

EVALUAR LAS EVALUACIONES

Una mirada política
acerca de las evaluaciones
de la calidad educativa



Gustavo Iaies
Javier Bonilla Saus
José Joaquín Brunner
Pablo Halpern Britz
Martín Granovsky
Alejandro Tiana Ferrer
Felipe Martínez Rizo
Juan Carlos Navarro
Emilio Tenti Fanfani

Instituto Internacional de Planeamiento de la Educación



IIFE - UNESCO
Sede Regional Buenos Aires



EVALUAR LAS EVALUACIONES

Una mirada política
acerca de las evaluaciones
de la calidad educativa

EVALUAR LAS EVALUACIONES

Una mirada política acerca de las evaluaciones de la calidad educativa

Gustavo Iaies
Javier Bonilla Saus
José Joaquín Brunner
Pablo Halpern Britz
Martín Granovsky
Alejandro Tiana Ferrer
Felipe Martínez Rizo
Juan Carlos Navarro
Emilio Tenti Fanfani

Instituto Internacional de Planeamiento de la Educación



IIPE - UNESCO
Sede Regional Buenos Aires



EVALUAR LAS EVALUACIONES

**Una mirada política acerca de las evaluaciones
de la calidad educativa**

© Copyright UNESCO 2003

International Institute for Educational Planning

7-9 rue Eugène-Delacroix

75116, Paris, Francia

IIPPE - UNESCO - Sede Regional Buenos Aires

Agüero 2071

(C1425EHS) Buenos Aires, Argentina

Índice

Perfil de los autores	9
Prólogo, por Juan Carlos Tedesco	11
Evaluar las evaluaciones, por Gustavo Iaies	15
Encuentros y desencuentros con los procesos de evaluación de la calidad educativa en América Latina, por Javier Bonilla Saus	37
Límites de la lectura periodística de resultados educacionales, por José Joaquín Brunner	67
Indicadores y sus usos en educación: una tensión no resuelta, por Pablo Halpern Britz	85
La evaluación como fetiche, por Martín Granovsky	97
¿Qué pretendemos evaluar, qué evaluamos y qué conclusiones podemos extraer de la evaluación?, por Alejandro Tiana Ferrer	111
Una mirada técnico-pedagógica acerca de las evaluaciones de calidad educativa, por Felipe Martínez Rizo	133

La evaluación y las actitudes de los docentes frente a ella: dificultades y alternativas de política, por Juan Carlos Navarro	147
Los docentes y la evaluación, por Emilio Tenti Fanfani	165

Perfil de los autores

Gustavo Fabián Iaies: Es Licenciado en Ciencias de la Educación por la Universidad de Buenos Aires y cursó estudios de Postgrado en FLACSO, Buenos Aires. Fue Secretario de Educación Básica y en la actualidad es consultor del IIPE - UNESCO Buenos Aires y de la Secretaría de Educación Pública de México.

Javier Bonilla Saus: Es Licenciado en Sociología y Economía Política y candidato al Doctorado de Estado en Economía Política, París, Francia. Actualmente es Director Nacional de Educación Pública y Presidente del Consejo Directivo Central de la Administración Nacional de Educación Pública, Uruguay.

José Joaquín Brunner: Es sociólogo de la educación y profesor e investigador de la Escuela de Gobierno y Director del Master en Gerencia y Políticas Públicas de la Universidad Adolfo Ibáñez en Chile. También es Director del Programa de Educación de la Fundación Chile.

Pablo Halpern Britz: Es Doctor en Comunicación por la Annenberg School for Communication Pennsylvania, EE.UU. Actualmente es decano de la Facultad de Comunicaciones de la Universidad del Desarrollo, Chile.

Martín Granovsky: Es Periodista y subdirector del diario argentino *Página/12*. Fue ganador del Premio Rey de España a la excelencia periodística. Entre sus publicaciones se encuentra *El divorcio*. La historia secreta de la ruptura entre Chacho y De la Rúa, las coimas en el senado y la crisis en la alianza.

Alejandro Tiana Ferrer: Es profesor de la Universidad Nacional de Educación a Distancia, España. En la actualidad se desempeña como Director General de Innovación y Desarrollo de la Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI) y es presidente de la International Association for the Evaluation of Educational Achievement (IEA).

Felipe Martínez Rizo: Es Licenciado en Ciencias Sociales por la Universidad de Lovaina, Bélgica, presidente del Instituto Nacional de

Evaluación educativa de México, y profesor y miembro del Consejo Mexicano de Investigación Educativa, de la Association for the Study of Higher Education, y del National Council for Educational Measurement.

Juan Carlos Navarro: Es politólogo, realizó estudios de doctorado en la Universidad Central de Venezuela, Caracas. Actualmente, se desempeña como Jefe de la Unidad de Educación del Banco Interamericano de Desarrollo. Publicó recientemente la obra *Quiénes son los maestros: carreras e incentivos docentes en América Latina*.

Emilio Tenti Fanfani: Es Licenciado en Ciencias Política y Sociales y graduado en el Tercer Ciclo de la Fondation National des Sciences Politiques de París. Es profesor en la UBA, investigador del CONICET y consultor del IIPE - UNESCO Buenos Aires.

Prólogo

El ciclo de procesos de transformación educativa que comenzó en casi todos los países durante la década de los años '90, estuvo basado en la idea según la cual era prioritario modificar el diseño organizativo e institucional de los sistemas educativos. Uno de los argumentos sobre el cual se sostenía esta prioridad era el bajo nivel de responsabilidad por los resultados con los cuales operaban las administraciones educativas tradicionales. Según el enfoque dominante en esos años, para aumentar la responsabilidad por los resultados era preciso otorgar mayor autoridad a los actores del proceso pedagógico –y en ese sentido se ubicaban las estrategias de descentralización y de autonomía a los establecimientos escolares– y establecer mecanismos que permitieran medir sistemáticamente los logros de aprendizaje y difundir masivamente la información para provocar mayor y mejor demanda por parte de las familias, la opinión pública y el propio gobierno.

En el caso de América Latina, prácticamente todos los países crearon sistemas nacionales de evaluación de la calidad, se instalaron mecanismos regionales de medición de resultados y varios países se incorporaron a las mediciones internacionales. De esta forma, los sistemas educativos comenzaron a disponer de instrumentos técnicamente confiables para conocer sus resultados, especialmente aquellos referidos a las asignaturas básicas: Lengua, Matemáticas, Ciencias Sociales y Naturales.

