizadas hay dificultades. Por un lado, como no hubo piloteo previo, no se pudo controlar la proporción de ítems de distin-

ta dificultad. Por otro lado, como hasta el año 2002 se consideraron una sola dimensión lectura y matemáticas, se manejaron en una sola escala; en el año 2003 se separaron las escalas, y tampoco hay seguridad si se utilizaron hasta 2002 los ponderadores adecuados para calcular los distintos estadísticos de la prueba.

### **Análisis de los instrumentos por parte del INEE**

Las variaciones de los resultados de las pruebas aplicadas de 1998-2003 debían hacer dudar de su consistencia y llevar a revisarlos, ya que no es probable que ocurran realmente cambios semejantes. El INEE hizo la revisión, cuya conclusión es que no se cumplió la mayoría de los criterios necesarios para que los resultados puedan compararse, incluyendo varios de los más importantes.

Las altas y bajas que se observan en los resultados no pueden atribuirse, con base en estos datos, a cambios en los niveles reales de competencia de los sustentantes; la conclusión es fuerte, pero la evidencia parece incuestionable.

Por la naturaleza de las fallas, no se pueden corregir *a posteriori* y cito a uno de los especialistas más destacados en el tema: *si hay problemas con la construcción de la prueba o con los datos, entonces ninguna cantidad de manipulaciones estadísticas puede llevar a resultados adecuados de la equiparación.*

Este análisis no significa que no haya elementos valiosos en los resultados revisados, de ninguna manera; dada la calidad de las muestras utilizadas se pueden hacer comparaciones entre entidades federativas, y entre los servicios que constituyen los estratos de la muestra en cada año, lo que no podemos seguir son las tendencias de un año a otro.

Esos son análisis que el INEE ya ha hecho en parte y difundido desde noviem-

bre del año pasado, junto con resultados de las pruebas internacionales, pues permiten también algunas modestas comparaciones en el tiempo.

Por esta razón, dado que hay algunas deficiencias importantes, el INEE emprendió desde hace varios meses el desarrollo de nuevos instrumentos de evaluación, los cuales empezaremos a aplicar el año próximo, para tener series consistentes que nos permitan medir de manera confiable la evolución del Sistema Educativo Nacional.

Por otro lado, tampoco nos limitaremos a ello, trataremos de hacer lo que se pueda con lo existente; ya en noviembre del año pasado hicimos algunas comparaciones de resultados 1995-2000 con los resultados de TIMSS; este año, en diciembre, tendremos resultados de PISA 2003, los cuales nos permitirán hacer también algunas comparaciones con PISA 2000, además de darnos resultados por estado. El año próximo estamos planeando una aplicación extraordinaria de pruebas anteriores que nos permitan hacer una comparación 2000-2005.

### **Complejidad del problema**

Es importante observar que la comparabilidad es muy compleja; es explicable la dificultad para entender el tema y que las posturas o las afirmaciones sustentadas, se comprendan de una manera parcial. El INEE desea contribuir a desarrollar la cultura de evaluación en ese sentido, pasar a un pensamiento más complejo, que trate de evitar las dicotomías simples de blanco-negro, sí o no, porque la realidad es más compleja.

Cuando se dice que la línea está cambiando de postura y ahora sí se va a comparar y antes no se quería comparar, o ahora afirman que sí darán resultados por escuela, y el mismo día salen en distintos

medios declaraciones totalmente contradictorias, es que el tema es más complejo, no es un asunto de sí o no.

Por supuesto queremos comparar y vamos a tratar de hacer algunas comparaciones inclusive hacia atrás, pero no con los resultados que existen; esos ya no los podemos salvar por más manipulaciones estadísticas que hiciéramos, sino con una aplicación nueva de una prueba vieja que nos permita comparar del año próximo hacia atrás; pero con las aplicaciones que teníamos ya no se puede, sólo con una nueva aplicación.

Entonces, no ha habido contradicción en la postura del INEE ni modificación de criterio; ha habido la expresión de una situación compleja, que da lugar a que se entiendan y destaquen más unos aspectos que otros.

Es importante ver que de estos temas tan complejos y novedosos, es normal que no tengamos un consenso completo, pero las diferencias no deberían ocultar las grandes coincidencias que tenemos. Hay un enorme terreno de coincidencia: a todos nos interesa evaluar bien, que se conozcan resultados por escuela, pero no siempre se puede; a todos nos interesa que se compare en el tiempo, pero eso implica ciertas exigencias.

El que maticemos o enfatizamos más un ángulo de la cuestión que otro, no debería hacer perder de vista el gran terreno común que tenemos. Con esta preocupación termino: para que el Sistema Educativo Nacional pueda contar con evaluaciones que informen de manera válida, confiable y consistente sobre el nivel de aprendizaje alcanzado por los alumnos y su avance o retroceso, necesitamos referentes firmes, cuya construcción implica que se respeten las exigencias técnicas correspondientes, no es cosa de buena voluntad nada más.

La complejidad de las técnicas psicométricas contemporáneas, hace que es-



fuerzos considerables, bien intencionados, pero con deficiencias, no ofrezcan resultados de calidad suficiente para que sean útiles; por eso necesitamos tener, en ciertos casos, evaluaciones muestrales que nos permitan controlar con gran rigor todas las variantes que hay en esto y, por otro lado, evaluaciones masivas que permitan dar resultados por escuela, aunque no se puedan controlar de manera tan rigurosa todos los elementos.

La experiencia de los países que han avanzado más en este terreno, muestra también que el desarrollo de mediciones comparables implica esfuerzos importantes y periodos considerables.

Como vimos, el NAEP se empezó a planear en 1963, se creó en 1969, pasó al ETS; en 1986 detectaron que andaban mal en las equiparaciones, incorporaron cambios y ahora tienen un sistema de gran calidad.

El INEE está convencido de la importancia de que sus evaluaciones permitan a la sociedad mexicana saber cuál es la situación real del Sistema Educativo Nacional, y si esa calidad avanza o retrocede.

El Instituto, termino, está haciendo lo necesario para proporcionar tal tipo de información, sólo necesita que se le conceda el tiempo indispensable para una tarea delicada y compleja.

---

## PARTICIPACIÓN DE LOS COMENTARISTAS

JOSÉ MANUEL ÁLVAREZ MANILLA\*

---

**E**n el comentario escrito que hice al documento formulado por el Director General del Instituto, hacía yo mis observaciones respecto a dicho documento, que me parece excelentemente formulado como punto de vista metodológico y teórico para crear un marco de lo que pueden ser las comparaciones, no solamente en series de tiempo, sino en las comparaciones horizontales con formas diversas, de lo cual me permití felicitarlo porque considero que fue una magnífica elaboración.

Ahora bien, viendo el documento, el cual espero en algún momento tengan ustedes la oportunidad de revisar, contempla dos cosas: en primer lugar, como mencionaba, el enfoque teórico-metodológico; en segundo lugar, una buena des-

cripción, un buen análisis de por qué no se pueden comparar los resultados de pruebas de épocas anteriores, en lo cual no voy a redundar.

Sin embargo, creo que hay otro aspecto que es importante tocar. Hace un momento, y el día de ayer se ha mencionado repetidamente la cuestión de la cultura de la evaluación; en neurología se usa un término que es afasia, cuando el sujeto, por algún accidente vascular o traumático, pierde la capacidad de hablar.

Los lingüistas han descrito un fenómeno que se llama afasia semántica, en la cual los términos, por ser usados repetidamente, dejan de decir lo que querían decir inicialmente. Éste es el caso de *cultura de la evaluación*.

---

\*Presidente del Instituto de Evaluación en Gran Escala y miembro del Consejo Técnico del Instituto Nacional para la Evaluación de la Educación.

## **Cultura de la evaluación**

Por un lado, vemos los apuntes que hacía el licenciado Martínez Rizo respecto a la gran complejidad del área de equiparación, que podemos catalogar en, primer lugar, como un área de alta tecnología, un área sumamente compleja, muy especializada, y yo diría, sin temor a equivocarme, una de las áreas de frontera de la psicometría.

Esto significa que se necesita personal muy especializado. Como ayer hacía referencia el propio secretario y el doctor Salvador Malo, nosotros podemos contar los psicometristas que existen en el país con los dedos de las manos, y a lo mejor nos sobran.

En cambio, si ustedes ven una institución, como el *Educational Testing Service*, es una comunidad en la cual dominan los psicometristas. Entonces, tenemos situaciones muy dispares que de alguna forma nos hacen ver la necesidad que tenemos de personal calificado.

En un segundo nivel intermedio, estarían los interesados en el terreno de la evaluación, que sin llegar a saber cómo hacer, tendrían que saber qué hace por ejemplo la equiparación, sin necesidad de manejar las técnicas estadísticas depuradas, todos los conceptos de la psicometría; deberían tener una idea de qué hace la equiparación.

Por último viene el público en general, entre otros la prensa, donde es pertinente preguntar qué tendrían que saber sobre este aspecto; por ejemplo, por qué es comparable o no un resultado.

Entonces, aquí es donde vale la pena hacer intervenir además de los elementos propios –metodológicos, teóricos– los de desempeño, que serán los factores internos, los cuales nos van a permitir o nos van a impedir la comparabilidad de resultados.

## **Problemas de la comparabilidad de los resultados**

En primer lugar, ayer se hacía referencia a que es posible lograr la independencia de un organismo evaluador, sin que necesariamente haya un status jurídico que lo cree como una entidad totalmente autónoma, y se ponían ejemplos de esto. Sin embargo, hay situaciones en las cuales la creación de opinión pública informada, que de lugar a una cultura, a una base de cultura de la evaluación, tendría que tomar en cuenta algunos aspectos.

Por ejemplo, si analizamos los datos que nos presentaba el licenciado Martínez Rizo, en primer lugar hubo defectos técnicos, por ejemplo la falta de pilotaje; esto quiere decir que necesariamente tendría que crearse y mantenerse el capital humano de psicometristas, lo cual sería una virtud que debería tener el INEE.

Vamos a ver qué pasa en la maroma del sexenio, si un presidente lo creó, otro puede descrearlo; si un presidente designó un director, el presidente puede designar otro director. Eso puede deshacer todo un equipo que llegó, hubo necesidad de formar y desarrollar su capacidad.

Nada más la comparabilidad en series de tiempo exigiría alguna estabilidad, necesitaría periodos donde haya los mismos intereses, en el supuesto caso de que se sustituyera al equipo técnico.

Es una parte que me parece trascendente de considerar, no es solamente la necesidad de que el evaluador, como decían ayer, si no es independiente no es evaluador; no, no es solamente esa independencia, sino las circunstancias en las cuales se va a dar ese tránsito que hace necesario insistir y convencer a la opinión pública, de que es indispensable asegurar la continuidad de los esfuerzos.

En segundo lugar, ya lo hemos visto en muchas ocasiones anteriores y se apunta



en el trabajo de Felipe Martínez Rizo, hay otro elemento muy importante: bases de datos incompletas, lo cual es la garantía inmediata de que no habrá comparabilidad. Aquí la opinión pública debería tener un elemento muy importante, que sería la convicción de que una base de datos educativa es tan similar a un bien como lo es una silla, como lo es una computadora y que debe exigirse responsabilidad y debe castigarse por la irresponsabilidad cuando esto no se transmite adecuadamente.

Otro argumento sería, digamos, el bajarse del caballo a la mitad del río, como sucedió en el TIMSS. Una vez que se inicia el estudio debe concluirse, pues en ese caso

se destruyeron bases de datos, y debió hacerse una labor muy importante para tratar de encontrar alguna coherencia.

Entonces, creo que la parte importante no es solamente la metodológica, la cual nos permite saber cuándo algo está bien o mal hecho; y respondiendo a la pregunta que se hacía hace rato de quién evalúa al Instituto, pues todos, nada más que dispongamos de los documentos adecuados para evaluarlo, pero la parte importante radica en la creación de una opinión pública de los diversos niveles, en el sentido de las condiciones externas que garantizan que se pueda hacer la comparabilidad.



## ÁNGEL DÍAZ BARRIGA\*

**M**e da mucho gusto que se haga un evento como éste, pues hace mucha falta en el país la discusión al respecto. Aunque traigo mi documento escrito, al leerlo y obviamente escuchar su exposición me daban ganas de decir: no hay nada que añadir.

Sin embargo, hay preguntas y una que me hago es: ¿por qué hasta ahora se muestran datos como los que nos presentan? No sólo sobre estos datos, en general creo que el error de la evaluación en México, el error de las prácticas de evaluación es que siguen siendo prácticas ocultas, secretas, que se conocen poco.

La evaluación educativa en su complejidad se ha desarrollado de una forma muy reciente en el país. Muchos de nosotros, en nuestra formación profesional, estudiamos la evaluación del aprendizaje y la teoría clásica del test, la elaboración de prue-

bas objetivas, su procesamiento, etcétera.

Hoy estamos muy lejos de aquella época. Quizá la modalidad compleja de evaluación que conocemos tenga apenas unos 14 años en México.

Mucho se ha hablado en este periodo de generar una cultura de la evaluación, más allá de lo incómodo de esta denominación para cualquiera que respete el pensamiento clásico de las humanidades. Varios humanistas brincarían del susto al ver los usos que en la era de la modernización damos al término cultura.

Ciertamente debemos reconocer que la disciplina de la evaluación creció en los países desarrollados muy vinculada al pensamiento eficientista, pero también con la presión de identificar las mejores estrategias para mejorar la calidad de la educación.

Por otra parte, afirmamos que el establecimiento de las prácticas de evaluación

\*Investigador del Centro de Estudios Sobre la Universidad (CESU) de la UNAM.

en el sistema educativo mexicano, se ha realizado con un profundo descuido de la necesaria reflexión conceptual, ha quedado ausente la dimensión teórica. En ese sentido, la llamada cultura de la evaluación en México ha sido bastante inculta.

Reconozco que un acto académico como éste tiende a subsanar esa deficiencia. La primera pregunta que yo me planteo cuando se aborda el tema *comparabilidad de resultados de la evaluación* es: ¿cuáles son las finalidades de la evaluación?

### **Finalidades de la evaluación**

El tema de la comparabilidad abre algunas: identificar si hay avances o retrocesos en el sistema educativo a partir de los resultados de la evaluación; ayudar a los que toman decisiones, a los responsables del sistema educativo; informar a la sociedad para que forme una opinión, esperando que esta información coadyuve en la aplicación de políticas educativas, lo cual me preocupa porque es un poco repetir el modelo con el que Reagan utilizó la evaluación en los años ochenta en los Estados Unidos: colocar los resultados de la evaluación del lado de la sociedad, y desde ahí presionar al sistema educativo.

Pero las finalidades de la evaluación en este sentido, me parece que se quedan limitadas. Indicar si hay avance o retroceso, es como si se declarara al enfermo-enfermo, al carro descompuesto-descompuesto, es necesario entender que mientras se desatienda la función de retroalimentación de la evaluación, es prácticamente imposible mejorar el funcionamiento del sistema, si eso es lo que se quiere.

La comparación de resultados es un acto antiguo en los procesos educativos, de hecho cuando se creó la educación comparada, la cual empezó a contrastar el funcionamiento de la educación primaria o de la educación primaria pública en Eu-

ropa, fue a principios del siglo XIX. Donde estamos, yo diría, es en otra era, otra etapa en la que no comparamos solamente el modelo en el que estamos tratando de formar, sino estamos buscando comparar los resultados que se tienen del trabajo, y de hecho, pudiéramos decir que las pruebas internacionales, desde la década de los cincuenta del siglo pasado, se están empleando en el mundo, aunque en México se empleen tardíamente.

### **Comparabilidad de los resultados**

Pero la comparabilidad tiene sus límites y sus riesgos, no siempre reconocidos y no siempre trabajados, uno de los mayores es sólo generar una opinión negativa sobre el sistema y una opinión que de manera muy simple, ingenua yo diría, coloque en el banquillo de los acusados a un actor que ya está puesto en tela de juicio, que enfrenta serios problemas de identidad profesional y ese es el docente, porque si bien es responsable de una serie de actos que acompañan el proceso de aprendizaje, no puede desde el aula resolver un conjunto de problemas que tienen muy diverso origen.

Recordemos cómo ante el gobierno de Margaret Thatcher, maestros ingleses establecieron que si el gobierno seguía haciendo estas comparaciones, ellos iban a impedir que los niños hijos de inmigrantes entraran a su salón de clase porque bajaban los resultados escolares.

El propio pedagogo brasileño Paulo Freire, declaró en 1978 a la Revista *Proceso* que era muy difícil para un maestro, con cuatro horas de trabajo en el aula, deshacer lo que una hora de televisión había hecho en la mente del estudiante.

Se puede entender que la comparabilidad tiene diversos usos para las autoridades educativas; puede tener un sentido diagnóstico, lamentablemente muy des-



perdiciado en la historia de la evaluación mexicana. Desde 1970 la Dirección General de Evaluación ha aplicado pruebas, en la década de los noventa hacía reportes por estado, pero esos reportes, cuando los conocíamos los investigadores, siempre se nos decía: es confidencial, no se pueden fotocopiar, no se puede esto, no se puede aquello. Creo ha sido un gran error del sistema político mexicano.

Me preocupa que esta finalidad, mejorar pedagógicamente el trabajo de los profesores en el aula, esté ausente en la reflexión mexicana de la evaluación, en realidad lo pedagógico es un gran ausente en las reformas educativas mexicanas de la última década.

Tan delicado es este asunto que a nadie le importó, por ejemplo, que en las escuelas normales, y me refiero a preescolar, primaria y secundaria haya desaparecido la didáctica, es más en pro de una bandera luchar contra la teoría, todo mundo pasamos por alto esta atrocidad que se está cometiendo en el sistema educativo.

Es aquí donde los resultados de la comparabilidad pueden servir para tomar decisiones, para presentar una información que impacte al público, los alumnos mexicanos obtuvieron el lugar 32 en la prueba de la OCDE, pero no son considerados los resultados para retroalimentar la función docente.

Cuando un maestro lee, por ejemplo, y lo vemos en los periódicos, los alumnos obtuvieron el 4.2 en la prueba en matemáticas, en la prueba de ingreso al bachillerato, yo profesor de matemáticas con ese dato no puedo hacer nada; no entiendo por qué el sistema de evaluación es incapaz de informar: los diez temas de matemáticas donde el alumno tiene las principales dificultades para resolver el examen son éstos, si a mí me dicen como profesor de matemáticas: éstos son los temas donde el alumno tiene dificultades, puedo ha-

cer algo. Si me refieren: *el enfermo está enfermo*, no puedo hacer nada y eso es lo que me dicen cuando expresan: *México ocupa el lugar 32 o el alumno obtiene cuatro*.

El problema de la comparabilidad implica otro tipo de planteamientos que no se eliminan en el problema técnico de evaluación a gran escala, y es el de la igualdad de instrumentos de medición, la igualdad del metro, lo que algunos autores denominan también el isomorfismo en la evaluación.

Evidentemente, en este contexto siempre es necesario tener presente que la evaluación, además de una disciplina académica y en este sentido parte de supuestos conceptuales, tiene sus problemas técnicos; creo que la evaluación pasó de la teoría clásica a la teoría de la respuesta al ítem, no me queda claro si en las facultades de psicología, de pedagogía, en las escuelas normales, se enseña algo de esto. Supongo que con todo el antiteoricismo, en las normales no; pero no me queda claro si en las facultades de educación se enseñan los nuevos modelos de evaluación, y se está formando la nueva generación de académicos que ayudará a interpretar los resultados de evaluación.

Pero dejando un poco el problema, digamos disciplinario, la evaluación también tiene un conjunto de problemas técnicos específicos y creo han sido presentados de una manera excelente.

### **Contexto cultural y pruebas de conocimiento**

No se debe omitir que la evaluación también es una práctica social, y como práctica social siempre estarán presentes múltiples expresiones de la cultura de la sociedad, las cuales se manifiestan en los procesos de pensamiento de los sujetos y el lenguaje que utilizan, lo que desde hace muchos años también ha permitido relati-

vizar los resultados de los alumnos en los exámenes masivos.

Los niños que aprenden como lengua primera el náhuatl, no solamente tienen dificultades para el español, también aprendieron en el náhuatl un sistema de conteo que es vigesimal.

Tenemos un país con una gran diversidad cultural, donde el lenguaje castellano que empleamos es utilizado con múltiples modismos, todo esto afecta la construcción del examen. Me van a decir que es parte de la calibración.

No entiendo, por ejemplo, por qué en la prueba de la OCDE, y me refiero a ésta porque las pruebas mexicanas siguen siendo, en general, pruebas ocultas. La OCDE publicó 300 reactivos de su banco de reactivos; las pruebas mexicanas más allá que nos puedan decir que se aplican veinte millones o se aplica equis cantidad, en las pruebas mexicanas realmente los reactivos siguen siendo ocultos.

Pero en la OCDE yo no entiendo por qué para el tema de lectura se utilizan ejemplos que implican el conocimiento o el acercamiento al tipo de lentes en cámaras profesionales. No sé si esto sea un conocimiento típico para un estudiante de tercero de secundaria en México, no sé si se hubiera tomado otro objeto, si el estudiante hubiera tenido otro tipo de respuesta frente a esta pregunta.

Y esto es lo que quiero decir con el tema cultural, el tema cultural afecta la comparabilidad de resultados. Hoy Felipe dijo algo interesante, yo lo llamé decálogo no en sentido peyorativo; nos ofrece un conjunto de 17 criterios, en el sentido de normas a seguir.

### **Contenidos de aprendizaje**

Además de la gran ausencia en el tema cultural, ya referida, hay una segunda ausencia relativa a la teoría del contenido.

Cuando uno revisa la prueba de la OCDE, se encuentra, por ejemplo, que cuando hablan de habilidades y destrezas para la lectura, dicen: la capacidad de lector es compleja, lo que significa que la valoración de estas habilidades tiene diversos elementos, lograr una amplia comprensión global, obtener información, elaborar una interpretación, reflexionar sobre el contenido del texto, reflexionar sobre una forma del texto.

Por ejemplo, mientras en España, cuando se analizaban los contenidos, se trabajaba el tema de que existe contenido como dato y eso implica determinado tipo de procedimiento de enseñanza y determinado procedimiento mental, pero que también hay contenido como concepto y este contenido como concepto implica que el sujeto va realizando distintos desarrollos, distintos actos comprensivos en la construcción conceptual de un tema, por ejemplo, el tema masa.

Entre que pasa de la secundaria al bachillerato, del bachillerato a una carrera de física, etcétera, en México este tipo de temas no ha sido totalmente acertado. Basta con que el plan de estudios defina contenidos, no se hace ningún acercamiento sobre la diferenciación de éstos, sobre cómo el maestro debe trabajar algunos de ellos con determinadas estrategias para que sean memorizados, otros para formar un desarrollo conceptual y unos más para formar un nuevo desarrollo procedimental. Todo esto parece omitirse en México.

Tengo la sospecha de que la vieja *Taxonomía de Bloom* sigue siendo un modelo organizador en nuestros exámenes, y eso me preocupa porque no logramos avanzar a otro modelo más allá de que la política educativa declare las ventajas del modelo de enseñanza constructivista. Porque es muy fácil declararlo, mucho más difícil es entender todas las operaciones que implica.



A estos 17 criterios debería incorporarse un criterio cultural, y otro de lo que estoy tratando de definir en este momento como *teoría del contenido*. Para tener cier-

ta consistencia, debemos empezar a definir de qué tipo de contenidos hablamos cuando reportamos los resultados de una prueba.

---

## JOSÉ LUIS GAVIRIA SOTO\*

---

**A**lgunas de las limitaciones de las pruebas de estándares nacionales, tuve oportunidad de señalarlas en una reunión técnica sobre la calificación de estándares celebrada el año pasado y, por tanto, no voy a repetir estos aspectos técnicos que ya han sido mencionados.

Además, probablemente las ideas más importantes al respecto, tendré oportunidad de publicarlas en breve tiempo, pero sí quiero referir algunas cuestiones.

Ya el licenciado Martínez Rizo ha destacado que se trata de un proceso técnico complejo en el cual hay dos aspectos, uno relacionado con el diseño de las pruebas y el otro con el aparato técnico.

Insistiría en que en cualquier equiparación, en cualquier proceso de comparación de resultados, hay un error. Lo que importa no es tanto que haya error, como que la magnitud de éste sea aceptable por nosotros, y esta magnitud del error está a su vez relacionada con el cumplimiento de los supuestos en los que se basan las técnicas que utilizamos.

Para insistir en la importancia de los aspectos, no sólo técnicos sino también del diseño de la equiparación, voy a hacer uso de una metáfora, que me gusta utilizar con cierta frecuencia cuando hablo de estos términos, es una metáfora que tiene

que ver con la imagen, con el desarrollo de la imagen.

### **Una imagen instantánea proporciona mucha información**

Cuando nosotros hacemos una evaluación, una sola evaluación en un momento determinado, en un año, nos traemos una buena muestra, hacemos una buena prueba, aplicamos con un procedimiento riguroso, hacemos los cálculos necesarios. Lo que obtenemos es una instantánea y una foto instantánea puede proporcionar mucha información.

De hecho, probablemente la tecnología de la imagen, del control de la imagen fija, durante mucho tiempo consistió precisamente en captar la imagen estática, la imagen instantánea, y ésta es fuente de gran número de datos. Sin embargo, varias fotos consecutivas no necesariamente nos dan una imagen del movimiento. A los especialistas en la foto fija, lo que más les gustaría tener es una imagen en movimiento, y esto fue lo que durante mucho tiempo se estuvo preparando con el desarrollo del cine.

¿Qué ocurre con la evaluación?, conocemos varias evaluaciones consecutivas en años distintos, pueden ser al cabo de

---

\*Catedrático de la Universidad Complutense de Madrid.

un año, dos, o tres, pero hay aspectos que cambian; cambia el encuadre, cambia el enfoque, cambian las circunstancias que impiden tener una imagen continua. Estos aspectos naturalmente hacen que cambie la manera de interpretar la secuencia de imágenes.

Debo decir, continuó con mi metáfora, que hay también dos evoluciones en el desarrollo del cine. Una de ellas tiene que ver con el desarrollo tecnológico, y esto ha pasado también con la evaluación, cuyos avances ha presentado muy bien el licenciado Felipe Martínez Rizo al principio, donde se observa el paso desde una serie de técnicas que se desarrollaron empíricamente a principios del siglo XX, hasta la teoría clásica, luego se desarrollaron otras teorías afines, muchos modelos distintos.

### **El lenguaje de la evaluación**

Realmente lo que ocurre con la evaluación en este preciso instante, es que estamos pendientes de los desarrollos tecnológicos básicos, los cuales no son más que la infraestructura de lo verdaderamente importante que es el lenguaje evaluativo del cual se estaba hablando; es decir, el desarrollo de un lenguaje, de una estructura semántica en la que todos estamos de acuerdo.

Voy a mencionar tres elementos en esta estructura del lenguaje: léxico, sintaxis y ritmo, los lingüistas probablemente me dirían: no, eso ya está pasado de moda porque yo tengo tres hijos, se llevan dos años entre ellos pero cada uno estudia con un libro distinto, una lingüística distinta.

Yo iba a decir que el español no es más que el mexicano hablado a mucha velocidad, pero después de oír al doctor Ángel Díaz Barriga estoy dispuesto a cambiar mi definición y creo que no es exactamente así. Creo que todos ustedes están de acuerdo conmigo con que en el lenguaje

una cosa muy importante es el ritmo, y el ritmo al que se produce la información.

Cuando oímos hablar un lenguaje extranjero no es que no entendamos las palabras, es que se producen a más velocidad de la que somos capaces de absorber.

Bueno, en la evaluación los efectos aparecen a medio plazo y es que siempre tenemos tendencia a querer interpretarlo todo inmediatamente.

En cuanto al léxico, lo diré muy rápidamente, éste es exactamente igual, es arbitrario y compartido, lo realmente importante es que cuando yo utilice una palabra ustedes sepan a qué me estoy refiriendo y tengamos el mismo referente.

Por último voy a hablar de la sintaxis, la cual es el conjunto de reglas que nos permiten establecer sentencias que tengan sentido y en evaluación pasa exactamente igual.

Fijense, podemos estar de acuerdo en la sintaxis, los 400 puntos significan tanto que los alumnos saben hacer esta cosa, la otra, no saben hacer esto, no saben hacer lo otro; pero lo que no podemos cambiar, fijense ustedes, es algo fundamental y es que a pesar de que tendemos a hacer una atribución causal inmediata de los resultados, los resultados que se obtienen en evaluación no son de experimento, son de observaciones y esto es muy importante. Hay que tener mucha cautela cuando hacemos atribuciones causales.

Imagínense que quiero saber si hay alguna relación entre el hecho de llevar corbata y el salario mensual que uno recibe. ¡Ah! me voy a ir a un famoso centro comercial en España y voy a tomar nota de qué personas salen y voy a anotar las que llevan corbata y las que no, y voy a preguntarles además cuánto ganan al mes.

Al final de mi estudio observo que las personas que llevan corbata ganan como 100 mil pesetas más que las otras. Ésta es la razón por la que yo llevo corbata. Se han



reído ustedes porque no se lo creen, simplemente está claro que no hay una relación causal ahí.

Se puede hacer un experimento para estudiar esto mismo, claro, yo por ejemplo, esta audiencia pudiera dividirla en dos partes y a unos durante los próximos dos años les obligaría a llevar corbata todos los días y a los otros nunca, si hubiese alguna diferencia se hubiese hecho esto aleatoriamente, la diferencia que yo he podido observar es que realmente podría

atribuirla al hecho de llevar una corbata, pero de otra manera no, en absoluto, y es que corremos el mismo peligro con la evaluación.

Las observaciones que hacemos naturalmente se deberán a algunas causas, pero no tenemos ninguna seguridad de que conozcamos esas causas porque aparte nos faltan herramientas teóricas, nos falta conocimiento sustantivo del proceso por el cual se produce esa relación causa-efecto.

---

## MANUEL GIL ANTÓN\*

---

**M**i madre decía que las comparaciones son odiosas, no se si ustedes han escuchado la frase, por lo cual he titulado mi participación *Las comparaciones pueden ser odiosas o inapropiadas*.

Agradezco a Felipe Martínez Rizo su invitación a participar en estas jornadas, mi contribución está dedicada a la memoria de Javier Palencia por su aguante cordial en el periodo en el que participé junto con Juan José Rivaud y otros colegas, en el Comité Académico del EXANI II en el CENEVAL, le dimos una lata impresionante y por ahí estará oyendo esto.

Como contamos con diez minutos, mi participación será a manera de telegramas. Una primera serie, se dedica a puntos generales que no quiero dejar de lado sobre el tema y la otra, los otros seis, al tema preciso de la comparatividad y dos *E mail*. Como se dice en la fiesta taurina: al toro, telegramas generales.

Primero, ¿México necesita un Instituto Nacional para la Evaluación de la Educación, INEE, o debe conformarse con un Instituto Nacional de Evaluación del Aprendizaje, INEA?, no confundir con el de la Educación de Adultos, es una simple coincidencia de siglas.

La educación es un concepto más amplio que el de aprendizaje, aunque el primero incluye al segundo, pero su extensión, siguiendo a Aristóteles es menor; creo que debemos poco a poco ir en dirección de la evaluación de la educación que incluye además del aprendizaje, por ejemplo en lectura o matemáticas, la ponderación de nuestras escuelas, subsistemas, regiones y sistema nacional en tanto ambientes en los que los muchachos mexicanos son educados para la vida, la convivencia, la participación como ciudadanos, el respeto a los otros.

Es decir, como espacios de socialización en el sentido amplio del término, dimen-

---

\*Investigador del Departamento de Sociología de la Unidad Iztapalapa de la UAM.

sión en la que es central, si son sitios en que se aprende y ejercita la capacidad de una actitud crítica y constructiva de nuevas posibilidades sociales.

Quizá no existan pruebas estandarizadas para ello, pero hay que ser creativos y buscar aproximarnos a una noción amplia de educación que no se restrinja al aprendizaje como capacidad para resolver pruebas de opción múltiple, es más fácil decirlo que hacerlo, me hago cargo de ello, pero el telegrama surgió así y así lo dejo en la mesa.

### **Las comparaciones son necesarias**

Segundo, equiparar es una cuestión compleja, como hemos aprendido de la ponencia de Felipe Martínez Rizo, pero vivimos en un tiempo en que cualquiera, con la mano en la cintura y sin parar mientes en la superficialidad de sus juicios, equipara a las primeras de cambio.

Por ejemplo, en el campo que mejor conozco, la educación superior, con base en una supuesta capacidad de equiparar, muchos afirman que las universidades públicas son fábricas de desempleados y que las privadas finas, sí preparan para el empleo, pues sus egresados lo consiguen rápido y en mejores condiciones.

¿Son equiparables las condiciones socioeconómicas y culturales de los estudiantes y sus familias en estos dos contextos? Claro que no. El país está partido por la desigualdad social, pero como la superficialidad en el juicio no se da cuenta de la escasez de empleos durante más de veinte años, concluye sin problemas que unas son mejores que otras, menudo modo de comparar.

Tercero, a pesar de la conseja popular de doña Julia Antón, que declara que las comparaciones son odiosas, resultan necesarias pero hay que hacerlas bien, necesitamos comparaciones bien fundadas al

interior de cada escuela, entre escuelas y a lo largo del tiempo, ya sea a nivel de cada plantel, de regiones, subsistemas, modalidades y el sistema en su conjunto de la educación básica. Sí, es necesario pero no trivial: otra lección general en el texto de Felipe y yo la acepto.