Después de más de una década de la instalación de estos instrumentos, parece llegado el momento de evaluar a los evaluadores. Las razones que justifican este ejercicio son muy diferentes. Por un lado, las mediciones indican que la calidad de la educación no está mejorando. En general, todas

las evaluaciones indican que no hay cambios significativos. Estos resultados podrían ser vistos como un dato positivo si lo consideramos en relación al crecimiento de la matrícula. Desde este punto de vista, podríamos afirmar que los sistemas educativos de la región mantuvieron sus resultados, aun a pesar de incorporar alumnos de sectores con mayores dificultades socioculturales. Sin embargo, tampoco aparecen cambios sensibles en aquellas zonas en las que el aumento de matrícula no fue significativo. En síntesis, lo cierto es que, a pesar de los enormes esfuerzos realizados por los procesos de reforma educativa y, en algunos casos, por la significativa inversión financiera realizada, los resultados de los procesos de evaluación no parecen registrar cambios en los resultados educativos.

Este dato abre varios interrogantes: ¿las reformas educativas no han modificado lo que sucede en las escuelas o los instrumentos no están midiendo correctamente los resultados? La instalación de estos mecanismos, ¿ha provocado mayor responsabilidad por los resultados o ha servido, por el contrario, para segmentar aún más la demanda educativa?

Un hecho es innegable: los sistemas de medición han convertido los resultados educativos en un motivo de debate público. El impacto que tiene la difusión masiva de los resultados de las pruebas nacionales o internacionales a través de los medios de comunicación de masas es muy significativo, especialmente si se lo analiza desde el punto de vista político. En este sentido, los instrumentos técnicos de medición de resultados han adquirido un significativo papel político y la comunicación se ha convertido en una de las principales dimensiones que debe manejar toda gestión de política pública en educación. En síntesis, ya no es posible pensar la problemática de la evaluación exclusivamente desde el punto de

vista técnico-pedagógico, sino que se requiere otro tipo de mirada, mucho más amplia y compleja.

En esta ampliación de la mirada sobre los instrumentos de evaluación debemos considerar aspectos internos y aspectos externos a las acciones educativas. Desde el nivel interno, es necesario incluir la relación de las políticas de evaluación con los actores del sistema. Más allá de las intenciones, parece evidente que los docentes percibieron estos dispositivos más como una presión o una amenaza que como un insumo para el mejoramiento de las estrategias de enseñanza y aprendizaje. En ese contexto, la información de los resultados no se transformó en un insumo para el mejoramiento de las políticas, los datos no se elaboraron en las escuelas y el impacto sobre las estrategias de enseñanza fue muy débil.

En relación a los actores externos al sistema, es preciso considerar todo lo relativo a la comunicación de los resultados. ¿A quién, cómo y cuándo se deben comunicar los resultados? ¿Qué puede y debe hacer cada actor con esta información? Estas preguntas aparecen en el momento en que transformamos a las evaluaciones de la calidad y sus resultados en un producto de la comunicación masiva, en el momento en que interactúan con el conjunto de la información transmitida por los medios masivos de comunicación. El modo de presentar la información, la propia información presentada, los interlocutores a quienes se la presentamos, el hecho de que sea el propio gobierno y su ministerio de Educación los que la legitimen, obliga a otro nivel de análisis de la problemática. La comunicación masiva de esos resultados es uno más de los elementos que han transformado a las políticas de evaluación en un hecho político además de una construcción técnico-pedagógica.

Por último, también es preciso proponer una mirada

que vincule la evaluación con los objetivos de las políticas sociales. ¿Evaluamos para mejorar la eficiencia de la inversión educativa? ¿Para orientar la demanda? ¿Para compensar las diferencias sociales? ¿Para comprobar los niveles de segmentación social? Los resultados de las evaluaciones dependen de las decisiones que tomemos en relación a “Para qué” evaluamos, y esas decisiones suponen definir el escenario social para el cual estamos trabajando. Detrás de esta mirada, ha crecido un debate ideológico y político. El “para qué” se ha vuelto una pregunta central en el análisis, porque define una idea sobre la sociedad que queremos construir. Modelos que ponen la idea de equidad en el centro del debate, otros que ponderan el concepto de competitividad, aquellos que se centran en la idea de eficiencia, por ejemplo, tendrán distintas respuestas a la hora de pensar las políticas de evaluación.

Estas consideraciones estuvieron en la base de la organización del Seminario Internacional “**La dimensión política de la evaluación de la calidad educativa en Latinoamérica**”, cuyos resultados se presentan en este libro. El Seminario reunió a líderes políticos y Ministros de Educación de la región, comunicadores sociales y especialistas en evaluación, quienes expusieron sus puntos de vista en un diálogo franco y abierto. Este libro reúne las principales exposiciones presentadas durante la reunión y un resumen de los debates. Confiamos en que la publicación de estos materiales permitirá enriquecer tanto las discusiones como las estrategias que los distintos países adopten para una utilización más pertinente de estos instrumentos.

Juan Carlos Tedesco

Director

IPE - UNESCO - Buenos Aires

Evaluar las evaluaciones

Gustavo Iaies

“¿Qué nos aportan estas evaluaciones en relación al terrible costo que implican?”. Con estas palabras, Mariana Aylwin, ex-ministra de Educación de Chile, tradujo el clima crítico o de tensión que las evaluaciones han generado en los tomadores de decisiones. Dicho clima se percibió en el Seminario Internacional “Una mirada política acerca de las evaluaciones de calidad” que el IPE-UNESCO Buenos Aires organizó en Santiago de Chile, en diciembre del 2002. En este artículo intentaré volcar algunas de las ideas principales que formaron parte del debate de los tomadores de decisiones, al menos algunas de las que observé con mayor interés.

Introducción

“En una escuela que tensiona por la equidad, la evaluación, como dispositivo diferenciador, hace ruido en algún lugar”. Estas palabras de Javier Bonilla, Director Ejecutivo del CODICEN – Uruguay, se suman y permiten observar el tono de un debate. La idea de buscar esos “ruidos” o esos “costos” y compararlos con los beneficios se empieza a transformar en un eje de la discusión acerca de las evaluaciones. Parecen estar más claros los costos que los beneficios, y eso lleva a una revisión crítica; la idea es que las evaluaciones de la calidad empiezan a verse sometidas, ellas mismas, a un proceso de evaluación.

Estas políticas aparecieron como una de las innovaciones principales de las reformas implementadas en la década de los años '90. Y, paradójicamente, los propios tomadores de decisiones sienten que son estos dispositivos los que han puesto en "jaque" a las reformas. Han expuesto a la política educativa, la han sometido a la opinión pública y, en ese campo, las decisiones y el debate abandonaron la exclusividad del mundo educativo. Desde ese punto de vista, la capacidad de dar respuestas, de analizar datos y de producir cambios, se encuentra muy condicionada.

La centralidad que las evaluaciones de la calidad han asumido en la mirada que la sociedad tiene sobre la educación las vuelve un aspecto que merece ser analizado con mayor cuidado. Han dejado de ser un objeto técnico-pedagógico para transformarse en uno de carácter político.