Cuarto y último telegrama general: La condición para evaluar bien no es contar con una gran cantidad de información, no es una cuestión de aprovechar los millones de pruebas que se han aplicado, si no son comparables, la cantidad no supe a la calidad que puede lograrse con menos datos, pero bien construidos y estadísticamente pertinentes.

Por ejemplo, el Ceneval ha realizado millones de pruebas para entrar a prepas y a las universidades. ¡Ándale Manuel, te ofrecemos todos esos datos para que hagas un estudio y sepamos cómo están las secundarias o los bachilleratos!... Pero no.

Si los exámenes se hacen con el fin de discriminar a los estudiantes para asignar preferencias saturadas, la naturaleza del examen no conduce sin mediación reflexiva de fondo a concluir sobre la calidad de los ciclos precedentes, y esto no es fácil destrabar para muchos colegas.

Sin embargo, es necesario; ayer en el periódico donde escribo apareció que el setenta por ciento de los muchachos de la secundaria están reprobados en matemáticas, según el resultado del Ceneval. ¡No es cierto!

Ahora los telegramas específicos.

Comparto, luego de escuchar la ponencia de Martínez Rizo, que los datos con los que contamos no son en gran medida comparables, aunque sean muchos y pese a la prisa de los adoradores de los *ranking*.

### **Reconocer los errores**

Más vale dar a conocer las fallas de los datos que proceden de teorías y méto-



dos inapropiados, que ceder a la presión y ofrecer listados de la mejor escuela, con dudas sobre la pertinencia de la construcción. Paciencia, ni hablar.

El ejercicio democrático, la rendición de cuentas, incluye que informemos sobre los procesos y resultados de las actividades, y sus posibles deficiencias, por ejemplo, inconsistencias en el levantamiento de los datos; es una forma incómoda de concebir la entrega de cuentas, pero es imprescindible. Es mejor decir eso a hacer comparaciones que no se compadecen, como nos contó Felipe, con la experiencia. En dos años no puede caer cuarenta puntos el aprendizaje, salvo que haya habido una catástrofe. La catástrofe está en el instrumento.

Es correcta la pregunta central de la ponencia, ¿cómo elaborar pruebas para el año 1 y el año 2 que sean a la vez distintas en sus reactivos, pero equiparables, comparativamente pertinentes?

En otras palabras y recordando los cursos de metodología que uno toma, los indicadores han de ser válidos, que midan lo que se proponen medir, confiables, que no se muevan caprichosa o aleatoriamente en cada aplicación y, para acabarla de complicar, comparables en el espacio y el tiempo. Nada más, sí, pero nada menos.

Propongo un barbarismo para divertirnos en este asunto. Si en inglés el término para decir “comparabilidad”, esto es las teorías y métodos que aseguran la equiparabilidad es, lo voy a pronunciar según mi sangre gallega, “ecuating”, en mexicano podríamos convertirlo en verbo “ecuatizar” y someter al análisis de la conveniente “ecuatización” lo que hemos hecho y podemos hacer, y podríamos jugar con aquella adivinanza de que: “el que ecuatice con fundamento mejor ecuatizador será”.

Me parece en todo acertado el empleo de ítems ancla, entre una prueba y otra, esto permite analizar la comparabilidad

de las pruebas. El supuesto de que la tasa de cambio, el tipo de variación de un sistema educativo compuesto por millones de chavales y decenas de miles de escuelas es lento, me parece lógico, salvo catástrofes.

Por ende, si entre un año y otro la validación es desproporcionada, ha de buscarse la falla en el instrumento antes de imputarla a la realidad. Si no revisamos las teorías y técnicas derivadas, como hace Felipe, con las que elaboramos los instrumentos y las consideramos perfectas, las anomalías se pueden atribuir a la realidad y caemos en la frecuente sobreestimación de nuestros supuestos que se enuncia bien con la conocida frase: “pobre realidad, no se ajusta a nuestra sabiduría”. Vaya soberbia.

Penúltimo. Desde mi poco conocimiento en la psicometría, los criterios expuestos en las páginas 9 y 10 de mi versión, que expuso Felipe, resultan lógicos y es menester seguirlos, aunque esto no signifique que sea sencillo lograrlo, pero para esto tenemos al INEE, para hacer un trabajo serio que no eluda las fallas de la información precedente y las que seguramente será inevitable que vuelvan a presentarse. En este caso, una sana noción *Popperiana* de ejercicios de refutación, se antoja un excelente camino.

Último de los específicos. En conclusión, creo que este texto nos muestra que no hemos hecho del todo bien las cosas, aunque la intención ha sido buena. De bola a bola, se antoja decir que las buenas intenciones han empedrado el camino del infierno, pero eso es un lugar común.

Mejor, como afirma el autor de la ponencia, pongamos manos a la obra, demos el tiempo necesario para enmendar las fallas observadas y vayamos construyendo, con bases firmes, las condiciones de comparabilidad que requiere el país para debatir con un piso sólido informativo las maneras de mejorar el sistema.

Una nota final. Pienso, luego insisto, como diría el clásico.

Los resultados de la evaluación del aprendizaje son sólo una parte de la evaluación del proceso educativo, importante sí, pero no suficiente. ¿Cómo enfrentamos las otras esferas que incluyen el esfuerzo educativo y su inversión social? Hay que pensarlo bien y no dejarlo de lado.

Y dos breves *E-mail*. Ojalá el lema del INEE no fuese evaluar para transformar. A mí, quizá por desviación profesional, me gustaría *evaluar para comprender*, hay un montón de asuntos que comprender en este asunto de la educación.

Y el otro *E-mail* es que hay que salirnos del fútbol, cada vez que se presentan los resultados, los metemos como si fuese una tabla de posiciones en el fútbol.



## JORGE HERNÁNDEZ URALDE\*

**Q**uisiera, si me permiten, iniciar los comentarios relatándoles una anécdota que, pienso, va a orientar el sentido que le quiero dar a los mismos.

### Experiencia que reporta el EXANI I

En 1996, en la ciudad de México se crea la Comisión de Instituciones de Educación Media Superior, que se denomina la Comipems, y deciden que debe implantarse en forma generalizada un examen para determinar el ingreso a las instituciones de educación media superior públicas, que imparten este nivel educativo en la Zona Metropolitana de la Ciudad de México.

Deciden estas instituciones que se cree el Examen Nacional de Ingreso a la Educación Media Superior, que es conocido como el EXANI I, que diseña y elabora el Ceneval. Hasta aquí las cosas muy bien.

El primer ejercicio que hacemos en torno al diseño de este instrumento, las instituciones nos señalan, los funcionarios a cargo de este proceso nos señalan que

para el caso y para el efecto de este proceso, seguridad mata a comparabilidad.

¿Qué quería decir esto? Señores, no pueden ustedes incluir en sus versiones de exámenes un solo reactivo común en las distintas versiones que se apliquen año con año.

Muy bien, dijimos “tendremos que acatar esta disposición y entendemos que el asunto de la seguridad es el asunto crítico de este proceso”. Así lo hicimos, realizamos los primeros exámenes, tuvimos las primeras aplicaciones tres años consecutivos y después de estos primeros tres años, un funcionario perspicaz notó que la media del examen había subido un poco año con año, el resultado promedio era un poco más alto en el tercer año que en el primero y entonces discurrió que lo que estaba detrás de esto es que la calidad educativa, el rendimiento de los jóvenes era mejor, y se nos comentó entonces que hiciéramos público este resultado, que era muy conveniente comentarlo en ese sentido. Dijimos “no se puede”.

¿Por qué? Porque el examen no está diseñado para esos propósitos, está elabo-

\*Director General Adjunto de Programas Especiales del Ceneval.



rado con referencia a la norma, se recentra año con año para que el total de los jóvenes alcance un porcentaje aproximado del cincuenta por ciento de aciertos. Entonces, no es posible hacer afirmaciones en torno a esto.

Nos dijeron “pero cómo es posible, debe haber un conjunto de elementos comunes de las pruebas que nos permitan hacer alguna afirmación en este sentido”. Dijimos “no, señores, nos solicitaron que no pusiéramos ningún reactivo común en las pruebas”.

Se acuerdan ustedes, gran frustración de los funcionarios, no pudimos generar información al respecto y no pudimos cumplir con la expectativa que las instituciones educativas convocantes tenían al respecto.

El problema fue que la instancia evaluadora no pudo comunicar en forma suficientemente clara el propósito, sentido y naturaleza de la evaluación que se estaba haciendo.

Simultáneamente, con la prensa encargada de dar cuenta de lo que estaba ocurriendo en estos medios, año con año tuvimos que enfrentar un problema: cómo explicar que la mitad o más de la mitad de los jóvenes que presentaran el examen, reprobaban.

Por más que nos dedicamos, en talleres incluso, a explicar cómo la naturaleza del examen y la forma como está definido no permitía sacar estas conclusiones, la prensa siguió informando al respecto año con año.

Bueno, esta es la anécdota y el referente que deseaba tomar para mis comentarios al trabajo que nos ha presentado Felipe Martínez Rizo.

### **Los datos que actualmente tiene el INEE no son comparables**

La primera cuestión muy clara en relación con los datos presentados por Felipe,

es que efectivamente todos los que trabajamos con el tema de la evaluación educativa, nos damos cuenta de que los datos con que cuenta el INEE actualmente, impiden hacer inferencias respecto al avance o retroceso de la calidad del sistema educativo, en virtud de que los datos no son comparables, creo que esto es un asunto que se concluye muy fácilmente dada la prolijidad de los datos que nos presenta el estudio de Felipe.

Además, yo pienso que si el Instituto tendiera a exponer alguna inferencia en este sentido cometería un gravísimo error, y me parece muy bien que ahora sustente esta negativa explicando metodológica y técnicamente, en forma suficiente sus razones.

El segundo asunto que me parece claro en el trabajo de Felipe Martínez Rizo, es que el tema de la medición de resultados de aprendizaje es un asunto sumamente complejo, es prácticamente para iniciados que requiere el conocimiento teórico, metodológico, de asuntos sumamente sofisticados y entonces tenemos un grave problema porque nuestra audiencia no está compuesta por expertos en la materia, está compuesta por funcionarios públicos, padres de familia, alumnos, y directivos escolares; cómo hacemos para traducir toda esta complejidad metodológica y de datos en términos tales que hagan comprensible el problema.

Es un país donde hasta hace muy poco el tema de las pruebas educativas era inexistente, yo recuerdo que el asunto empezó a discutirse en los medios, justamente a partir de la implantación de este concurso metropolitano en la zona de la ciudad de México.

Tenemos, entonces, audiencias poco entrenadas, tenemos personal poco calificado para contender con este problema. Tenemos un reto de comunicación que para las agencias encargadas de este tipo

de desarrollos es una necesidad insoslayable e imposterable.

Si se considera además que las pruebas están ya instaladas, el tema de las pruebas está incorporado en el debate público y no con poca frecuencia las inferencias sobre los resultados de las mismas, hechas por políticos, comunicadores y el público en general son simple y llanamente erróneas.

Para evitar interpretaciones equivocadas es importante que los ciudadanos in-

teresados, comunicadores y formuladores de políticas entiendan qué significa realmente la terminología de las pruebas.

Por ello, las agencias y especialistas dedicados a la evaluación educativa, tenemos una enorme responsabilidad de atender urgentemente esa necesidad educativa de la sociedad, para permitir interpretaciones pertinentes e inteligentes acerca de los datos y estudios que presentamos a la opinión y al escrutinio público.



## ARTURO DE LA ORDEN HOZ\*

**D**esde la perspectiva estrictamente de evaluación, ojo, no estoy hablando de psicometría, ya hablaremos un poco, sino de evaluación, la comparación es un elemento básico.

Ayer les decía que la evaluación era fundamentalmente un juicio de valor sobre una determinada realidad; pero claro, un juicio de valor necesariamente se apoya en una comparación.

Siempre que hablo de esto tengo la tentación y muchas veces no resisto, es un chiste y no muy bueno, pero recuerdo que en los años setenta los libros de evaluación que aparecían en el mundo anglosajón en Inglaterra y en Estados Unidos, iban acompañados de unas viñetas humorísticas para que los estudiantes entraran en esto.

Y hablando de evaluación y tratando de precisar qué era eso, aparecía una viñeta en un libro donde dos amigos se encuentran en la calle y uno le pregunta al otro: oye y cómo está tu mujer; el otro dice: comparada con quién; evidentemente la respuesta es muy diferente, al consi-

derar con quién la compares o qué quieres saber, pues esto pone de manifiesto lo que es el papel, la función de la comparabilidad en la evaluación.

### **La comparación en evaluación**

Qué es lo que realmente debe compararse para que haya evaluación, pues lo que tenemos que hacer para llegar a un juicio acerca de la adecuación de una realidad evaluada, el aprendizaje de dos alumnos, la educación, una escuela, un profesor, cualquiera que sea el objeto de evaluación, cuando hacemos evaluación, lo que hacemos es ver si es coherente de comparar con un criterio en el que realmente ya está, de alguna manera se supone, el ideal del maestro sobre aprendizaje, pero para comparar al maestro, el aprendizaje o la escuela, tengo que recolectar información, para qué, para definir precisamente el objeto evaluado.

Y entonces, una de las formas de recogida de información es la medida, por tanto estamos... y quiero distinguir que

\*Profesor de la Universidad Complutense de Madrid y miembro del Consejo Técnico del Instituto Nacional para la Evaluación de la Educación.



estamos hablando de comparabilidad en términos psicométricos, pero es paralelo al fenómeno básico de la comparación en la evaluación, la comparación y el juicio de valor, son los elementos nucleares del concepto de evaluación y por tanto, es importante que pueda hacerse; podemos usar la medida y entonces viene una segunda comparación, usamos la medida como una descripción peculiar de la realidad, la medida no hace más que hacer una descripción métrica del aprendizaje, poner unas puntuaciones, pero si no vamos más allá, la medida no hace juicios, y si no hace juicios no precisa comparaciones.

Lo que ocurre es que después, estos problemas de tipo psicométrico, que llevan a efecto comparaciones que son necesarias para que puedan hacerse realmente las comparaciones sustantivas, son a las que he venido aludiendo últimamente, es decir, comparar y ver la coherencia entre la descripción de la realidad a evaluar y el criterio, que es el elemento que consideramos básico y que sería el ideal; la comparación es uno de los elementos importantes y adquiere un valor añadido.

### **¿Qué tipo de evaluación vamos a realizar?**

Elaborar un juicio acerca de si el progreso de los estudiantes a lo largo de un periodo es el satisfactorio o no, ese es el juicio y en esta línea necesitamos dos series de medidas y estas dos series de medidas tienen que ser hechas con instrumentos que estén equiparados, está así de simple la cuestión y claro, así de simple a ustedes les han puesto de manifiesto toda la complejidad que hay detrás de la equiparación, de la comparabilidad y naturalmente, nosotros no podemos prescindir de que los resultados de las pruebas deben ser comparables, si queremos evaluar el progreso de los estudiantes.

La dificultad de establecer una especie de paralelismo en estas dos formas o líneas de comparación, aumenta porque son los factores que están implicados, ambas son verdaderamente altas.

Por otra parte, nos encontramos que en muchos casos la propia medida se identifica con la evaluación y de hecho decimos: voy a evaluar, y simplemente voy a aplicar una prueba y ¿qué estoy evaluando?, está usted midiendo. Las pruebas tienen en sí mismas unos referentes, para saber si los mismos dicen algo acerca de la evaluación.

Cuando un profesor evalúa un examen de un alumno que tiene ahí y le pone siete o le pone diez, no sólo está describiendo métricamente el aprendizaje de los alumnos, está evaluando porque siete en el contexto que nos movemos significa que sabe el alumno una cantidad determinada, tiene un nivel de conocimientos y si tiene diez tiene otro, está incluida ya la evaluación, por tanto, se ponen en conexión la comparación en ambas series paralelas.

La literatura psicométrica sobre equiparación ya señala la gran cantidad de factores a tener en cuenta en la elaboración y aplicación de pruebas para lograr unos niveles aceptables de equiparación.

La explicación de que exista ya esta literatura y señale tantos factores es muy simple, en el fondo todas las facetas y operaciones de la elaboración y aplicación de la prueba, todas inciden en la equiparación, y naturalmente si esto es así, no es extraño que este magnífico documento que ha partido ahí, haga referencia a prácticamente todos los documentos, o sea, hacer una buena prueba, es hacerla de una determinada manera, que está implicando la equiparación.

Quiero terminar casi en la misma perspectiva que he empezado, y es que entre estos factores hay uno esencial que apare-

ce también en primer lugar, pero no sólo es esencial para la equiparación, sino para la elaboración de la prueba también, que es en el fondo: ¿cuál es el propósito de esta prueba?

Y el propósito de la prueba, es el que va a determinar cuáles son los aspectos críticos para tener una prueba que dé resultados iguales o equiparables.