Más allá del consenso existente sobre su importancia y sobre las ventajas de su instalación, resulta fundamental volver a pensarlas, principalmente, desde sus dimensiones política, comunicacional, pedagógica y en el modo en que estos dispositivos interactúan con el conjunto de los actores educativos.

Se trata de proponerse una evaluación de las evaluaciones, a diez años de su instalación en los sistemas educativos de la región, y de revisar críticamente la experiencia.

Una visión política

Las evaluaciones se han vuelto un hecho político. Los gobiernos y sus políticas son analizados a partir de los resultados de la aplicación de estos dispositivos. El mundo de la educación ha creado un instrumento sobre el que parece haber per-

dido el monopolio del debate. En el momento en que las evaluaciones han llegado a los medios de comunicación masiva, han quedado de algún modo fuera del campo exclusivo de los gestores educativos o de los pedagogos. La comunidad educativa ha “bendecido” un indicador capaz de dar cuenta de la situación de la educación en materia de aprendizaje de los alumnos y en esos términos lo ha recibido la sociedad. Desde ese momento, se han vuelto un procedimiento de las políticas públicas en su conjunto, y eso implica la participación de otros actores. Las desigualdades sociales, la competitividad de los países, la eficiencia del gasto, entre otras variables, empiezan a incluir la información de los operativos de evaluación de la calidad para ser analizadas.

En ese contexto, el mundo educativo ha perdido el manejo exclusivo de la herramienta. Ya no se trata de un dispositivo que produce información para el debate educativo exclusivamente, ha dejado de ser un objeto de discusión técnica, el mundo político se ha apropiado de sus resultados y los ha incorporado a sus debates. En muchos casos, con análisis apresurados, poco apropiados e, incluso, revelando muchos de los errores cometidos o la necesidad de producir cambios.

¿Cómo recuperamos la posibilidad de pensar la herramienta? ¿Cómo volvemos a transformarla en un dispositivo capaz de brindarnos información para el mejoramiento de la calidad y la equidad? Tales parecen ser las preguntas de los tomadores de decisiones. Retomando las palabras de Martín Granovsky, en el Seminario: “Si no llenamos la evaluación con información y análisis, los resultados están sustituyendo a la política”.

Parece necesario recuperar el sentido. ¿Evaluamos para mejorar la equidad y calidad de nuestros sistemas educativos?, ¿lo hacemos para juzgar al sistema y acusarlo? Por-

que de acuerdo a las intenciones que tengamos, los instrumentos serán las propias políticas, no serán un instrumento sino un contenido. Los sistemas educativos dejarán de trabajar para mejorar la calidad y la equidad educativa y pasarán a trabajar para el mejoramiento de los resultados de las evaluaciones.

Los dispositivos y sus productos han tenido mayor impacto en la construcción del imaginario educativo de la sociedad, que en la transformación de las estrategias educativas. En el plano de lo simbólico vivimos, como expresó Juan Froemmel en el Seminario: “el síndrome del análisis melancólico, todos creen que nunca tuvo fiebre porque nunca se la midieron”; la sociedad no cuenta con registros históricos, con lo cual “se pelea” con los actuales. En este sentido, Ottone dijo en el Seminario: “Existe la idea de un pasado «glorioso», una escuela que convivía cotidianamente con la excelencia. Con ese pasado es imposible compararse porque ese era un mundo más feliz. No creo que la felicidad sea un efecto de las políticas públicas, la felicidad es compleja. Es muy complejo hacer políticas públicas contra ese síndrome de la nostalgia”. No existen registros históricos en la región que permitan comparaciones serias con el pasado, pero, de todos modos, el imaginario social los ha construido. Y resulta complejo relativizar esas ideas incorporando el dato de las tasas de escolarización o de la complejidad de los planteos curriculares o de la menor desigualdad, en algunos casos. Parece haber un dictamen que establece que la escuela actual es peor que la del pasado. Y ese dictamen parece asociarse a la calidad de nuestros gobernantes.

Las evaluaciones parecen haber sido funcionales a este síndrome, quizás, hasta lo hayan alimentado. Acaso hayan confirmado a la sociedad que ya no tendrá una escuela de excelencia como la que tuvo.

La sensación es que esos resultados han venido a confirmar el peso de ciertos condicionantes estructurales. Así lo indicó J. J. Brunner, en el Seminario: “Ningún indicador logra imponerse a los de los condicionantes sociales. En el mejor de los casos se podría llegar a equiparar el efecto del origen social. La escuela no puede compensar a la sociedad, el sentido común le ha ganado a las comprobaciones científicas”. Estas afirmaciones tienen un enorme peso en términos de la ejecución de políticas, es casi una justificación para disminuir las inversiones en educación y centrarlas en las políticas sociales. Podrían permitir razonamientos que justifiquen el hecho de que si los resultados educativos están condicionados por los sociales, no conviene invertir en políticas educativas, sino en aquellos programas sociales que garanticen las condiciones de educabilidad, dado que hasta que impactemos la situación social de los beneficiarios, no podremos mejorar sus resultados educativos.

El pesimismo pedagógico de final de década, la evidencia de que enormes inversiones y una importante apuesta política no han producido resultados significativos en materia de calidad, hace que este tipo de argumentos ganen espacio. La idea de que el sistema tiene muy baja permeabilidad a los cambios, que es muy difícil producir transformaciones cualitativas y que requieren gran cantidad de recursos, parece ser un obstáculo para volver a construir un clima favorable a la inversión educativa.

Al mismo tiempo, el análisis de los resultados de las evaluaciones internacionales muestra que los promedios de la región se componen de sectores acomodados de la sociedad que brindan a sus hijos una educación que puede ser ubicada dentro de los estándares internacionales, y otros cuya calidad se aleja cada día más de dichos estándares.

Esta evidencia anima a algunos sectores a preguntarse si es posible garantizar una educación de calidad para todos o si deberemos conformarnos con garantizársela a los grupos más dinámicos de la sociedad, aquellos que parecerían en condiciones de darle competitividad a las economías de sus países. Estos peligrosos razonamientos son algunos de los elementos que nos obligan a asumir un análisis más cuidadoso de los resultados de las pruebas. En ese sentido es que los mismos han trascendido el análisis del mundo pedagógico: no sólo tienen impacto sobre las políticas curriculares o de gestión institucionales, también pueden tenerlos sobre las políticas de empleo, de inversión industrial, etcétera.