A partir de ahí es cuando empieza todo lo que conocemos como la literatura técnica, de diseños, metodologías, etcétera. Pero es un tema para discutir, trabajar, o estudiar en común, y me parece

que la idea de que la evaluación en México y en este Instituto, entre otras funciones tiene la de evaluar el aprendizaje, concretamente nuestra perspectiva es el progreso del aprendizaje de los alumnos, la idea del progreso adecuado que ahora se ha puesto tan de moda en los Estados Unidos como consecuencia de esta ley de: "ningún niño debe quedar retrasado".

Considero que la evaluación del progreso exige necesariamente una atención precisa, detallada, minuciosa, del problema de la equiparación.



## PREGUNTAS Y RESPUESTAS

### MANUEL ÁLVAREZ MANILLA

**Quisiera conocer la comparabilidad de la evaluación en Ciencias Sociales, Historia, etcétera. ¿El ser humano sólo puede ser evaluado en matemáticas y lectura? ¿No es esto reduccionista? ¿Hacia dónde se dirigen las escuelas normales sin considerar a la didáctica?**

Bueno, realmente la evaluación, como se dijo ayer en la mesa, puede evaluar todo, no nada más es el propósito de la evaluación, por ejemplo, la escuela o el profesor de historia, pues es importante emitir los juicios de valor sobre lo que pasó con los alumnos.

Pero cuando se trata de un proceso que fue señalado repetidamente el día de ayer, que es la evaluación del sistema educativo, qué resultados está produciendo en términos de aprendizaje, no de educación como decía Gil Antón sino de aprendizaje; realmente sucede que hay bastantes evidencias empíricas de que la evaluación en

matemáticas y lectura es un indicador de cómo está el resto del sistema.

Por ejemplo, nosotros en una investigación que hicimos vimos la correlación que existía entre la calificación en el examen para ingreso al bachillerato y la retención, o sea, la posibilidad, la probabilidad de que el alumno se quedara en la escuela.

La correlación con todas las preguntas que utilizaban, excepto la de razonamiento verbal y la de razonamiento matemático era bajísima, casi no existía. En cambio, la posibilidad de que el estudiante quedara en el bachillerato era muy alta si tenía calificaciones elevadas en español y matemáticas.

Esto lo ratifica un estudio reciente, en el cual se plantearon esa interrogante, vieron la proporción de varianza que en un momento dado podría ser explicado por lectura, matemáticas y por el contenido de las materias; se encontraron que la calificación de las materias, cualquier diferencia de matemáticas y español, obedecía a la explicación de un componente G



general que estaba dado por matemáticas y lectura, más que por los contenidos específicos.

Entonces, no es que sólo se puedan evaluar matemáticas y español, sino que son indicadores razonables de cómo va evolucionando el sistema en su conjunto en materia de aprendizaje. Realmente me siento incompetente para contestar hacia dónde se dirigen las escuelas normales.

### ÁNGEL DÍAZ BARRIGA

#### **¿La evaluación va a modificar el plan de estudios o va a modificar las formas de enseñanza?**

Creo que ese es un riesgo gravísimo de la evaluación, la relación histórica es al revés, la apuesta educativa ha sido al aprendizaje.

Cuando Comenius hace *Didáctica magna* en 1657, y perdonen que tenga una mirada histórica, el tema *examen* forma parte del tema de *Metodología de enseñanza*; históricamente después se separó, ya tiene su sentido y ya tiene su forma de ser, pero donde veo la perversión es en la posible pretensión de regular la enseñanza desde la evaluación.

Creo que el sistema educativo y el sistema educativo mexicano deben volver a apostar a la reorganización de los sistemas de aprendizaje, y al principal en los sistemas de aprendizaje, que en mi opinión es el docente.

#### **¿Hay que homogeneizar calificaciones en primaria y secundaria?**

No creo que sea el tema fundamental, me preocupo por la palabra homogeneizar. Precisamente cuando estamos hablando de diversificar, cuando estamos hablando de que un problema en la comparabilidad es reconocer lo singular.

**Hay otra pregunta que habla de escuelas de Chiapas, y otra que se refiere a contexto.**

Todos los estudios internacionales de la OCDE, los estudios del TIMSS, antes de dar el reporte de evaluación hacen un planteamiento de contexto, y es algo que ha faltado en nuestro medio.

Quiero añadir otro aspecto, el Laboratorio de Evaluación de América Latina, que se hace en Santiago de Chile, por las condiciones de América Latina, que no son las condiciones del primer mundo, promovió el estudio regional *Escuelas Destacables*.

Es una manera muy pertinente de emplear los resultados de la evaluación, no *escuelas excelentes*, sino *escuelas destacables*; es decir, aquellas escuelas que se encuentran en un contexto adverso, pero sus alumnos obtuvieron resultados por encima de lo esperado.

La pregunta formulada en el estudio es: ¿cómo trabajan los maestros, qué pasa en la escuela, cómo es la organización escolar para explicar este resultado?

Considero que en México la evaluación sigue funcionando con un esquema muy sancionador. Manuel lo definía con un concepto de partido de fútbol, pero creo que en el fondo y lo más grave es que es muy sancionador: están mal, vean qué puntaje tan malo tienen.

En vez de decir: "Esto es un dato, cómo lo comprendemos, qué nos quiere decir; sobre todo, qué podemos hacer frente a este dato". Mientras no se pase a la segunda estrategia, no vamos a poder hacer nada.

#### **¿Por qué la postura de no proporcionar los resultados de evaluación en México?**

Considero que parte del problema es una falta de cultura de la evaluación, se segmenta la evaluación, todos los que estudiamos evaluación sabemos que tiene una función de retroalimentación y una función formativa. ¿Dónde está en la política mexicana la función de retroalimenta-

ción y la política formativa? Informar no es retroalimentar. Informar es decir: el resultado está así.

La función de la retroalimentación está pensada *al garete*, perdón que lo diga con este término, porque me van a decir, no es una función que corresponda a Ceneval ni al INEE, y creo que tienen razón, pero tampoco corresponde al diseñador de política en el caso de las distintas subsecretarías; entonces queda *al garete*, queda en el aire. Y la función sancionadora es la que va teniendo peso socialmente, esa es una de las cuestiones altamente preocupantes.

Y una última pregunta, y Manuel también lo decía, y lo decía también Roberto, evidentemente que para obtener un dato de evaluación hay que tener formación en estadística para poderlo interpretar, y si no se tiene formación, se pueden hacer múltiples generalidades sobre este dato, múltiples simplificaciones sobre el mismo.

Ahora, ¿qué es lo que ha pasado en la política educativa en el país? Que no se ha dedicado a hacer una formación social; quiere que la sociedad maneje estos datos, hay que desarrollar la formación social para su manejo, pero tampoco le ha preocupado formar a los cuadros directivos y docentes de las escuelas.

Mi experiencia en investigación me confirma, que cuando al directivo le llega un reporte, y he logrado entrevistar un gran número de ellos, dicen: ¿a poco el reporte contiene otros datos?

Entonces, se habla mucho de la cultura de la evaluación y creo que por cultura de la evaluación se quiere decir una cosa muy simple: que los mexicanos aceptemos que las escuelas sean evaluadas, pero no se quiere decir que manejemos conceptualmente el tema de la evaluación.

Y un directivo que recibe un reporte de evaluación de su institución debe saber interpretarlo, debe tener un apoyo para la interpretación del trabajo que tiene fren-

te a sí. Por otro lado, el único que considero puede transformar el sistema educativo en serio, es el maestro, el que va a estar con los niños 200 días de clase, es el único que va a empezar a incidir en un cambio real.

Recuerdo en alguna investigación, lamentablemente no la publiqué, que frente a exámenes departamentales había maestros que decían: mañana vemos de la pregunta 1 a la 75, de los instrumentos que venden de mil preguntas y respuestas; como código con los estudiantes, había maestros que recomendaban: tomen ese manual y lo estudian en su casa porque si prueban les va a ir mal.

Finalmente, en quien aterriza toda esta problemática es en el maestro y no perdamos de vista que un gran problema que enfrentan países como Chile es que si ustedes van hoy a una librería, lo que encuentran son libros de mil preguntas escritas como pruebas objetivas para preparar todos los exámenes, de cada una de las materias de secundaria y ahí hay una veta que no se ha explotado en México y que si la evaluación sigue, la dinámica puede ser la salida de maestros y de estudiantes.

Si me están diciendo que soy buen maestro porque mis alumnos responden bien la prueba, entonces los entreno para que respondan este tipo de instrumento solamente.

### **JOSÉ LUIS GAVIRIA SOTO**

#### **¿Qué pasa con el problema de las comparaciones estadísticas como un modelo? ¿Qué es lo rescatable para el caso de México?**

La respuesta es muy rápida, ya lo dijo el licenciado Martínez Rizo. A partir de nuevos diseños en futuras evaluaciones pueden hacerse comparaciones de datos que se recogieron anteriormente, con datos que se recogerán con posterioridad.



Lo que no se puede hacer ahora mismo, es reanalizar los datos que se han retomado aunque cambie la metodología, o el modelo, da lo mismo; pero sí se pueden comparar los datos que se tomaron en el pasado con datos que se tomen en el futuro, utilizando pruebas diseñadas para tal efecto.

**¿Qué relación hay entre los resultados de evaluación que se realizan en las escuelas?...las escuelas reflejan niveles de aprendizaje y manejo de habilidades en niveles de excelencia, mientras que las evaluaciones internacionales, de las grandes evaluaciones, presentan resultados negativos.**

La respuesta en este caso es que las calificaciones que asignan los maestros en las escuelas son totalmente contextuales, todos estamos perfectamente conscientes de que en cada país, en cada lugar existe un código distinto, por ejemplo, en España el rango de aprobado va desde cinco hasta diez.

Entonces, qué hace realmente un maestro, cuando asigna un nueve ¿quiere decir que asignaría la misma nota a cualquier alumno que supiese exactamente lo mismo que saben éstos? No, realmente los maestros en una escuela lo que hacen es ordenar a sus alumnos, todos los maestros ordenan a sus alumnos de mejor a peor y en este caso, las calificaciones escolares son totalmente contextuales; esto hace que haya calificaciones que son, por ejemplo, de nueve en una escuela que realmente reflejan un contenido que no pasa de siete en una escuela distinta, depende del nivel de la escuela.

Un alumno mediocre en una escuela con alumnos muy malos tendrá notas bastante buenas. Un alumno en una escuela de alumnos buenísimos tendrá notas peores que la anterior.

Las notas de las escuelas son contextuales y no sirven en este sentido para hacer comparaciones. Ésta es la razón de que haya necesidad de realizar evaluaciones en las que realmente la comparación sea posible.

## MANUEL GIL ANTÓN

**¿No sería más valioso que la evaluación se oriente más y mejor a que nuestros estudiantes logren actitudes y conocimientos que les faciliten acceder a actitudes de mayor comprensión, ampliación de nuevos saberes y apertura a cambios en ciencias, artes o técnicas?**

Creo que sí, pero como refiere la mala broma repetida, *el descuartizador diría que hay que ir por partes*, porque creo que la temática que hoy vimos, se circunscribe a una problemática específica y muy seria: la sociedad le está pidiendo a usted que compare, si vamos mejor o si vamos peor, si esta escuela es mejor o no.

Primera cuestión, que el instrumento sea confiable y válido, pues si no lo es, en caso de que se transmita esa información vamos a enviar señales equívocas.

Más allá del detalle, yo pensaría mucho en el lema de Felipe, de *evaluar para transformar*. Me parece que la evaluación se está convirtiendo en un fin, no en un medio, y creo necesario evaluar para comprender la complejidad del proceso educativo.

También podemos acercarnos a los rendimientos, pero cuando uno ve el informe del INEE que tuve el gusto de estudiar, no soy psicometrólogo ¿se dirá así?, si nosotros vemos los resultados crudos del aprendizaje de las escuelas, sin tomar en cuenta el contexto socioeconómico y otras dimensiones, la lectura puede cambiar. Se ha generado en México a través de Tabaré Fernández una tesis doctoral en el Colegio de México que es extraordinaria,

porque lo que hace Tabaré es ver cuántos factores están influyendo para llegar a ese pequeño dato de: *esta escuela sacó en promedio tal nota*.

Hay un cuadro en el informe del INEE, no me acuerdo de la página, ya sería demasiado para mí, en que dice que el ochenta por ciento de las escuelas rurales de México, están ubicadas en contextos socioeconómicos desfavorables o muy desfavorables y que por el contrario el ochenta por ciento de las escuelas privadas o particulares en México están ubicadas en contextos favorables o muy favorables.

No tengo *a priori* ningún problema con que haya escuelas particulares, pero no creo que sean mejores o peores por ello, igual que las públicas no creo que sean mejores o peores por ello, lo que sí es cierto es la fuerza del impacto de la condición socioeconómica, el impacto del capital cultural de la familia, el impacto del ambiente escolar, el impacto del tipo de profesor. No soy psicólogo, ni pedagogo, ni didáctico, a mí me enseñaron a hacer sociología y la sociología es la ciencia, dicen mis profesores, el profesor Max Weber dice que es la ciencia o la disciplina que trata de responder por qué un hecho se dio así y no de otra manera.

Y la pregunta es muy buena, porque imaginemos el experimento, como experimento mental, que a una escuela muy deteriorada, en una comunidad muy deteriorada llega un muchacho al que pudiésemos medir con cuatro de origen y sale con siete; y a una escuela de otra zona con mejores servicios, llega con ocho y sale con nueve, el valor agregado de la escuela en el primer caso es mucho mayor, aunque la nota comparativa los ponga en prelación distinta.

Por ello uno de los hallazgos más importantes de la tesis de Tabaré, de la que fui lector, es que en países con menor desigualdad la escuela juega menos, en paí-

ses tan desiguales como México la escuela juega muchísimo.

Trabajemos para que la escuela pública, la escuela de todos los mexicanos, sea de la mejor calidad posible y utilicemos la evaluación como instrumento para retroalimentar esa escuela. A diferencia de Ángel que se centra mucho en el docente, de lo cual me surge una idea de círculo, yo propondría que pensáramos en una elipse, en la cual los dos focos sean estudiantes y profesores. Le hemos apostado mucho a los profesores, por lo menos en la educación superior y lo que hemos generado –por lo menos en la educación superior– es un conjunto de soberbios que con el doctorado consideran a sus estudiantes indignos de su sabiduría.

Creo que debemos pensar en la forma de una elipse con dos focos, y para terminar, perdonen, hay que trabajar mucho para mejorar la calidad de nuestras escuelas, y la evaluación nos sirva para que como sociedad ponderemos lo que agregan las escuelas: la telesecundaria, la escuela atendida por Conafe; no es simplemente escalar y ver que los del Colegio Americano salen mejor que los de una escuela indígena de Chiapas, pues eso resulta intrascendente.

Pero estudiar cómo en la comunidad indígena de Chiapas, los muchachos pasan de un nivel a otro en su capacidad de pensar, de divertirse, de pasarla bien en la escuela, eso sí me parece que es el sentido profundo de la evaluación; no directamente transformarla porque nos podríamos convertir en agentes para que los muchachos respondan pruebas, y eso sería muy feo e indeseable, estoy seguro, ¿verdad?

La otra opción es que todos los que estamos interesados en la educación, si queremos que ésta mejore, debemos pelear con el mismo ahínco contra la oprobiosa y vergonzosa desigualdad social que tiene este país. Es ineludible un compromiso



político para cambiar un modelo de desarrollo que cada vez nos separa más.

Y luego le echamos la culpa a la escuela, y decimos: “fábrica de desempleados, produce reprobados”. No friegue señor, hace veinte años no hay empleo, y el empleo se consigue más por el color de piel y las relaciones en el *Tec* que por el conocimiento de las Leyes de Newton. Todo esto demanda nuestra reflexión.

### **JORGE HERNÁNDEZ URALDE**

#### **¿Por qué los reactivos de los exámenes –supongo que los exámenes del Ceneval– siempre son de opción múltiple?**

La respuesta sería la siguiente: en primer lugar, no todas las pruebas que se diseñan en el Ceneval son de opción múltiple, tenemos de pregunta abierta, de ensayo, incluso pruebas que exigen la entrevista de expertos. Dependiendo de la naturaleza de la evaluación se definen las características de los instrumentos de medición.

En el caso particular de los exámenes nacionales, la justificación fundamental son las condiciones bajo las cuales se realizan estas aplicaciones, su carácter masivo y el corto tiempo que reclaman las instituciones en la presentación de los resultados.

Y dentro de las pruebas de tipo objetivo, los reactivos de opción múltiple son los que han demostrado ser más confiables, y permiten una mayor flexibilidad en el formato de presentación.

### **DR. ARTURO DE LA ORDEN**

Tengo cuestiones de tipo doctrinal, político, social, de estas grandes dimensiones que han estado pululando en torno a nosotros.

Me preguntan también si puedo recomendar un libro de evaluación, pero no sé si desean un libro para aprender los procesos de cómo evaluar; o si nuevas concepciones de evaluación, que es en donde yo me movía en la interpretación. En cualquier caso, quien lo haya escrito se acerca a mí, le doy mi dirección y le envío una lista completa.

También puede usted consultar la Enciclopedia de Evaluación Internacional: tiene ahí un montón de temas con su biografía, seleccione los aspectos de la evaluación que le interesen, o las últimas obras del mundo anglosajón, que es nuestra nodriza pues nos está alimentando en...no tanto la teoría, pero sí en elementos para realizar una buena teoría que deberíamos ser capaces de interpretar a la luz de nuestra realidad.

En el fondo los problemas de evaluación pueden considerarse generales y generalizables, de lo contrario no tendríamos un conocimiento relativamente científico de la evaluación, pero como se ha visto, la evaluación está metida justamente en todo lo que es el núcleo educativo, social y problemático de una sociedad. Y para eso no hay ningún libro que pueda resolverlo de momento, probablemente lo tendrán que escribir ustedes en México.

### **FELIPE MARTÍNEZ RIZO**

Voy a resumir todo lo que pueda; hay varias preguntas que se refieren a la difusión de los resultados, si se van a poder consultar en Internet y la respuesta es: sí, desde luego que sí, incluyendo las ponencias del evento e inclusive, en relación con esto que se le preguntó al doctor de la Orden, esperamos pronto tener el catálogo de la biblioteca del INEE consultable en Internet, para quienes se interesen en estos temas, puedan acceder fácilmente a los mismos.

Otra serie de preguntas, se relacionan con: si se pueden evaluar los valores, si se piensa evaluar a las escuelas privadas, si se buscará sinergia entre diferentes tipos de evaluaciones.

Y lo respondo refiriéndome al concepto manejado ayer sobre el sistema nacional de evaluación, con la idea de que la evaluación es un concepto muy amplio y complejo; incluye gran número de modalidades, enfoques, aspectos, metodologías, lo cual hace absolutamente impensable que una sola instancia se haga responsable. Por ello debe considerarse que la evaluación se realice por varias instituciones, pero con un sentido de congruencia y coordinación que impidan caminar en direcciones distintas.

Esas diferentes instancias incluyen de una manera fundamental al maestro, y creo que el maestro es la instancia básica, no sólo para la enseñanza sino también para la evaluación, considerando –coincido ampliamente con Ángel Díaz Barriga– la evaluación como un elemento del proceso de enseñanza.

Si ustedes ven la parte final del informe, el primer informe que difundió el Instituto, al que se refirió Manuel Gil desde noviembre del año pasado, que está en Internet y que se les ha distribuido en forma impresa a muchos ahora, termina con una cita de un profesor de la Universidad de Harvard, que dice que en el corazón de un buen sistema educativo hay maestros profesionales y familias informadas.

Precisamente, creemos que una concepción equivocada, que refiero como una visión *Orwelliana* de la evaluación, en el sentido del *Big Brother*; es una concepción según la cual la evaluación sustituye al maestro, porque parte de que él no tendría la capacidad suficiente, no sabe, no es competente. Por tanto, deben hacerse evaluaciones externas todo el tiempo, para saber cómo están los niños o cómo

están las escuelas: que alguien vaya, tome medidas y ponga orden.

Yo estoy convencido también, como Ángel, de que el maestro es la pieza fundamental y la evaluación será positiva si apoya para que el maestro haga mejor su trabajo, y contribuye a su profesionalización.

Siguiendo la analogía que se mencionó ayer por Salvador Malo, la evaluación proporciona unos elementos; es un laboratorio de análisis clínicos, que le dice al maestro, por ejemplo, que el niño trae el azúcar alta, pero el laboratorio no prescribe un tratamiento porque no conoce al niño, no sabe si es hombre o mujer o si tiene ocho años ó 14, y con el ejemplo que tomé del azúcar, más bien si es un adulto o joven el paciente, si es una señora embarazada o si es un anciano o un niño, o tiene diabetes muy tempranamente, y eso sólo lo puede hacer quien conoce al niño.

El maestro puede recibir elementos de la evaluación, que como mencionaba José Luis Gaviria, le pueden aportar datos que no extrae de su propia aula, –porque no tiene todos los elementos para comparar a nivel nacional y sus comparaciones son más reducidas–, pero nada sustituye, no podría hacerlo, toda la riqueza de elementos que él posee.

Se pregunta sobre el papel del INEE en cuanto a la mejora de la calidad educativa, y también en este sentido coincidiría mucho con Manuel; la postura que siempre hemos tenido es que el INEE no debe ni quiere tener autoridad. Se me pregunta si puede influir, intervenir, pero también se me pregunta si puede exigir. No, nosotros no somos una autoridad, no tenemos esa facultad. Es importante deslindar los terrenos, precisamente para que la evaluación cumpla su función claramente.

Uso otra metáfora o comparación si ustedes quieren: en un avión hay muchas piezas, alas, colas, fuselajes, motores...y tiene también radar; cada pieza tiene su



función. En este caso el INEE debe de ser radar. Si un avión no vuela y le ponemos radar nuevo, sigue sin volar, porque el radar no hace volar al avión, para eso necesitaría motor nuevo u otra pieza. Pero si tiene un motor muy bueno o varios motores muy buenos y no tiene buen radar, va a chocar por ahí, porque la potencia de sus motores no podrá ser dirigida adecuadamente si no se tienen los otros elementos.

Entonces, los organismos de evaluación no deben ser instancias de decisiones, no deben conducir el currículum; efectivamente, totalmente de acuerdo con Ángel: el currículum tiene que ser establecido de otra manera y los organismos de evaluación no deben de pretender eso, no deben pretender sustituir a la autoridad, sino dar buena información para que cuenten con ella desde los maestros y padres de familia. Reimers dice: buenos maestros, maestros profesionales y familias informadas conscientes de lo que pueden esperar y exigir de la escuela, y capaces de aportar elementos a la escuela para su mejora.

Estamos totalmente de acuerdo con este concepto y con un concepto de cultura de evaluación muy rico, no reducido a información y con un concepto de rendición de cuentas muy positivo y muy amplio, que no es simplemente hacer *ranking*.

Hay otra cita que utilizo en el informe –a mí me gusta mucho– refiere que la postura según la cual mejorar la educación es muy fácil, pues basta hacer unas pruebas, aplicarlas, hacer un *ranking* y ya, parten de un supuesto falso: que hacer buena educación en cualquier contexto es fácil. Si somos conscientes de la dificultad de hacer buena educación en contextos diversos, tenemos que ser muy prudentes y respetuosos para que la evaluación no sea destructiva.

Se me hacen varias preguntas sobre evaluación con enfoques cuantitativos y

cuantitativos, disyuntiva que deberíamos haber superado hace buen rato. Se reflexiona sobre las razones de los bajos resultados de alumnos indígenas y se hacen consideraciones similares a las que ya se hicieron sobre la necesidad de que las pruebas tengan en cuenta las diferencias culturales, etcétera.

Quiero decir que comparto todo esto, precisamente la visión que el Instituto quiere llevar adelante es de una evaluación muy rica, la cual no sea reduccionista, simplista y precisamente eso es algo donde requerimos un esfuerzo muy complejo.

Los 17 criterios que se presentaron en este trabajo sólo se refieren a comparabilidad, nada más, pero no se piensa que es lo único que demanda atención, de ninguna manera.

Como decía el doctor de la Orden, dado que la comparabilidad empieza desde el diseño de las pruebas, de alguna manera los criterios recorren todo el camino que se tiene que hacer, pero están centrados en ese aspecto.

Ayer mencionaba que el Instituto pretende tener evaluaciones de muy buena calidad no simplemente por un buen deseo, sino porque pretende seguir un modelo de calidad muy rico que incluye criterios muy amplios, no sólo de comparabilidad sino también de validez e implica que no haya sesgo cultural, se respeten las diferencias y se tomen en cuenta para evaluar. También considera que el currículum es muy rico y que si bien todo se puede evaluar en principio, no todo se puede evaluar de la misma manera ni a cargo de la misma instancia.

La evaluación que puede hacer el Instituto, para aportar elementos al conjunto del sistema educativo, es muy particular en algunos aspectos, y debe desarrollarse muy bien para enriquecer en lugar de desorientar o distorsionar. También debe rea-

lizarse en otros ámbitos: desde cada aula con su maestro, cada escuela, cada supervisor, etcétera, para que el sistema en su conjunto funcione.



## **Comparabilidad de los Resultados de las Evaluaciones**

Preguntas y  
Respuestas





Los especialistas en evaluación explicaron al público por qué es comparable o no el resultado de una evaluación.



Integrantes del panel en el que se abordó el tema de la Comparabilidad de los Resultados de las Evaluaciones.

QUINTA MESA REDONDA

## **Las Comparaciones Entre Entidades: Alcances y Límites de los *Ranking***

**Ponente:**

Dr. Eduardo Backhoff Escudero

**Comentaristas:**

Dr. Eduardo de la Garza Vizcaya

Dr. José Ángel Pescador Osuna

Dr. Roberto Rodríguez Gómez

**Moderador:**

Mtro. Andrés Sánchez Moguel





## Las Comparaciones entre Entidades, Alcances y Límites de los *Ranking*<sup>1</sup> Por EDUARDO BACKHOFF ESCUDERO\*

**A**bordaré un tema que ha causado mucha polémica en México y en cualquier sistema nacional de educación: *ranking*, como usualmente se le denomina en inglés, en español es el ordenamiento de las unidades que se tienen o se pueden comparar. Un punto importante a tratar es la dificultad que tienen las comparaciones entre unidades, así como el interés nacional que despiertan los resultados de las evaluaciones educativas, y los estudios internacionales del logro del aprendizaje.

Respecto a los *ranking* debemos decir que se parecen mucho los de deportes a los de la educación, pero veremos los problemas que existen en las comparaciones, en términos de los ordenamientos de las entidades federativas.

Trataremos también en esta plática, de lo difícil que es el tema debido a las inconsistencias en los resultados, los errores en las evaluaciones, la influencia que tiene el contexto en los resultados del logro de aprendizajes, y algo muy importante: qué evalúan las pruebas estandarizadas del logro educativo.

### Gran interés por los resultados

Es en los últimos años, que en México los resultados de las evaluaciones educativas nacionales e internacionales han despertado gran interés en el público, esto se debe, entre otros aspectos, a la importancia de saber si la educación ofrecida por

un sistema educativo es de mejor o peor calidad que la del vecino, o de otra entidad federativa, o de un país en especial.

Sin embargo, la dificultad radica en que comparar de manera confiable la calidad de dos sistemas educativos de dos regiones o dos escuelas, en ausencia de mecanismos apropiados es injusto, inválido y no pertinente, por lo cual, uno debe poner atención en la calidad del instrumento y de las evaluaciones mismas.

Por tal razón, hace medio siglo se iniciaron proyectos internacionales, los cuales permiten hacer comparaciones consistentes, aplicando pruebas de aprendizaje a distintos alumnos de cierta edad y/o de determinados grados.

Los proyectos internacionales más conocidos son PISA, TIMSS y ahora el Laboratorio Latinoamericano, con los cuales desde 1958 se iniciaron los primeros intentos por poner de acuerdo a distintos países participantes, para tratar de hacer evaluaciones equivalentes entre ellos, llevar a cabo algún tipo de comparación, y conocer el grado de avance de su sistema educativo.

Sin embargo, a pesar de que hace casi cincuenta años empezó este programa, es a partir de la segunda mitad de la década de 1990, que la OCDE comenzó un proyecto similar, el cual ahora se conoce como PISA y cuyas pruebas se han aplicado por primera vez en el año 2000 y por segunda ocasión en el 2003, y podría decir que igualmente sucedió con TIMSS en 1995,

<sup>1</sup> Dado que no existe una traducción exacta de la palabra *ranking*, se decidió utilizar el término original en lugar de aproximaciones como *tabla de posiciones* u *ordenamientos*.

\*Director de Pruebas y Medición del Instituto Nacional para la Evaluación de la Educación.



cuyos resultados no se dieron a conocer oportunamente en México.

Es desde 1995 y más en estos últimos años, que se producen pruebas más o menos comparables, las cuales se pueden sustentar en términos de su calidad técnica. Asimismo, podemos decir que apenas a fines del siglo pasado y a inicios de éste, internacionalmente tenemos evaluaciones que aún se debaten a niveles nacionales e internacionales, y las cuales comienzan a consolidarse.

Las comparaciones internacionales de estos proyectos han permitido ordenar a los países en función de su puntuación promedio, obtenida por los alumnos en cada una de las pruebas. Este procedimiento ofrece una visión que parece clara e indiscutible de la calidad del sistema educativo de los países participantes, y se ha hecho muy popular.

### **¿Los ranking deportivos son similares a los de evaluación educativa?**

Vamos a ver ¿por qué los ranking deportivos son similares o parecen ser similares a los de evaluación educativa? En un deporte es sencillo plantearlo, pues para sobresalir en él es necesario tener un rasgo en cuanto a destreza, rapidez, fuerza, precisión o en algún aspecto destacado internacionalmente respecto de un deportista con otros.

Los resultados deportivos definen posiciones en los competidores de manera oficial, clara e inequívoca, más allá de la distancia precisa que puede haber entre ellos, es decir, por décimas, milésimas, diezmilésimas de segundo, uno puede quedar en segundo o tercer lugar, independientemente de ello siempre hay un ganador y un perdedor. Entonces se glorifica al triunfador y se critica muchas veces despiadadamente al perdedor, aunque

sea también excelente y aunque haya perdido por poco tiempo.

Los *ranking* de los resultados educativos se han basado en los deportivos, nuestros titulares de prensa así lo ilustran: México, penúltimo lugar del mundo; México primer lugar de América Latina; Vence México a Brasil, pierde con Hong Kong; Cae México del lugar 31 al 34; México mejor, por encima de Argentina y Chile.

Si ustedes le quitan el contexto evaluativo a esto, parece una nota completamente deportiva y es que hay una transferencia muy sencilla y muy simple de ambos *ranking*.

Sin embargo, se requiere conocer cuáles son los problemas de las instituciones educativas, ¿por qué no deben hacerse esas transferencias sencillas y simples con los deportes? Daré algunos ejemplos:

*Primero.* La educación no es unidimensional, como en el caso de los deportes, la educación es multidimensional; un alumno, una escuela, una entidad federativa pueden tener diferentes y distintos desempeños, unos en algún aspecto, otros en otro, y cuando promediamos todos esos desempeños obtenemos un solo puntaje. En realidad no podemos estar hablando de que el puntaje promedio nos dé un indicador justo y equitativo.

Las evaluaciones tampoco alcanzan un grado de precisión tal, el cual permita diferenciar con claridad el nivel de desempeño alcanzado por alumnos y escuelas o las mismas entidades federativas. Ya comentó Felipe Martínez Rizo lo difícil que es construir buenos instrumentos; les puedo decir que es muy, pero muy difícil construir elementos precisos y confiables a lo largo del tiempo.

Se debe considerar que la equidad es un aspecto muy importante al juzgar y al hacer comparaciones y, no debemos dejar de vincular el contexto. Si ustedes me preguntaran en este momento: ¿En el país

cuáles son los mejores estados o entidades federativas? Sin tener una bola mágica les podría decir cuáles son, y no tengo que realizar una evaluación, simplemente debo considerar los niveles de pobreza y marginalidad existentes en cada una de las entidades federativas, y les puedo asegurar que el promedio de equivocación oscilaría entre 10 ó 15 por ciento.

Con respecto de las pruebas de lectura, por darles un ejemplo, podemos decir que las primeras tres entidades en 2003, evaluando sexto de primaria fueron Distrito Federal, Coahuila y Sinaloa. Y por otro lado, las últimas son Veracruz, Oaxaca y Chiapas. Vemos entonces que el Distrito Federal está en primer lugar y Chiapas en último, en términos de contexto sonaría lógico, pero no debemos olvidar que esas son puntuaciones promedio; bastaría para que en los periódicos se publicara: *Distrito Federal cuenta con el mejor nivel educativo y Chiapas con el peor*, pero todos sabemos que es una falsa idea.

Otros ejemplos serían: Aguascalientes tiene un lugar 27 y un tercero; Baja California un 26 y un sexto; Coahuila un segundo y un 28, y Durango un 29 y un cuarto. Queriendo decir esto que no hay uniformidad, y todo depende de la materia y/o grado evaluados.

También sabemos la existencia de diferencias entre primaria y secundaria; Colima, Durango y Quintana Roo obtienen buenos lugares en el nivel de secundaria, mientras que en primaria esos resultados son más bajos. Esto nos lleva a un sesgo en las interpretaciones, el carácter multidimensional de las evaluaciones hace que el resultado de una sola materia o un solo grado pueda ser engañoso.

La tentación de una entidad por destacar sus mejores resultados es natural, pero, si se ignoran los menos favorables, se estará presentando en realidad una verdad a medias. Entonces, es muy riesgoso hacerlo.

## El margen de error

Hay otro aspecto, al que voy a dedicarle, por desgracia, menos tiempo, me refiero al margen de error de los resultados. Sólo les quiero comentar que toda medición –como lo es una prueba y sobre todo las pruebas de rendimiento– tiene lo que llamamos *margen de error* y éste puede ser muy grande, lo suficiente como para no poder sacar conclusiones simples. Éste depende de muchos factores, entre ellos, calidad de la prueba, tipo de muestreo y condiciones de aplicación.