¿De qué hablan los tomadores de decisiones cuando afirman que hay que apropiarse de las evaluaciones? El Presidente de la IEA, Alejandro Tiana, dijo: “Hay una cuestión de la complejidad de la evaluación que tiene una vertiente de conocimiento y otra de valoración. La vertiente política es la de la valoración y la ponderación”. Por su parte, Martha Lafuente expresó: “Creo que llega el momento de empezar a llamar a las cosas por su nombre, nos hemos propuesto evaluar calidad y estamos midiendo, no evaluando. Evaluar exige tener un patrón con qué compararse y ese patrón no está del todo claro”. Aparece la idea de que se ha concebido una política como un hecho técnico y, después de una década, parece claro que es necesario revisar esa decisión, entender que esas decisiones técnicas deben dar cuenta de definiciones filosófico-políticas. Se construyeron indicadores que se definieron técnicamente, y que consideran casi con exclusividad las habilidades académicas. Nuestros índices no consideran el aumento de las tasas de escolarización, la capacidad del sistema para homogeneizar actores de una sociedad cada día más segmentada, dar cuenta de los nuevos públicos que la

escuela ha sido capaz de albergar, de la capacidad de contener otras realidades sociales, etc. Y esas definiciones implican una toma de postura ideológica, utilizar unas variables y abandonar otras; lo cierto es que la experiencia de los '90 hace pensar más en una "no toma" de posición política, en el sentido de que los tomadores de decisiones no se posicionaron en ese punto.

Se ha construido un indicador de calidad que es puramente académico, se han considerado fundamentalmente los aciertos en Lengua y Matemáticas como indicadores de una buena educación. Pero no es que, en consecuencia, esa sea la demanda que se le formula a la escuela, no ha sido ese el único objetivo de las reformas. Y en el caso de que ese hubiera sido, tampoco se le garantizaron las condiciones de educabilidad mínimas para producir esos aprendizajes.

Se le ha pedido a la escuela que enseñe contenidos definidos, y se formularon evaluaciones acerca del grado de adquisición de los mismos por parte de los alumnos. Pero también se le demandó que les dé de comer a los niños y jóvenes, que controle su vacunación, que les haga los documentos de identidad, que presione a los padres para que eduquen a sus hijos como corresponde, etcétera. Y ni siquiera los sistemas educativos fueron capaces de liberarla del peso administrativo de las burocracias, no se han dictado normativas y regulaciones que faciliten la tarea, no se han generado programas de capacitación eficientes para sus docentes, de modo que pudieran enseñar dichos contenidos.

Y en ese sentido, la escuela ha quedado sola adelante de la sociedad; se la midió con un indicador de nivel académico similar al de otras regiones del planeta en las que la escuela cuenta con una cantidad de variables garantizadas. No se han construido indicadores capaces de dar cuenta del con-

junto de las misiones y funciones que se le encomendaron al sistema educativo. Ni siquiera, que consideren el contexto en el que se le han pedido dichos resultados.

Las evaluaciones de la calidad han lanzado al sistema educativo a la discusión política y, por lo tanto, las consecuencias de dichas discusiones generan un fuerte impacto sobre la imagen de los gobiernos, sobre el campo de la política educativa y sobre el de la política en general.

No es posible enfrentar una discusión política con enfoques técnicos solamente. Quizás haya llegado la hora de asumir la discusión de sentido, la definición de objetivos de política y de someter las herramientas a esos principios. Quizás el problema resida en que, tal como dijimos previamente, se ha medido pero no evaluado, y medido determinadas variables y no otras. Y más allá de que los equipos políticos no hayan construido esas decisiones, las mismas han adquirido ese carácter.

Quizás la primera conclusión de este análisis sea que, más allá de la politización “de hecho” que sufrió el mundo de las evaluaciones, ha llegado el momento de que los tomadores de decisiones en materia de políticas educativas, los especialistas del campo y los equipos de los organismos multilaterales asuman que este es un debate de sentido, y que entrar al mismo por el “costado” técnico en lugar de hacerlo por el político es un error y nos conducirá a conclusiones falsas.

La visión comunicacional

“Los diarios informan sobre la lista de resultados, las radios llaman a la directora de evaluación, la televisión prepara un

programa a la noche y al otro día desapareció”. De este modo, Martín Granovsky describe algo así como el circuito de la evaluación en los medios, la idea de que el tema se vuelve una noticia más, no hay proceso, no hay actores, son sólo resultados. Por su parte, Javier Bonilla Saus agregó: “Las evaluaciones clásicas son generadoras de buenos y malos humores. Dificilmente son generadoras de grandes movimientos de opinión. La publicación de que los índices de repitencia han subido puede desembocar en largos cuestionamientos sobre el ministro, los datos, las políticas educativas, la crisis de la educación, etcétera”.

Como señaló J. J. Brunner, en el Seminario: “Esta claro que la comunicación masiva le ganó a los sociólogos y educadores”. Pareciera ser que el mundo educativo ha perdido el control de ese canal de comunicación y, encima, es el único que está utilizando.

Ahora, ¿por qué el mundo educativo creyó que el tratamiento de esta información iba a ser distinto? ¿Es otro el abordaje que los medios masivos proponen para otras temáticas? ¿Por qué los medios se habrían dejado condicionar por la lógica del pensamiento pedagógico?

La escuela tuvo una cierta centralidad en el comienzo del siglo XX y finales del XIX. Era una institución con capacidad de construir muchos de los discursos que circulaban por la sociedad. Pero esa capacidad ha disminuido, por pérdida de peso propio, pero, también, por la aparición de otras agencias de transmisión de discursos en la sociedad. La capacidad hegemónica de la escuela se encuentra debilitada, y, por el contrario, es ella la que se ve sometida a otras agencias sociales que le atribuyen sentidos.

Las evaluaciones de calidad siguen un proceso similar al de los indicadores de aumento del costo de vida, del de-

empleo, del producto bruto. ¿Por qué sería distinto? ¿Acaso esos indicadores no impactan del mismo modo en la vida de los miembros de la comunidad? Ocurre que, en algunos casos, las personas logran construir el relato que “rodea” al indicador estadístico, porque participa de sus vidas en forma cotidiana, como es el caso del índice de desempleo. Pero en la mayoría de los casos un indicador no es más que eso, no es un relato, no es un análisis, sino una síntesis. Es limitada la capacidad que tiene un dato para representar una realidad compleja, para dar cuenta de la multiplicidad de relatos que lo componen, y del conjunto de los matices que le aportan sentido.

Otro de los problemas de la información que aportan los operativos nacionales de evaluación de la calidad, es que no cuentan con el consenso del campo. El indicador que se ha construido no parece contar con el acuerdo del mundo educativo acerca de que expresa una síntesis del estado de la calidad educativa. Los indicadores económicos no generan discusión per se, el debate generado por el índice de variación del costo de vida puede vincularse a las razones de las variaciones, al modo de construcción de la muestra, pero nadie cuestiona al propio índice. Nadie cuestiona el hecho de que mide lo que se propone medir.

Sin duda, una de las deudas en materia de evaluación de la calidad es el consenso acerca del indicador propuesto, las variables y la metodología con la que se construirá el mismo.

Esta discusión se deriva del planteo anterior; respecto de la definición de sentido, es preciso consensuar una definición de calidad antes de empezar a medirla. Ese consenso es lo que el sistema educativo debe volcar al resto de la sociedad, de modo que el mismo sirva como principio de un

diálogo y una comunicación con ella. También debe ser discutido el hecho de medir solamente la calidad, más allá del consenso alcanzado acerca de dicho indicador.

Pero definido dicho indicador, cabe la pregunta respecto de si deben ser los medios de comunicación el medio elegido para la transmisión de los resultados, o cuáles resultados deben ser comunicados a través dichos medios. En el Seminario, Juan Carlos Tedesco señaló: “La duda es cuál es el impacto de los medios sobre la educación. La gente, en general, no lee la información escrita sobre educación, la leen los dirigentes políticos. La gente se apoya mucho en la experiencia concreta y eso explica y da pistas para el diseño de una política educacional. Es preciso reforzar el papel comunicador de la escuela. Los datos no les sirven a los docentes tampoco, tal como están presentados. Los ciudadanos no nos guiamos sólo por información racional”. El aporte de Juan Carlos Tedesco obliga a pensar si se ha elegido el medio más indicado para comunicar los resultados, por su capacidad de llegada, por el tipo de discursos que admite y el modo de procesarlos, por la capacidad de consensuar con los actores el modo de comunicación de los mismos, etcétera. Más aún cuando la televisión, la radio y los periódicos han sido, en algunos casos, los únicos canales elegidos para la transmisión de la información.

Es preciso evaluar si el tipo de comunicación y el modo de organizar los discursos han sido los apropiados. La posibilidad de evaluar modelos de comunicación que no trabajen sobre los paradigmas de la velocidad, inmediatez, impacto, parecen más apropiados para comunicar la información educativa con mayor cuidado. Se trata de encontrar los canales para transmitir relatos, con tiempos y espacios que permitan la reflexión.

Al mismo tiempo, si reunimos el análisis de eficiencia con el de pertinencia, daría la sensación de que el propio sistema puede ser un canal más interesante de comunicación.

En este sentido, no parece un dato irrelevante que no se haya tomado al propio sistema como canal para la comunicación de los resultados.

Pero la reflexión acerca del canal de comunicación obliga a volver al plano de las opciones de sentido. El modo y la vía que se utilizará para comunicar los resultados debe estar vinculado a los objetivos propuestos para los dispositivos de evaluación. Es preciso preguntarse “para qué” antes de definir el “cómo”; la comunicación debe tener capacidad de construir sentido con el propio discurso que se está intentando transmitir, debe existir relación entre los contenidos y el modo de transmitirlos.

Si los objetivos que el sistema se propone pasan por producir cambios en las estrategias de trabajo de los docentes, si se está pensando que los resultados sean un feed back que mejore las estrategias áulicas, las políticas de capacitación, de currículum, y el planeamiento en general, se hace necesario buscar una estrategia en la dirección y la lógica de nuestros destinatarios. Ahora, si lo que se busca es un insumo informativo para un sistema que avance hacia una organización a partir de la demanda, un modo de que la sociedad y los propios padres puedan aumentar la presión sobre la oferta, si la hipótesis es que dicha presión sobre la escuela la obligará a mejorar, los medios masivos pueden ser un canal adecuado.

Pero, sin duda, vuelve a aparecer la demanda por una definición de sentido. Y en el caso de la comunicación, los especialistas del campo le aportan complejidad al debate, obligan a integrar al análisis la elección del canal, el modelo de

comunicación, las intencionalidad, tanto como los propios contenidos. Pareciera que los tomadores de decisiones son nuevamente demandados respecto de decisiones de valores y de sentido; así, el debate lleva a la necesidad de que los especialistas encuentren un marco ordenador de su tarea.

Una mirada técnica

“La mayoría de los sistemas educativos latinoamericanos ha incorporado dispositivos y estrategias de evaluación al repertorio de políticas que constituyeron sus reformas en la década del ‘90”. Los modos de organización, los niveles alcanzados, el tipo de instrumentos, la periodicidad, y otras variables, muestran diferencias entre ellos. Pero lo cierto es que la evaluación se instaló como una política y adquirió un fuerte reconocimiento social.

Más allá de las diferencias encontradas, pueden observarse elementos comunes entre los diferentes países, que dan cuenta de las opciones técnicas adoptadas, o de las que más allá de haber sido definidas explícitamente, resultaron de los modelos de implementación puestos en marcha.

- En ninguno de los casos se han observado variaciones significativas de los resultados a lo largo del período de implementación. Los niveles de acierto se han mantenido estables tanto en relación a los indicadores cuantitativos cuanto a los cualitativos. Es posible afirmar que los alumnos siguen logrando niveles similares de aciertos en el mismo tipo de ítems. Estos resultados podrían permitir dos tipos de inferencias iniciales, en el sentido de que los cambios curriculares en particular y los programas de re-

forma en general, no han impactado los aprendizajes que los alumnos adquieren en las aulas, o que las preguntas que estamos formulando no son capaces de captar las diferencias que se están produciendo.

- La mayoría de los sistemas han trabajado con mayor dedicación en la producción de información que en el análisis de la misma y en su comunicación a los actores. La región ha implementado millones de pruebas por año, pero es relativamente escasa la cantidad de información que se ha producido, al menos la que se ha utilizado para la reformulación de políticas o estrategias pedagógicas.

- No se ha logrado que los docentes incorporen los resultados de los dispositivos como insumo de trabajo para la reformulación de su tarea. Esto tampoco ha ocurrido en los propios ministerios que las implementan. Los equipos de currículum o de capacitación no hacen de los resultados un insumo principal de sus trabajos. Los operativos de evaluación aparecen como elementos con gran autonomía del resto de las políticas implementadas.

- Los resultados se han comunicado, casi exclusivamente, a través de los medios de comunicación masiva. Se han difundido los promedios nacionales con mayor o menor desagregado, pero en muy pocos casos se han construido informes que puedan orientar el trabajo de la escuela o, al menos, que las escuelas los vivan así. Se han resaltado los casos extraordinarios, se han comparado con la línea histórica de los mismos. En los casos en que se ha trabajado en la distribución, la temporalidad de dicha devolución es tal, que se vuelve difícil transformarla en estrategias de mejoramiento institucional. Los informes llegan tarde o, al menos, desajustados con la posibilidad de transformarlos en herramientas de planeamiento.