Una buena medición debe incluir siempre el *margen de error*, es decir, *error estándar de medición*. Y para interpretar correctamente las puntuaciones debemos tener en cuenta no sólo los aciertos, sino también los errores estándar. Por ello es muy difícil hacer *ranking* simple, porque siempre debemos considerar este factor.

El contexto es un aspecto muy importante a valorar, los resultados se explican en gran medida por las diferencias de éste entre los alumnos. Una gran mayoría de las escuelas privadas se sitúa en un contexto socioeconómico muy favorable, alrededor del ochenta por ciento de este tipo de escuelas están en dicho rubro o en el favorable. Es decir, el 96 por ciento de las escuelas privadas tiene condiciones muy favorables.

Por otro lado, escuelas indígenas y cursos comunitarios se sitúan en más del 85 por ciento, en contextos desfavorables o muy desfavorables. Las diferencias contextuales en las entidades federativas deben tomarse en cuenta a la hora de realizar comparaciones; la proporción de escuelas públicas y privadas, urbanas y rurales, cursos comunitarios y escuelas indígenas que hay en las entidades del país es muy distinta; en el Distrito Federal no hay escuelas rurales ni indígenas ni cursos comunitarios; en cambio existen más escue-



las privadas, en proporción, que en cualquier otra entidad federativa.

En los estados del sureste del país, principalmente Chiapas, existe un elevado número de escuelas rurales indígenas y muy pocas privadas. Entonces estoy comparando peras con manzanas o, si fuera en el aspecto deportivo, estaría comparando primera fuerza con tercera.

Esto es algo que se debe considerar a la hora de hacer comparaciones, saber cómo hacerlo, ya sea por tipos de escuela, estratos socioeconómicos o nivel de marginalidad, por ejemplo, para que comiencen a tener sentido los datos; si no, es injusto e inequitativo.

### Las pruebas estandarizadas de aprendizaje

Una última consideración ¿qué miden las pruebas estandarizadas de aprendizaje? Lo pregunto porque los *ranking* parten de una premisa falsa, la cual señala que los resultados de las pruebas estandarizadas de aprendizaje son el producto de la calidad de las escuelas ¡no hay cosa más falsa que esto! y lo demuestro fácilmente.

Las pruebas de rendimiento, como las pruebas estandarizadas de PISA, TIMSS y aun las del IDANIS, en realidad miden tres tipos de rasgos de las personas: condición biológica, dotación biológica –inteligencia, capacidades y destrezas con las que cuenta el niño– contexto sociocultural: lo que aprenden de su medio, de la biblioteca, de sus papás o de la ausencia de ellos y, por supuesto, lo que aprenden en las escuelas. Esas tres dimensiones son casi indivisibles.

Sin embargo, si queremos evaluar la calidad del sistema educativo, deberemos medir sólo lo que los estudiantes aprenden en las escuelas. Deberíamos suprimir todas aquellas preguntas que miden habilidades intelectuales, conocimientos o destrezas adquiridas en casa o en su entorno. De otra manera, estaremos evaluando no al sistema educativo en particular, sino a la sociedad mexicana en general.

¿Por qué digo a la sociedad mexicana? Porque la sociedad mexicana tiene un contexto, una cultura, un nivel de marginación, y estamos evaluando aspectos aprendidos dentro y fuera de las escuelas.

### Conclusiones

- La evaluación educativa que realiza el INEE debe ser una herramienta valiosa para apoyar los esfuerzos de mejora de la calidad del Sistema Educativo Mexicano.
- Es necesario que dicha evaluación sea de alta calidad y sus resultados sean interpretados adecuadamente, como ya se dijo anteriormente.
- La comparabilidad de los resultados entre entidades federativas debe sustentarse en argumentos técnicamente válidos, y no en posiciones simplistas como se ha hecho últimamente.

Este trabajo ha pretendido mostrar lo inadecuado de presentar y analizar los resultados en ordenamientos simples de escuelas, entidades o sistemas, inclusive. Esperamos que la ponencia ayude a corregir el rumbo y resulte un aporte serio a lo que ahora se denomina cultura de la evaluación.



## PARTICIPACIÓN DE LOS COMENTARISTAS

EDUARDO DE LA GARZA VIZCAYA\*

El documento que se nos ha pedido comentar, las *Comparaciones entre Entidades, Alcances y Límites de los Rankings* elaborados por el INEE, es un inicio conveniente de la reflexión sobre el tema.

Su propósito, como el mismo documento lo expresa, es mostrar lo inadecuado de presentar resultados mediante un ordenamiento simple con base en las puntuaciones promedio, sin hacer matices ni salvedades.

El documento logra convencer de la inconveniencia de proceder así, sin embargo, suscita preguntas, las cuales quedan aún sin contestar. No niega la posibilidad de hacer *ranking*, sino plantea cómo debemos realizarlos de manera que sean significativos para la toma de decisiones.

### Las pruebas internacionales de rendimiento

El documento inicia refiriéndose a las pruebas internacionales de rendimiento, aplicadas a muestras de alumnos de distintos países en las cuales ha participado México en los años recientes. El TIMSS, que podría traducirse como Tendencias en los Estudios Internacionales de Matemáticas y Ciencias de la Asociación Internacional para la Evaluación del Logro Educativo, y el Programa para la Evaluación Internacional de los Estudiantes, conocido como PISA, de la OCDE. Ambos programas ordenan a los países en función de la puntuación promedio obtenida por los estudiantes.

El documento menciona que la popularidad de presentar y entender los resul-

tados mediante estos ordenamientos, es alimentada por la costumbre de hacerlo en otros ámbitos, como el deportivo. En el ámbito deportivo se llega a ordenamientos únicos basados en criterios decisivos, por ejemplo el número de goles anotados y recibidos que definen la posición de los competidores de manera clara e inequívoca, y son acompañados de la glorificación del triunfador y la crítica del perdedor.

Con ligereza, en no pocas ocasiones se transfieren estos rasgos a las evaluaciones educativas; hice el ejercicio de imaginarme un PISA del fútbol, pensé que cada tres o cuatro años, una oncena internacional recorre el país midiéndose con los equipos de los distintos barrios en ciudades y pueblos, en una hora los equipos deben meter el mayor número de goles, recibir el menor número de ellos, conservar el mayor tiempo el balón, hacer el mayor número de disparos al arco y hacer el menor número de faltas.

Deberíamos cuidar que todos los equipos del barrio jueguen a la misma hora, proporcionarles el balón cuando menos seis meses antes para que se acostumbren a él, desde luego no sería relevante si algunos equiperos juegan descalzos. La dificultad de hacer un *ranking* de los resultados por países, salta inmediatamente.

El documento aclara más adelante que las evaluaciones educativas, a diferencia de las deportivas, tienen rasgos propios que hacen inadecuada la presentación de los resultados en la misma forma; primero, las evaluaciones educativas son multidimensionales, es decir, atienden niveles de desempeño en diferentes áreas: matemá-

\*Académico de la UAM y Consejero Técnico del INEE.



ticas, lectura o ciencias; segundo, el grado de precisión de las evaluaciones hace difícil afirmar que un estudiante se ubica ligeramente por arriba o por debajo del otro; y tercero, las consideraciones de equidad exigen matizar los resultados con aspectos del contexto de los alumnos, de las escuelas o de los sistemas.

Con base en los resultados de las pruebas nacionales de lectura y matemáticas aplicadas en 2003, de los últimos tres años de primaria y los tres de secundaria, la parte central del documento muestra la dificultad de presentar e interpretar dichos resultados, en ordenamientos simples de las entidades federativas a partir de las puntuaciones promedio.

### **Multidimensionalidad, error estándar de medición y factores del contexto**

En primer lugar, hace ver que la multidimensionalidad de los resultados de desempeño, lectura y matemáticas, en seis grados distintos, relativiza cualquier ordenamiento parcial, por ejemplo lectura en sexto grado de primaria; no hay total coincidencia en los dos ordenamientos aunque sí se muestran algunas tendencias. Saltan a la vista inmediatamente aspectos contrastantes, en donde hay entidades que tienen el lugar 27 en un resultado y el tercero en otro.

En segundo lugar, el documento se detiene en el *margen de error* que presentan los resultados; la precisión de los resultados depende de varios factores, entre los cuales se haya la calidad del instrumento, el número de observaciones realizadas y el cuidado en el control de las condiciones de la aplicación.

Además, las evaluaciones que hacen mediciones sobre muestras, deben cuidar la confiabilidad de éstas, de ahí se deduce que toda medición necesita estimar su *margen de error* con una proba-

bilidad determinada: el *error estándar de medición*.

Los buenos reportes de resultados deben dar a conocer, no sólo los promedios alcanzados por las muestras, sino el error de medida en el procedimiento, lo cual conduce en el caso de los resultados de las pruebas nacionales, a que con un 95 por ciento de probabilidad, la puntuación real de una entidad federativa pueda estar en un rango de dos errores estándar de medida, arriba o abajo de la media de la propia institución.

Esto quiere decir que entidades con resultados por encima de otra, ubicadas dentro del rango de los dos errores estándar de medida, podrían estar técnicamente empatadas con la primera. El documento propone presentar los resultados añadiendo los valores máximos y mínimos de las puntuaciones, es decir, estableciendo el rango en que los resultados podrían moverse hacia arriba o hacia abajo para ayudar a una mejor lectura o interpretación.

Una vez presentados los resultados de esta forma, resulta claro que las entidades se agrupan en tres: las que están por encima de la media nacional; las que están por abajo, y el mayor grupo sin diferencia significativa respecto a ella. Finalmente, los ordenamientos que identifican a las entidades por encima o debajo de la media, con base en cierta probabilidad, parecen más robustos.

Por último, se analizan los factores del contexto, se afirma que hay diferencias considerables en el tipo de servicio más que entre las entidades federativas. Así, hay diferencias entre escuelas privadas y públicas, y en éstas, entre urbanas y rurales, cursos comunitarios y, sobre todo, escuelas indígenas. Lo anterior tiene su explicación en los contextos favorables o desfavorables de las escuelas.

Asimismo, se debe considerar que la proporción de cada una de estas escuelas

en las entidades del país es muy distinta; de ahí se sigue la necesidad, atendiendo a la equidad, de hacer comparaciones que incluyan sólo escuelas del mismo tipo.

A estas alturas, no parece posible evitar presentar los resultados mediante *ranking* por entidad federativa, sobre todo ahora que los estudios internacionales nos familiarizan con otro tipo de éstos: los de países.

Convendría, por otra parte, analizar con mayor detalle las consideraciones que dichos estudios hacen para justificar la presentación de los resultados mediante *ranking*. En éstos pareciera que las instituciones dejan a las lecturas las interpretaciones y matices que incluyen los contextos en cada país.

Una conclusión importante a partir de las consideraciones del documento, es la necesidad de cuidar las condiciones de calidad de las pruebas, para reducir al máximo el error estándar de medida y obtener así mediciones más precisas.

No sólo la lectura e interpretación de los *ranking* que haga el público más amplio se ve afectada por ese factor, sino que incide directamente en la toma de decisiones de los responsables educativos, pues poco pueden éstos inferir de una evaluación a otra cuando los rangos de lectura de los resultados son amplios.

### Formas idóneas de reportar resultados

Parece también conveniente estudiar las formas idóneas de reportar los resultados a partir de las situaciones del contexto mexicano. Esto me parece que es el mayor reto que se tiene.

Considerar, por ejemplo, la oportunidad de reportar medianas, en lugar de medias, como lo hace el Laboratorio Latinoamericano de la UNESCO en alguno de sus informes. Se observa que esto da mejor cuenta de las desigualdades; también acudir a medidas de la dispersión de la población y el lugar relativo manifestado en deciles del conjunto muestreado.

Considero, por tanto, que la pregunta permanece: ¿cómo hacer el *ranking* de tal manera que sea apropiado para la información amplia y la toma de decisiones? La pista, me parece, está clara en el contexto que se debe recuperar.

El documento, concluyo, es un inmejorable punto de partida para ahondar en la reflexión y discusión, sobre los alcances y limitaciones de hacer *ranking* de entidades federativas a partir de los resultados de desempeño educativo; sobre todo, encuentro su mayor utilidad en la de ilustrar la dificultad de hacer lecturas superficiales sobre los mismos.

---

JOSÉ ANGEL PESCADOR OSUNA\*

**A**ntes de hacer mis comentarios al documento de Eduardo Backhoff, haré algunas aclaraciones.

Soy miembro del Consejo Técnico. Éste se compone por 16 personas, tres no asistieron a esta última reunión, 13 estamos

aquí y voluntariamente aceptamos participar para hacer comentarios sobre los documentos. Somos representantes de un Organismo Colegiado, independiente, autónomo, tenemos nuestros puntos de vista y deseo dejarlo perfectamente claro, para

---

\*Docente, experto en educación, exSecretario de Educación Pública y miembro de El Colegio de Sinaloa, y del Consejo Técnico del INEE.



que no se piense que venimos a defender una postura del Instituto.

En segundo lugar, quiero hacer el comentario, voy a hablar de cosas muy sencillas, porque parece que en los dos días hemos hablado de todo lo que es difícil en materia de evaluación.

Creo que los maestros, y aquí hay muchas personas que están ligadas con el sector educativo, me corregirán o a lo mejor reafirmarán lo que digo. Evaluaciones se han hecho por mucho tiempo, a lo mejor no con el rigor técnico ni con la calidad que ahora se plantea, pero nos han servido para tomar decisiones; probablemente, no siempre las más apropiadas, pero siempre se ha hecho evaluación.

Ahora que se está hablando tanto de pruebas, recuerdo que en mis años en la Normal, terminábamos un curso de psicotécnica pedagógica donde veíamos cómo se elaboraban las pruebas. Los criterios de confiabilidad, validez y objetividad siguen siendo los mismos; probablemente, como se decía muy bien esta mañana, la psicometría ha avanzado significativamente, pero quiero decirles que no es tan difícil pensar en tener especialistas en este campo, creo que es mucho más difícil estar frente a grupo en el aula.

Y un tercer comentario, el cual me parece importante: decía Eduardo Backhoff, a propósito de las comparaciones hechas, algunas con limitaciones, con riesgo; sí, pero se han hecho todo el tiempo.

¿Saben por qué PISA comenzó en el 2000? Porque la OCDE se preocupó durante muchos años de los indicadores de economía y de establecer ordenamientos, no me gusta usar la palabra *ranking*, ordenamientos entre países, precisamente en términos de otras variables que resultaban más importantes.

De manera tal que, como sucede en otros campos, llegamos tarde en educación, vean el periódico de hoy, toda la nota

se la lleva la seguridad pública y los mil millones que se le asignarán, y prácticamente pocas reflexiones en torno a un asunto tan importante como es la educación.

### **Sí existe una cultura de la evaluación**

Eso me lleva a un primer comentario, nunca hay que ver la evaluación de la educación exclusivamente en términos de lo que significa calidad, –creo que es nuestra principal tarea– sino cómo afecta al bienestar de la población. Y ahí tenemos que admitir que el Instituto, como muchas otras instituciones, tiene que aceptar que hay diferentes puntos de vista. Por eso ustedes han escuchado aquí, que todavía se sostienen opiniones según las cuales el Instituto debería tener una conformación distinta. Yo no las comparto, pero debo admitir que tienen derecho a decirlo.

Como se decía esta mañana, los resultados obtenidos hasta la fecha y las evaluaciones conocidas, han tenido tantas limitaciones que de alguna manera generan desconfianza, pero al mismo tiempo una cultura de la evaluación.

No comparto la tesis de la no existencia de una cultura de la evaluación, en todo caso diría *a lo mejor hay una cultura equivocada o a lo mejor una subcultura*. Cuando queremos hablar de los contrastes en términos de seguridad, hablamos de la subcultura de la violencia o el narcotráfico.

Probablemente deba plantearse en esos términos y por eso es muy importante lo que el Instituto está proponiendo hacia el futuro en materia de estas correcciones, fundamentada en dos criterios, ¿qué es calidad? y ¿qué es perfeccionamiento en cada uno de sus trabajos?

Quiero decirles que comparto plenamente el trabajo de Eduardo Backhoff, es riguroso y exacto, no hay ninguna duda de que el planteamiento es cierto; pero hace muchísimos años conocemos el *error es-*

*tánder*, no es novedoso en educación. Lo que él está agregando es que las comparaciones realizadas hasta la fecha, por no considerarlo, tenían un sesgo que podría orientar a tomar malas decisiones. Eso me parece un apunte muy importante.

Es más significativo lo que dice al final, si nos preocupamos por las pruebas, si hemos hablado durante casi dos días de estos instrumentos tan importantes para la evaluación, pero resulta que no miden exactamente lo que estamos tratando de medir, esa es una de sus conclusiones más importantes para tratar de orientar una decisión distinta.