- Las escuelas han vivido los operativos como dispositivos externos a los que han dedicado mayor o menor tiempo de preparación, pero en muy pocos casos han percibido esas actividades como propias de la institución.
- Los ítems de contenidos evaluados han sido construidos en el contexto de un proceso de transformación curricular de gran alcance, razón por la cual resulta difícil evaluar los resultados de dichos ejercicios.
- La comunicación de los resultados ha transmitido la sensación del deterioro, la idea de que los sistemas educativos son muy poco eficientes en términos de calidad, y que se encuentran en un proceso de deterioro respecto de los resultados que obtenían en el pasado.

Es preciso relativizar muchas de las críticas que se efectúan sobre los equipos técnicos. Lo cierto es que la región no contaba con ningún tipo de experiencia previa acerca de la organización de operativos de este tipo y la experiencia europea se daba en contextos muy diferentes, al menos de menor heterogeneidad y en condiciones de organización del Estado mucho más sólidas.

Al mismo tiempo, los equipos técnicos no contaron con directivas políticas que les permitieran orientar el trabajo. El pedido de organizar dispositivos de evaluación no contó con el planteo de direccionalidades claras respecto de lo que se haría con dichos resultados, del impacto que se pretendía generar con los mismos, etcétera. Pero, en el caso de los equipos, éstos tampoco lo demandaron y en este sentido es interesante reflexionar acerca de los grados de autonomía de los equipos técnicos y los riesgos que esto genera.

El avance técnico parece haber sido construido con autonomía de las decisiones de sentido. De hecho, la variación

de los instrumentos desde el comienzo de su implementación es baja. Las que se produjeron han tenido que ver con ajustes estadísticos o técnico-pedagógicos. EL elemento innovador apareció en el caso de algunos dispositivos internacionales que incorporaron nuevos modelos y contenidos a los instrumentos.

En síntesis, la reflexión acerca de nuevas decisiones técnicas, ajustes y revisiones que se realizan, debe estar precedida de las reflexiones y decisiones políticas que orienten el trabajo técnico posterior.

A diez años de la implementación de estos dispositivos de evaluación, parece interesante comenzar por una indagación profunda de los resultados, un intento de encontrar pistas que aporten a una revisión de las políticas implementadas y de las nuevas que se elaboren. Resulta importante formular un análisis profundo de los resultados, realizarlo con participación de los actores del sistema, y construir la cultura de la evaluación como insumo para el debate y la construcción de políticas educativas.

En cuanto a las estrategias de futuro, es necesaria una revisión que aproveche los avances alcanzados. Puede resultar riesgosa la tradición de fractura en el modelo de construcción de programas de política que ha tenido la región. No se trata de producir una nueva ruptura, se trata de un proceso de construcción, de re-orientar, de potenciar la herramienta en función de los objetivos que se definan. Latinoamérica requiere institucionalizar sus operativos de evaluación, incluyendo en dicha institucionalización los propios mecanismos de revisión y re-orientación de las decisiones técnicas. Y esos mecanismos deben ganar coherencia con los objetivos de política que se definan.

Una visión acerca de las relaciones entre las evaluaciones y los actores

Las evaluaciones de calidad son uno de los componentes de las políticas encaradas en la década del '90 que más críticas reciben entre los docentes y los directores. Sin duda, el modelo de implementación y construcción no ha logrado que los mismos se sientan incluidos en dicho proceso.

Este dato resulta interesante de ser analizado por el hecho de que tal como lo explica el artículo de Emilio Tenti, en este mismo libro, la visión crítica no se extiende a la política de evaluación como tal, sino al modo en que la misma fue implementada. Los docentes parecen acordar con la idea de evaluar, aunque tienen sus diferencias con el modo en que dichas evaluaciones se han implementado.

Por otro lado, manifiestan que la tarea de evaluar es de las que más dificultades les generan dentro del repertorio de las que implica su propio trabajo. En este sentido, parece interesante observar el hecho de que ellos mismos parecen requerir una reflexión más profunda de para qué evaluar y cómo hacerlo, y esa reflexión no debería estar divorciada de la que se formula alrededor de los operativos de evaluación en un sentido más general, como política global del sistema.

Al mismo tiempo, la propia exposición a la que la escuela se ha visto sometida por la comunicación masiva de los resultados, las sospechas que ha generado en la sociedad acerca de la capacidad de los docentes de garantizar aprendizajes significativos de sus alumnos, la puesta en duda de la idoneidad de los mismos para su tarea, pueden explicar la mirada crítica acerca de los dispositivos de evaluación de la calidad.

Sin duda, las evaluaciones son vividas como un ele-

mento externo que, de algún modo, se interpone entre las autoridades y las escuelas. Deja a los docentes solos y rompe cierta visión de unidad del sistema. Como expresó Javier Bonilla: “Se nota un grado de exterioridad muy grande entre el aparato de evaluación y el de gestión habitual del sistema. La pregunta es cómo hacemos para que el dispositivo sea vivido como un dispositivo interno. Cómo hacemos para que el aparato de evaluación sea amigable para la cultura del sistema educativo”. La sola idea de dos aparatos a los que se debe volver “amigables” y hacerlos dialogar entre sí, implica una ruptura en la lógica de un sistema que se ha caracterizado por importantes grados de endogamia, como es el sistema educativo. Las evaluaciones aparecen como un dispositivo externo al propio sistema, que han adquirido el poder de validar o no sus acciones y efectos. La escuela parecía un ámbito invulnerable y ha quedado sometida a un procedimiento que no sólo la evalúa, sino que expone los resultados de dichas evaluaciones al conjunto de la sociedad.

El hecho de que la discusión previa de las evaluaciones acerca del sentido haya resultado tan débil, hace que tampoco hayan podido participar los docentes de la misma. La revisión que se realiza de las mismas, la discusión de sentido en la que se avance, debe incluirlos, debe contar con ellos en el análisis del “para qué” y del “cómo”. En caso contrario, resulta complejo construir una “amigabilidad” a posteriori. Lo cierto es que la posibilidad de lograr que el sistema viva a los dispositivos como propio, debe incluir la participación de la discusión del para qué, que sienta como propios los objetivos que se proponen.

El Jefe de la división educación del BID, Juan Carlos Navarro, señaló en el Seminario: “Los docentes sienten que entran en un juego en el que los datos están cargados con-

tra ellos (...) Creen que es justo que sean evaluados, no están en contra de esto, por principio”. Tal como afirmamos anteriormente, se trata de comprender que la revisión de sentido en materia de políticas de evaluación, va a incluir tanto las cuestiones sustantivas como las metodológicas, y que la participación de los actores en las distintas instancias de ese proceso es fundamental. Por otro lado, es preciso que los tomadores de decisiones sean transparentes, en el sentido de que no se está construyendo un dispositivo para confrontar con los docentes, sino uno capaz de producir los insumos que permitan mejorar la calidad y la equidad, en los aprendizajes de los alumnos.

Hacia una búsqueda de sentido

La propuesta de realizar una mirada política de las evaluaciones de calidad ha permitido revisar distintos aspectos de estas políticas y proponer miradas novedosas de las mismas. Sin duda, la intuición respecto de que no podía empezarse la revisión por una mirada exclusivamente técnica las acertada. El debate con los tomadores de decisiones les permite salir del “enojo” con el instrumento y pasar a la apropiación de los mismos respecto de su direccionalidad y conceptualización. Resulta claro que en materia de evaluación hemos carecido de una toma de decisiones política, de sentido, y que en realidad hemos asistido a una “no política”. Es decir, las decisiones no tomadas fueron ocupadas por opciones técnicas.

Probablemente este fenómeno no sea exclusivo de las políticas de evaluación de calidad, pero en este caso adquiere especial relevancia por el peso político que han tenido las

mismas en los medios de comunicación, en los espacios políticos y en la visión que las sociedades han construido sobre el estado de la educación.

Aunque en cierta medida estas políticas pueden ser consideradas un caso común entre las que impulsaron las reformas de la década del '90, el hecho de la construcción de cursos técnicos con cierta autonomía de la reflexión política y, en algún caso, de la construcción de sentido compartida con los actores, parece ser una consecuencia más generalizada.

Lo cierto es que en el caso de las políticas de evaluación, resulta importante volver a evaluarlas desde la lógica de la revisión de sentido. Y en este punto, es una responsabilidad de los equipos políticos de los ministerios esta definición, que no puede encontrarse aislada del resto de las decisiones de política educativa. Esa revisión debe incluir, en el caso particular de la evaluación, respuestas a las preguntas de:

- ¿Para qué evaluamos?
- ¿Qué evaluamos?
- ¿Cómo lo hacemos?
- ¿Quiénes participan del proceso de construcción?
- ¿Qué información comunicamos a cada uno de los distintos actores del proceso?
- ¿Qué hacemos con los resultados?

Estas preguntas, que pueden parecer elementales, permitirán ordenar las revisiones técnicas que se realicen. Será distinto afirmar que se evalúa para establecer parámetros de comparación que obliguen a las escuelas a mejorar, que sostener que se lo hace para mejorar el conocimiento del sistema de modo de ajustar las estrategias de política.

Será distinto evaluar a los docentes como modo de mejorar sus estrategias de carrera y de formación, que hacerlo solamente para establecer premios y castigos.

Estas decisiones permitirán decidir por dispositivos muestrales o censales, por instrumentos individuales, grupales, por ítems cerrados o abiertos, etcétera. Lo cierto es que es fundamental recuperar los modos de construir políticas de sentido, definición de fundamentos que puedan sostener los acuerdos técnicos posteriores.

Algunas de estas reflexiones aparecerán en este libro, que se propone abrir un debate sobre las evaluaciones que sume miradas y resulte un modo de enriquecer la política educativa en general y la de evaluación en particular.

Encuentros y desencuentros con los procesos de evaluación de la calidad educativa en América Latina

Javier Bonilla Saus

I. Introducción

En las últimas décadas, tanto los responsables de los sistemas educativos públicos y privados, como la mayoría de los especialistas en educación, han ido adoptando, de manera crecientemente generalizada, la idea de que es necesario evaluar la calidad educativa de la enseñanza que se imparte a la población de los países de la región.

En América Latina el fenómeno es relativamente nuevo puesto que comienza a expandirse aproximadamente a partir de la década del '90 y lo hace, indudablemente, como la continuidad de un movimiento relativamente generalizado y mucho más amplio que comenzó a desarrollarse en los Estados Unidos en los años '50 para luego extenderse a toda Europa en la década del '70 y, posteriormente, a un buen número de países del Asia y Oceanía.

Algo más de una década después de la aparición de aquellos primeros entusiasmos en nuestro continente, hemos aquí reunidos para plantearnos una serie de preguntas que, bien miradas, deberían quizás haber precedido a aquellos enfáticos compromisos con el movimiento generalmente com-

partido por la utilización de los sistemas de evaluación de la calidad educativa.

II. Los problemas planteados por la evaluación educativa

Efectivamente, estamos aquí reunidos porque el desarrollo de las herramientas de evaluación de los sistemas educativos, no solamente ha constituido una operación costosa –tanto en términos de recursos económicos como en materia de esfuerzos y de recursos humanos–, sino que ha resultado ser una opción, aparentemente técnica, de efectos políticos problemáticos.

Hoy se nos pregunta si creemos que las evaluaciones han afectado la inversión pública de los gobiernos en educación, si las evaluaciones conllevan riesgos políticos, etc. Pues bien, en el Seminario que aquí nos congrega, de lo que se trata es de preguntarse por qué razones la opinión pública –y/o los medios de **prensa**¹– de nuestros países encuentra tantas dificultades para interpretar los resultados de distintas pruebas de evaluación de la calidad educativa de un determinado aspecto del funcionamiento del sistema educativo y, por ejemplo, muchas veces, terminan confundiéndolos con un

¹ Estando este artículo en vías de terminación, acaba de aparecer una publicación cuyas conclusiones podrían haber enriquecido nuestro trabajo: *El papel de los medios de comunicación en la formación de la opinión pública en Educación*, de **Juan Carlos Tedesco** y **Roxana Morduchowicz**. Tanto por nuestra voluntad de conservar el espíritu de las discusiones del Seminario que dieron lugar a esta publicación, como por razones de tiempo, no nos pareció pertinente introducir las hipótesis y conclusiones de esa publicación en este artículo.

verdadero diagnóstico general del estado de la educación en nuestros respectivos países.

En efecto, estos son algunos de los problemas que parece acarrear la multiplicidad de evaluaciones de los resultados educativos que hemos llevado a cabo y estamos llevando adelante en los últimos años. Pero, muy en especial, creo que lo que más preocupa es el último de los problemas mencionados. Es decir, el problema que conlleva la dificultad de transmitir a la opinión pública cuál es el verdadero estatuto del resultado de una evaluación y la aparente imposibilidad de que ésta dimensiona exactamente su significado, sus alcances y sus reales implicaciones.

II.A. Los resultados de las evaluaciones educativas ante la opinión pública

En primer lugar, es necesario señalar que el fenómeno es preocupante porque, en buena medida, la opinión pública parece reaccionar de manera relativamente *diferente* ante los resultados de las evaluaciones educativas que frente a otro tipo de mediciones. En efecto, en ocasiones, hay mediciones que son portadoras de indicadores de realidades económicas o sociales mucho más trascendentes –y por ello más inquietantes– que la persistencia de una alta tasa de repetición en primer año de Primaria o de la tasa de deserción del segundo ciclo de Enseñanza Media. Sin embargo, es necesario constatar que muy pocas veces las reacciones de los medios y de la opinión pública alcanzan, ante la difusión de distintos tipos de indicadores, la virulencia que suelen adquirir frente a algunos resultados provenientes del sector educativo.