### **¿Qué pasa con las comparaciones?**

Hechas estas aclaraciones, déjenme decirles qué pasa con las comparaciones. Mi primera opinión es precisamente que las comparaciones entre las entidades siempre se han dado, no hay nadie –yo mismo hice un trabajo en 1972, en que comparé a todas las entidades federativas– que nos dijera cuál era el *status* de ellas. Claro, no teníamos indicadores tan sofisticados, como el índice de desarrollo humano o el índice de marginación, hoy actualizados año con año, pero utilizábamos niveles de productividad, población ocupada en el sector agrícola, incorporación de valor agregado en la industria, con una variable en educación, que era el rendimiento académico medido por la tasa de eficiencia terminal.

Entonces, ¿por qué el interés en perfeccionar cada vez más estas comparaciones? Por varias razones, la primera es por el mal uso que se hace de ellas. Los ordenamientos son utilizados normalmente por los gobernadores de los estados o los secretarios de educación para tratar de ubicar a su entidad donde quieren que esté.

Voy a narrar varios ejemplos, pero comienzo con el de mi estado, porque ahí

vivo. En algún momento escribí una ponencia en términos de la evaluación, que se llamaba *La Honrosa Medianía de la Educación en Sinaloa*, es una frase juarista, muy elegante y de un significado muy propio, si ustedes la conocen. No quiere decir mediocridad, la honrosa medianía tiene que ver con un concepto ético, especialmente para los servidores públicos.

Cuando dije aquello y me di cuenta que en todos los indicadores –ahí están los libros que ustedes tienen– Sinaloa aparecía en medio, casi siempre cercano a la media nacional, con la excepción de primaria que se encuentra por arriba, pero en secundaria baja y en preparatoria baja más en promedio. Entonces, todo mundo decía *ah, caray, qué crítica tan fuerte hizo Pescador*. Pues no, lo dicho fue que a lo mejor en términos de los estándares de comparaciones que hemos tenido, Sinaloa se mantenía dentro del promedio nacional, lo cual a mí me preocupaba, porque teniendo los recursos humanos que tiene Sinaloa, debería estar, por ejemplo, como Nuevo León.

Nadie puede negar que las entidades de mayor desarrollo económico han tenido históricamente mejores índices en educación, de modo que las comparaciones tienen esa limitación, pero al mismo tiempo ese propósito. ¿Cómo se utilizan? De varias maneras.

La segunda experiencia la tuve en Tamaulipas; he recorrido todo el país hablando precisamente de los indicadores y de lo poco que tenemos. A lo mejor yo soy responsable de difundir esa subcultura de la evaluación, pero a las autoridades educativas les gusta mucho que se les diga y se reflexione en torno a lo que está pasando en sus sistemas educativos.

Lo que no les gusta es cuando uno corrige una afirmación hecha por ellos. Oiga, yo vi en el periódico ayer, porque estaba en plena campaña, que usted ya tiene el



98 por ciento de cobertura. Caray, eso no lo tiene ninguna entidad y me llama mucho la atención, ¿podríamos revisar la cifra? Sí, cómo no –hoy en la mañana, en la sesión de indicadores seguramente les explicaron cómo se ajustan estas tasas brutas y cómo se calculan otras tasas– pues obviamente, haciendo las estimaciones apropiadas se bajaba como hasta el 80 por ciento. Ah caray, entonces no lo diga, porque aquí están 400 maestros que ya leyeron el periódico y saben perfectamente que tenemos eso resuelto.

En otro estado me pasó con la preparatoria, con la media superior y utilizamos los índices del Ceneval. Ya explicaba Salvador Malo ayer y hoy en la mañana seguramente hubo un refrendo de lo que se dijo, cómo están calculados de acuerdo con las pruebas que ellos utilizan, los promedios y las referencias que aprovechan. Dije, fíjense qué curioso, en esta entidad algo me llama la atención, los estudiantes de mayor edad tienen, de acuerdo con la puntuación, un rendimiento menor, no quiero hacer ninguna reflexión adicional porque el propio Ceneval dice que sus anuarios no son para hacer comparaciones y que el número de cuestionarios aplicados en cada entidad no es significativo, etcétera, pero si uno lo toma como una aproximación para generar una reflexión, vieran qué reacción provocó entre las personas de Tamaulipas y de Nuevo León porque lo único que dije fue: ahora Tamaulipas está arriba de los neoleoneses.

Por supuesto no lo hacía con mala intención, pero los gobernadores sí, en muchas ocasiones lo que tratan de hacer es lo siguiente: primero, asignan recursos en función de los buenos resultados y entonces se alteran las evaluaciones o las mediciones; segundo, se utilizan para castigar, y ahí están los usos inapropiados y, tercero, siempre se tiene la pretensión de que se debe estar en los primeros lugares. Ahí

es donde es útil el conjunto de reflexiones que nos acaba de presentar Eduardo, para así poder situarnos en una realidad mucho más exigente, pero no se van a dejar de hacer las otras comparaciones.

Ahora bien, se ha hablado mucho de las cargas excesivas de opiniones superficiales que alimentan esta falta de cultura. Bueno, creo que también es la forma en que nosotros difundamos la información. Yo les preguntaría a todos ustedes, a dos años de la creación del INEE, cuál es el conocimiento en profundidad que tienen del INEE, qué es lo que saben de sus acciones.

Ayer se hicieron comentarios sobre la estructura de gobierno –que tampoco comparto– pero entiendo que se hacen porque la evaluación de la educación en su conjunto, como ocurre, por ejemplo, con la preocupación del nivel de vida de un país, tiene concepciones diferentes y debemos aceptarlo, eso lo va a vivir el INEE durante los próximos años mientras no llegue su primera etapa de consolidación en el 2006, y la segunda para el 2012.

### **Ver hacia adelante, no hacia atrás**

Primero, por favor lean los documentos del Instituto, ahí tienen el *Panorama Educativo*, ya en la mañana se platicó de él, lean el Primer Informe Anual, el que se llama *La Calidad de la Educación Básica en México*, no hay trabajos como esos. No quiero caer en el error de: *todo lo viejo es malo y todo lo nuevo es bueno*, ¡jamás!, hay muchos elementos buenos en el pasado, y estoy seguro de que algún investigador acucioso va a encontrar en todas esas limitaciones, y en todas esas pruebas con sus fallas alguna utilidad.

Creo que eso es lo más importante, ver hacia adelante, no hacia atrás, pero al punto que me estaba refiriendo es al de la lectura de los documentos editados por el

INEE, de esa manera conoceremos qué se está haciendo, porque no es suficiente el Internet.

Quisiera plantear, de lo que se ha formulado aquí, algunos desafíos que enfrenta el INEE, en términos de lo que he visto de cómo se ha planteado aquí su futuro.

En primer lugar, déjenme decirles que como miembro del Consejo Técnico comparto su visión y misión, pero sobre todas las cosas comparto el modelo de calidad que se ha impuesto el INEE, porque ese modelo exige la conformación de una masa crítica de recursos, que estoy seguro no lo va a tener ninguna institución.

Comparto sus resultados, estoy plenamente de acuerdo con ellos. Habrá algunas diferencias y matices porque estoy muy convencido de algunos cambios, necesarios de realizar independientemente de los señalados para su primera etapa de consolidación en el 2006, y para la etapa de consolidación plena en el 2012.

Primer desafío, debemos reconocer la existencia de paradigmas distintos y formas de ver la evaluación educativa, no es única la que está detrás del Instituto Nacional para la Evaluación de la Educación, porque por supuesto existe una forma de ver las cosas y en ese sentido hay que seguir aceptando aún los cuestionamientos y las opiniones, en las cuales yo, en las afirmaciones hechas por Eduardo, no observé ninguna mentira, nada más presté atención a las afirmaciones con base en documentos que ellos vieron y sí claro, México está arriba de Argentina y de Chile, dijeron, México está arriba de Argentina y de Chile; y si México está atrás de Hong Kong, dijeron, México está atrás de Hong Kong.

Y como siempre digo, Estados Unidos en las comparaciones internacionales, siempre aparece en medio cuando están los países europeos e industrializados y sobre todo los tigres asiáticos. A Estados Unidos sí le preocupa, por supuesto, pero

sigue siendo la primera potencia mundial y en términos de ingreso *per cápita* creo que es poco y en términos de otros tipos de satisfactores. De modo que hay algo, lo cual se debe completar con lo que está pasando con su sistema educativo.

Nosotros tuvimos un programa de calidad en educación en 1981, nunca lo evaluamos, nunca supimos lo que pasó, pero Estados Unidos inició en 1983 su gran transformación educativa, tiene veinte años y no ha mejorado prácticamente nada.

Segundo desafío, el INEE debe garantizar su permanencia y su continuidad, se dijo ayer y se dijo hoy, y ahí sí vale la pena apoyarlo, pero sobre todo, es muy importante que se generen los mecanismos institucionales y la solidaridad de grupos específicos, ustedes, investigadores, maestros, medios de comunicación, hay que hacerlo por el bien del país, no porque la institución subsista –porque también haré unos comentarios en ese sentido– sino porque es un elemento esencial para el diseño de políticas públicas en esta materia: educar, evaluar para transformar es vincular evaluación de calidad con educación, es reconocer ese carácter multidimensional, sobre todas las cosas, mejorar y hacer propuestas.

Tercer desafío, se debe poner énfasis en lo cuantitativo pero no excesivamente en los números. Hay que reconocer las estadísticas y los indicadores, pero también incorporar elementos cualitativos, los cuales vayan enriqueciendo esta cultura de la evaluación.

Cuarto desafío, debemos empezar a depender menos del Internet, pues sólo un porcentaje muy reducido de la población tiene acceso al mismo, debemos buscar la forma y en eso ustedes nos pueden ayudar.

Quinto desafío, evitar el ensimismamiento y sobre todo, la soberbia académi-



ca, esto es muy importante porque sucede que las concepciones de las instituciones de esta naturaleza se vuelven más cerradas, hay que abrirse, compartir espacios, no es un centro de investigación tradicional como cualquier otro, este centro de investigación llamado INEE, tiene una responsabilidad social muy importante.

No hay que alienar al maestro, no utilizarlo únicamente como aplicador, sin explicarle con claridad qué, para qué, cómo y cuándo se quiere evaluar; si los niños dijeran qué tipo de evaluación quieren, lo primero que nos dirían es: no queremos tantos exámenes.

Los resultados deben llegar a los maestros, a los formadores de docentes y a las instituciones, de lo contrario no vamos a poder formar esos recursos de los que tanto se ha hablado, es lo que más escuché entre ayer y hoy, formar y formar recursos, llegar a transformar las instituciones formadoras, darle un significado distinto al concepto nacional, es un instituto nacional, no es un instituto central, no abusar insisto, porque los niños se cansan y los maestros se molestan.

Debemos hacerle adecuaciones al marco jurídico, no para ciudadanizar, no pretendo eso, en primer lugar lo que debemos de hacer es consolidar esta coordinación y articulación de la que se habla y, por supuesto, avanzar en la consolidación del *sistema nacional de evaluación educativa*; debemos avanzar, y es responsabilidad colectiva la homogenización de un sistema heterogéneo, porque esto restringe las posibilidades de comparación como lo acaba de decir Eduardo.

Es necesario admitir, insisto, la existencia de esta cultura de la evaluación, probablemente equivocada, pero a veces nosotros también la alimentamos; qué pasó con el TIMSS, no se dieron a conocer resultados porque la Secretaría de Educación Pública simple y sencillamente los ocultó,

probablemente había razones de fondo, pero hubiera sido mejor darlos a conocer.

Debemos atender y respetar el sentido común, la evaluación nacional llega a maestros y padres de familia, no todos son especialistas, ni van a tener la capacidad para entender lo avanzado de los descubrimientos y la tecnología en la psicometría. Por eso, el sentido común a veces es importante, les puedo decir lo que los maestros y autoridades opinan en 14 entidades de la República, avalado con dos mil 300 preguntas, donde se enumeran las características de una buena escuela y la calificación promedio curiosamente es homogénea en el país.

Algo debe estar pasando, probablemente las comparaciones hacia el futuro nos van a enriquecer, les puedo decir que los maestros no están conformes con muchos de los elementos, los cuales afectan la efectividad de una escuela y que por cierto están en la parte última del Primer Informe Anual del Instituto.

Y finalmente, evaluar y evaluar bien, pero ligado como ya dijimos con la calidad; hay que publicar mucho y distribuir mejor, no importa que lo financie el gobierno, no importa quién lo haga, por ejemplo, ayer se decía, ¿quién financia al Instituto Federal Electoral? pues los impuestos, los contribuyentes, además recuerden que no todos los recursos públicos implican dominación, tampoco todos los recursos privados garantizan autonomía.

Por eso, como servidores públicos debemos estar atentos a estas condiciones. Cuidémonos de la dependencia excesiva en los instrumentos y en las instituciones del extranjero. Esto se percibe cuando se dice: todo lo del 2000 para atrás tiene fallas, ¿el mundo va comenzando en el 2000? No es cierto, hace mucho tiempo que lo comenzamos, es como el cambio democrático.

Por otra parte, acudimos al *College Board*, acudimos al *Education Testing Servi-*

ce, y recordemos que estas son instituciones autónomas, pero fíjense dónde empieza a impregnar el modelo de la evaluación y de la medición, en los organismos internacionales, aquí se mencionó a la OCDE, yo conozco el Banco Mundial, el Banco Interamericano de Desarrollo. No hay pues lejanía de esa concepción.

Por eso, lo más importante es entender

lo que hoy pasa con la evaluación y la rendición de cuentas, a lo mejor lo que antes era virtud hoy es vicio. Esperemos que la virtud siga siendo virtud.

Éste es precisamente el significado del Instituto y por lo cual nosotros pedimos la colaboración de toda la sociedad, por su trabajo, por su calidad y por lo que puede generar en beneficio de la educación nacional.

---

## ROBERTO RODRÍGUEZ GÓMEZ\*

---

**P**ara hacer un poco de justicia a la excelente ponencia del doctor Bachhoff, preparé un comentario en torno a los sesgos del *error estándar* debidos a la concatenación de varianzas de los segmentos que buscan representar.

El muestreo es un tema técnico, pero para mi sorpresa he oído mucho de fútbol en estas presentaciones y pues me animo a comenzar por ahí. Como muchos de los aquí presentes, yo siempre le he ido a *las Chivas* desde que era niño, gane o pierda. Casi independientemente de cómo juega, si sale campeón qué bueno, si no, la afición no cambia por ese detalle.

Digo eso sólo para mostrar que no tienen nada que ver estas preferencias, con aspectos tales como puntuaciones, posiciones en una tabla, puntos obtenidos, calificaciones técnicas, etcétera; ese tipo de decisiones que a veces tomamos como un ejemplo muy simple de *ranking*. Hasta el *ranking* futbolístico, cada quien lo sabrá según el equipo al que le vaya, no necesariamente tiene que ver con esto, hay otros aspectos y hay otras razones.

### Determinismo y relativismo

Creo que parte de esta discusión, en sus aspectos más serios y sustantivos, está reflejando uno de los debates que han durado casi un siglo en las Ciencias Sociales, y es el dilema entre determinismo y relativismo.

El determinismo dice: todo se puede medir, todo se puede evaluar, una vez evaluado se puede comparar y diagnosticar, y a partir de estos diagnósticos transformar la realidad. Es el punto en el que está situado el positivismo; y es por cierto, donde también está parada la ciencia.

El otro extremo es el del relativismo: si cada cabeza es un mundo, nada se puede comparar, todo depende de circunstancias, contextos, tiempos, la realidad es de una complejidad tal que resulta casi desconocida.

Hay razones para justificar un extremo y el otro, pero en realidad el conocimiento procede, si es serio, de una especie de punto intermedio de esa gama; con una suerte de determinismo crítico, una suerte de relativismo que no renuncia a las po-

---

\*Investigador del CESU y Presidente de Observatorio Ciudadano de la Educación.



sibilidades de conocimiento que dan las mediciones, aunque no se les toma tan en serio. Dicho esto, creo que la ponencia afilia a esta especie de posición intermedia, la cual también es la de este comentario.

Como se sabe, en la teoría clásica del muestreo, la medida de precisión de los resultados obtenidos en determinada prueba, experimento o encuesta, se expresa en función de la impresión. Es decir, no sabemos, no podemos saber qué tan precisa es una medición, lo que sí podemos saber es cuán imprecisa es ésta. Es, para los que están familiarizados con el tema, la teoría de los errores estadísticos; generalmente la medida de imprecisiones descrita a través de unidades de desviación estándar de los resultados, de tal manera que a mayor desviación estándar, menor precisión experimental y viceversa.

Conviene recordar, en primer lugar, que el valor asignado a la magnitud materializada por un patrón de referencia, en un lugar y en un tiempo determinado, se define como el valor verdadero convencional, lo cual quiere decir que, si bien los valores asignados a los sujetos participantes en una prueba –por ejemplo las puntuaciones totales y parciales en una prueba de conocimientos– son apenas una aproximación cuantitativa a tales conocimientos, no son los conocimientos mismos; se acepta como valor verdadero convencional algo que refleja o se aproxima al valor real.

En cada caso, esta suposición es verdadera, falsa o sesgada, en la medida en que la teoría de la que se desprende la medición es a su vez verdadera, falsa o sesgada. Este punto es de la mayor importancia, lo trató con abundancia el doctor Díaz Barriga, y no es nada trivial, quiere decir que hay estrictamente una relación entre teorías del aprendizaje, currículum y formas de medir.

Cuando se aplican pruebas que no tienen que ver con teorías del aprendizaje explícitas y con formas de presentación del

currículum, hay anomalías que no permiten comparar, ni siquiera entre lo que saben los estudiantes y lo que deberían saber, porque lo que deben saber es una cosa y lo que se mide es otra muy distinta; o entre lo que el maestro está enseñando y lo que el estudiante efectivamente aprende.

Entonces, una medición ideal que permitiera en efecto evaluar conocimientos, debería de tener una correspondencia mucho más estricta entre teorías del aprendizaje en acto, es decir, currículum y formas de medición.

### **Imprecisiones e incertidumbres**

Voy al asunto de las imprecisiones e incertidumbres. Decíamos que, de la teoría clásica del muestreo se desprende que la incertidumbre de la medida es un parámetro asociado al resultado de una medición, la cual se caracteriza por la dispersión de los valores, los cuales pueden, razonablemente, ser atribuidos al objeto mismo de la medición.

La incertidumbre de medida comprende, ya se dijo, varios componentes, algunos de los cuales se pueden evaluar a partir de la distribución estadística de los resultados de series de mediciones y se suelen caracterizar por sus desviaciones estándar, experimentales, empíricas, o bien, suponiendo distribuciones de probabilidad basadas en la experiencia adquirida u en otras informaciones, suponer por ejemplo, como se hace en las pruebas de evaluación educativa, que los resultados de aprendizaje se distribuyen en forma asimilable a la curva normal.

En conclusión, se parte del supuesto de que el resultado de la medición es la mejor estimación del valor del objeto de estudio, que todos los componentes de la incertidumbre contribuyen a la dispersión, y que ésta puede ser determinada bajo la forma de una probabilidad de error.

La dispersión se expresa en el muestreo a partir del error estándar, el cual se supone equivalente a la medida de desviación estándar de una población. Operacionalmente la desviación estándar es definida como la raíz cuadrada de la varianza empírica; es decir, la media de los cuadrados de los datos centrados, o de manera más sencilla: en la distribución muestral la desviación estándar o el error estándar es una medida que condensa la variabilidad teórica de las medias, de todas las medias posibles de un tamaño dado y por eso equivale a la desviación estándar de la media de la distribución empírica. Esto en términos conceptuales, pero el error estándar o la medida de desviación estándar es un promedio de variación de los datos con respecto a su valor central.

La conceptualización supone que, la media de la distribución muestral equivale a la media de la población objetivo, lo cual ocurre estrictamente en el caso de una distribución normal perfecta; no obstante, si la distribución experimental en la población objeto de estudio no es normal, se da por válido o se apoya el llamado *teorema del límite central*, que en términos no matemáticos indica que si el tamaño de la muestra es suficientemente grande, entonces la distribución muestral de la media se aproxima sintéticamente, se aproxima mucho a los valores de la normal teórica.

Por eso se utilizan convencionalmente los valores teóricos de una *Campana de Gauss*, porque por el *teorema del límite central*, aunque en la realidad no tenga esa forma la distribución medida, sus valores se pueden aproximar por un procedimiento de cálculo a esos valores.

### **Intervalo de confianza**

Además, la medida de la imprecisión experimental se vale de nociones que ya aquí fueron presentadas, tales como el *in-*

*tervalo de confianza* o el nivel de confianza. Como se sabe, éste es equivalente al rango de valores en que con un cierto nivel de confianza está presente el parámetro que se está estimando en el muestreo.

Por ejemplo, un nivel de confianza de 95 por ciento como el que indicaban los datos presentados y que es el utilizado en el muestreo de las pruebas 2003 de la SEP, acepta al menos dos lecturas.

La primera es: con razonable certeza por lo menos el 95 por ciento de todas las muestras posibles de seleccionar de la población objetivo, van a presentar valores dentro del intervalo de confianza del muestreo.

Pero acepta otra lectura más interesante y es que los valores centrales de la muestra, es decir, su media, tienen una exactitud entre más y menos cinco puntos con respecto a los parámetros de los valores reales convencionales, eso trataba de mostrar Eduardo.

Todos estos supuestos son estrictamente válidos en estudios con diseños no complejos, básicamente los derivados de un muestreo aleatorio simple; las pruebas de 2003, aunque tengan un diseño de muestreo simple en realidad no lo son, porque buscan representar distintos niveles de conglomeración de los resultados: por ejemplo, entidades federativas, escuelas o conglomerados de estudiantes.

Para cada uno de esos conglomerados, entidades, escuelas o grupos de estudiantes, hay varianzas diferenciadas y la varianza total de donde se obtiene la desviación estándar; por tanto, el cálculo del error estándar es la suma compleja de las varianzas de cada uno de los conglomerados.

En realidad, es la raíz cuadrada de la suma de las varianzas parciales de los conglomerados a los que previamente se sustrajo la unidad para tener una expresión del tipo 1.

La manera de medir convencionalmente el error estándar en un muestreo que no



es aleatorio simple como el de las pruebas, puede estar sobrestimando la magnitud del error estándar porque está sumando las varianzas de los distintos componentes, y hay una serie de alternativas para ajustar mejor la medición del error estándar y permitir una comparación más justa entre los valores puntuales. Por ejemplo, si comparo entidades federativas.

Se está experimentando de diferentes maneras con esto. Por supuesto existen las técnicas tradicionales de utilizar valores de correlación de resultados, de hacer uso de la *caja de herramientas* de la teoría de la correlación, la reducción de mínimos cuadrados, las regresiones, etcétera, para buscar mejores ajustes, pero producen resultados ilegibles. Es decir, no es lo mismo medir la diferencia en puntos que podemos entender, y medir la diferencia en valores de raíz cuadrada entre las distintas partes de una prueba.

Pasa por ejemplo con las pruebas del Ceneval, dicho con todo respeto; la escala del Ceneval cuenta con una media de mil 24 para el examen –eso no lo entiende nadie– y no es por falta de cultura o de familiaridad con las escalas de medición, son escalas complejas para asegurar una mayor precisión y no vérselas con problemas como los sesgos en los errores estándar.

Eso nos lleva a pensar que se necesitan valores que entienda toda la gente, por ejemplo, puntos en una escala de uno a cien, y la manera de estabilizar los errores de muestreo mediante otros procedimientos que no impliquen estadística sofisticada.

Hay técnicas muy interesantes para lograr esto y ajustar el error estándar concebido de la manera del muestreo clásico; programas, por ejemplo de cómputo, los cuales permitan redistribuir la muestra de varias maneras, para que la varianza o el error estándar de cada comparación sea el que efectivamente corresponda a la comparación, y no a la muestra como tal o a la suma de las varianzas de los conglomerados en la misma.

Lo que quiero decir con esto es que algunos de los problemas técnicos de la comparación son solucionables en términos técnicos; otros de los problemas de la comparación, como por ejemplo, aquellos que tienen que ver con el ajuste a través del aprendizaje, no tienen una solución técnica.

Con esta apresurada discusión sobre términos técnicos, trato de aproximarme al contenido de la ponencia, pero de ninguna manera considero que la solución de los problemas técnicos es la solución del otro rango del problema.



## PREGUNTAS Y RESPUESTAS



### EDUARDO DE LA GARZA VIZCAYA

**¿Por qué es importante comparar los resultados de una entidad a otra o de una escuela y otra? ¿Esta comparación en qué le sirve a la evaluación?**

Primero, toda evaluación debe terminar en un juicio, y éste al final de cuentas descansa en una comparación, no necesaria-

mente la comparación tiene que ver con el comportamiento de otra entidad o de otra persona, yo puedo comparar mi conducta con respecto a ciertos estándares o criterios que me he fijado.

La situación cambia cuando los objetivos que se pretenden alcanzar son comunes, cuando hay un acuerdo, como en el caso de las escuelas, de cubrir un currí-

culum común, es ahí donde tiene sentido conocer los elementos que inciden en el buen o mal desempeño, en otras palabras, poder explicar el aprendizaje.

**¿Es la comparación la que va a ayudar a mejorar las realidades concretas de cada entidad o escuela, o debemos buscar cambios y mejorar nuestras realidades educativas, independientemente de los logros o retrocesos de otras realidades?**

Volviendo a estos aspectos comunes, no hay que perder de vista el momento que estamos viviendo de globalización, donde de alguna forma los estados necesitan atender estándares de competitividad que están ahí presentes, y asegurar llegar a la población con ciertos logros comunes y esto, queramos o no, está pesando a veces para bien o para mal, por desgracia no todas las veces es para bien. Hay muchos trabajos sobre esto, pero es innegable que en los objetivos comunes que se plasman en el currículo, está detrás esta necesidad de atender estándares más allá de lo que nosotros mismos hemos establecido libremente.

**JOSÉ ÁNGEL PESCADOR OSUNA**

**¿Cómo se aplican los exámenes para los niños de las comunidades rurales donde su percepción del conocimiento es distinta? Por ejemplo, la leche que ven todos los días obtener de la vaca la deben comprar en la tienda.**

Insistí muchísimo en el modelo de calidad que viene operando el INEE y con el cual va a regir sus acciones para el futuro, éste es un modelo que todo mundo debe conocer, es casi perfecto, probablemente hay algunos aspectos que no puedo detectar yo, quizás los especialistas; pero uno de los puntos que se incorporan en las áreas de trabajo es cuidar los sesgos, si

eso es lo que me quieren preguntar.

Ustedes observaron que se han presentado los resultados –los cuales están plasmados en los documentos del Instituto– de escuelas públicas y privadas, rurales y urbanas, cursos comunitarios, secundaria general, telesecundaria, secundaria técnica, etcétera, son las modalidades contempladas; pero hay algo muy importante que está trabajando de manera muy profunda el Instituto, al menos así se nos informó a todos los miembros del Consejo, y es lo referente a que el conocimiento de un niño depende de sus características biológicas, del ambiente cultural de la familia, y luego, de lo que pasa en la escuela.

Una buena medición debe identificar estas variables y llegar al punto pretendido, si es comparación de niños de zonas urbanas y rurales, debe precisamente hacer esa separación, por eso se insiste en que las comparaciones con instrumentos –lo explicaba el Director General del Instituto– no son equiparables, tienen muchísimas fallas; no es que sean malas, él lo dijo claramente: simple y sencillamente no hay ese conjunto de 17 criterios, cuatro básicos, para hacerlas comparables.

Por eso señalaba francamente en mi participación: no es que el mundo haya comenzado en el 2000, lo que pasa es que la exigencia tecnológica y el tipo de investigación desarrollada, deben de apuntar a instrumentos que puedan visualizar esas diferencias.

**EDUARDO BACKHOFF ESCUDERO**

**En el caso del estado de Chiapas ¿Por qué ven el contexto en relación con la imposición de una cultura? ¿No sería mejor elaborar el instrumento respetando a la interculturalidad? Sin embargo, ¿ver a Chiapas en el último lugar debe explicarse? ¿el dato ha sido mal utilizado o interpretado? ¿el rezago es**



### **del estado o de quienes no han comprendido su contexto?**

Es una pregunta bastante compleja, pero la contestaría así, tratando de interpretar al autor de las preguntas. Yo estaría de acuerdo en el fondo de las preguntas.

Diría que el contexto en que se desarrollan los individuos es esencial para explicar los resultados de aprendizaje, como ya lo vimos. ¿Por qué? Simplemente porque lo que evalúan estas pruebas nacionales o de aprendizaje, es el cúmulo de conocimientos del niño adquiridos tanto dentro de la escuela, como fuera de ella. Entonces, como aprenden en un contexto determinado, hay que tomar en cuenta el mismo para interpretar los resultados.

Definitivamente como dice la pregunta, las pruebas no deben sesgar o castigar a los estudiantes por su cultura o lengua, en eso estamos totalmente de acuerdo.

Sin embargo, les quiero comentar que en México este tema apenas se está tomando seriamente, caso contrario al de Estados Unidos, en donde por su condición intercultural o multicultural, ha cobrado mucha relevancia. Por ello es muy importante demostrar que las pruebas utilizadas para evaluar el rendimiento no castigan al niño por la cultura, sexo o situación en la que se encuentre.

Finalmente, no creo que las pruebas puedan explicar el retraso educativo de Chiapas, sin embargo, debe de ser explicado totalmente por las condiciones generales de marginalidad en que se encuentran las distintas comunidades.

### **ROBERTO RODRÍGUEZ GÓMEZ**

#### **¿Por qué son tan difíciles de entender los rankings en educación?**

El problema no es su dificultad, por el contrario, el problema es su aparente facilidad, es decir, no tiene ningún misterio leer los ordenamientos, no importa en qué

puntos se ubiquen, pero que el primer lugar es Finlandia, segundo Noruega, el lugar número 8 es Estados Unidos y México aparece en el 32, eso es fácil de leer y se interpreta a la primera.

El problema es si ello hace justicia al objeto de la medición y tiene las implicaciones que suponemos tiene. Sin embargo, no vamos a renunciar a las posibilidades de conocimiento que ofrece la medición cuantitativa; y no solamente de conocimiento, también de crítica, esto es fundamental, cómo desde una posición crítica se pueden empezar a manejar los indicadores para exigir; si no, es solamente una herramienta para la imposición de políticas.

Otro asunto que nos debiera preocupar mucho, es cómo en estos últimos años está creciendo el analfabetismo. Ese indicador numérico nos sirve para una vigilancia crítica de las políticas, y muestra que no solamente del lado de la elaboración de éstas, es posible utilizar los indicadores.

### **FELIPE MARTÍNEZ RIZO**

#### **¿Cuál es el objetivo de realizar las Jornadas de Evaluación Educativa en vacaciones de alumnos y maestros?**

Hay una razón muy precisa: anteaer y ayer hubo reuniones del Consejo Técnico, al encontrarse aquí sus integrantes aprovechamos para que estuvieran en esta reunión. Pero, entre las cosas que pedimos nos aporten en unas hojitas de evaluación que distribuimos, es precisamente las fechas que les parecen más adecuadas, para tomarlas en cuenta y logremos organizar unas jornadas que correspondan mejor a sus necesidades.

Ahora me gustaría hacer algunos comentarios: quiero agradecerles su presencia, decirles que me sorprendió muy gratamente la respuesta que tuvimos a la invitación a estas jornadas; ello me hace

pensar que el interés que hay por el tema es mucho más consistente de lo que yo creía. Por tanto, debemos redoblar esfuerzos para difundir lo que hacemos.

El reto que tenemos es, por un lado, hacer evaluación de muy buena calidad, lo cual no es fácil, como hemos visto, pero no termina allí; incluye, además, hacer llegar los resultados a las instancias correspondientes, para desarrollar la cultura de la evaluación en el sentido amplio de la capacidad de comprender, analizar e interpretar los resultados, de tal manera que nos lleven a juicios adecuados.

Lo dicho por Roberto hace un momento me parece muy importante, el problema no es la complejidad de interpretar, sino pensar que es muy fácil; ese es el problema que enfrentamos, creer que los resultados de las evaluaciones son sencillísimos y nos permiten decir sin ningún matiz, sin ninguna contextualización, que tal país está en tal lugar, tal estado está en tal sitio, tal escuela está bien, regular o mal; y no es así, es algo mucho más complejo.

Parte de la cultura de la evaluación, y por ahí tenemos que empezar, es que to-

dos tomemos conciencia de lo difícil que es el tema y de la necesidad de ser cautos y prudentes, para evitar un mal uso de la evaluación.

Cuando en Inglaterra, hace cerca de veinte años, hacia 1987, se hizo el primer *Sistema de Indicadores de Evaluación de Educación Superior* y salieron los resultados, un analista decía: *a estas listas hay que ponerles un letrero que diga: "El uso inapropiado de estos indicadores puede ser dañino para la salud de su institución"*.

Efectivamente, el uso inadecuado de la evaluación puede ser dañino para la educación, así como el uso apropiado de la evaluación puede ser muy saludable. Necesitamos hacer buena evaluación y desarrollar la cultura de la evaluación, para que sus resultados sean interpretados correctamente y usados adecuadamente.

Las diferentes posturas manifestadas en las jornadas, son reflejo de una realidad muy clara: hay diferentes puntos de vista, y eso es muy bueno. Debemos discutirlos, dialogar para llegar a consensos entre todos, partiendo de la base de que tenemos muchos puntos de acuerdo, y éstos son más que los desacuerdos.



Fue necesario habilitar espacios adicionales por la numerosa asistencia de personas interesadas en lo temas de las Jornadas.



En los últimos años la evaluación de la educación ha despertado gran interés en la sociedad.