Es así que debemos reconocer que las evaluaciones

educativas son relativamente nuevas en el horizonte de la gestión gubernamental. Los estados modernos contemporáneos –y con ello quiero decir, *grosso modo*, los aparatos estatales concebidos y diseñados después de la Segunda Guerra Mundial– hace tiempo se empeñaron en cuantificar los procesos públicos y privados que tenían lugar en nuestras sociedades. En gran medida esta tendencia está ligada, en un sentido, al proceso de tecnificación de los aparatos estatales y al incremento de la importancia de los medios de comunicación de masas. Cabe simultáneamente destacar, en un segundo sentido, que la tendencia a informar sobre los resultados de la marcha de los principales procesos que se desarrollan en la sociedad está íntimamente vinculada al proceso mismo de consolidación de la democracia.

Esta requiere, de manera regular –y yo diría que **inevorablemente**²–, de un flujo de información continuo hacia los medios para permitir la conformación de ese actor decisivo de todo régimen democrático que es la opinión pública.

Es así que la tasas de inflación, la de desempleo, la de inversión, los niveles de ingreso, el índice de crecimiento o

² El tema de la “publicidad”, y hoy diríamos de la “transparencia” de una gestión democrática y de sus acciones y resultados, ya ha sido implacablemente planteado en múltiples ocasiones. Quizás la referencia al tema, a la vez más seria y más conocida, sea la de **Norberto Bobbio** –véanse los capítulos *El poder invisible* y *La democracia y el poder invisible*; pp 22-24 y 65-83 en *El futuro de la democracia*, Ed. Fondo de Cultura Económica, México, 1984-. Véase igualmente, para una perspectiva más histórica del problema; **Dunn, John et al.**: *Democracy: The unfinished journey*, especialmente el capítulo 13 de conclusión; pp. 239-266. Ed. Oxford University Press, 1992 y **Eccleshall, Robert et. al.**: *Ideologías políticas*, Ed. Tecnos, Madrid, 1993, pp 182-183, que aborda el problema desde la perspectiva de la necesidad de la democracia de suprimir el “secretismo” y fomentar distintas formas de participación.

disminución del salario **real**³, etc., son indicadores vinculados a la economía que se publican rutinariamente en todos los medios de prensa de nuestros países sin que ello signifique sorpresa o alerta alguna siempre que se mantengan dentro de los rangos más o menos “esperados” por la opinión pública.

Algo muy similar sucede, por ejemplo, con los indicadores vinculados a la salud pública. Aunque menos concurridos que los indicadores económicos por la curiosidad de los medios de prensa, no es menos cierto que los índices de natalidad, los indicadores de mortalidad infantil, los de desnutrición o los de morbilidad vinculados a determinadas enfermedades o accidentes parecen haber ganado una suerte de carta de ciudadanía entre la opinión pública y se han integrado al lenguaje corriente de nuestras sociedades.

Que lo que estas mediciones indiquen sea “bueno” o “malo” resulta para la opinión pública relevante pero, en algún sentido, no precisamente *inquietante* si esos resultados están dentro de un rango considerado por **ésta**⁴ como “normal”. Pero, en todo caso, difícilmente el asunto no será más que noticia pasajera y, por lo general, en ninguno de estos procesos las informaciones y mediciones transmitidas a la ciudadanía son generadoras de grandes movimientos de opinión que pretendan poner en cuestión las raíces mismas del sistema económico o las grandes políticas que rigen el sistema de salud pública.

Pero con las evaluaciones de la calidad de la educación

³ Recordemos que acabamos de incorporar durante la última década, en la terminología periodística, nada menos que el sorprendente indicador de “riesgo país” dentro de la batería de informaciones económicas con la que alimentamos a la opinión pública.

⁴ Y debemos admitir que, a veces, la definición de esa “normalidad” se lleva a cabo algo arbitrariamente.

las cosas parecen ser, muchas veces, distintas. Si, por alguna razón, el sistema educativo público tiene que informar que las tasas de repetición han subido algún punto porcentual, que el “*drop-out*” se ha incrementado o que los resultados de las evaluaciones en Matemática deben ser considerados insuficientes, esa información tiene una alta posibilidad de desembocar en una verdadera campaña de prensa acompañada de largos cuestionamientos sobre la política educativa en general, sobre la idoneidad del Ministro de turno, sobre la validez de los datos, sobre la capacidad del cuerpo docente, etc.⁵ Y seguramente, además, nunca estarán ausentes como infaltables anexos, argumentaciones no menos extensas sobre “*la crisis*” de la educación, los efectos aparentemente devastadores de las nuevas tecnologías de la información globalizada y los problemas más abstrusos de la civilización contemporánea.

Incluso, *a contrario sensu*, cuando los resultados que arrojan las evaluaciones sobre la calidad educativa resultan ser **positivos**⁶, las reacciones de los medios y de la opinión

⁵ Creo que conviene, desde el inicio, señalar que esta peculiar hipersensibilidad ante los resultados de las mediciones en educación no son privativas de los países de nuestra región. Los recientes resultados obtenidos por Alemania en las pruebas del sistema PISA han significado un verdadero escándalo nacional y recordemos que Francia entera sigue, año con año, los resultados del examen final de la Enseñanza Media que son, íntegramente y nombre a nombre, publicados por la prensa nacional.

⁶ Como simple anécdota referida al sistema educativo de nuestro país, basta recordar que, hace escasamente unos meses, las evaluaciones de desempeño en Lengua y Matemática de 6o. año de Enseñanza Primaria hechas en Uruguay arrojaron, respectivamente, mejoras de casi el 10% y el 14% con respecto al desempeño obtenido en iguales mediciones del año 96 y, también respectivamente, mejorías del 5% y del 7.5% con respecto a mediciones del año 1999. Alguna prensa sencillamente ignoró la información mientras que otra, en lugar de destacar la mejoría alcanzada y sobre todo la consistencia

pública pueden llegar a ser sorprendentes: la opinión pública parece reaccionar como si los resultados positivos fuesen los “naturalmente esperables” por lo que, de manera explicable, dichos resultados difícilmente constituyen una “noticia” significativa para los medios.

II. B. Algunas tentativas de explicación

Creo que si todos los aquí presentes, que provenimos de países muy disímiles, estamos más o menos de acuerdo sobre la existencia de este peculiar fenómeno que podríamos llamar de “desinteligencia” para con los resultados de las evaluaciones educativas; si estamos genéricamente de acuerdo que parece haber un problema a nivel de las lecturas que la opinión pública hace de los resultados de las evaluaciones de la calidad de la educación, corresponde que ensayemos al menos alguna explicación sobre la cuestión.

II.B.1. En primer lugar creo que es necesario recordar algo que, por obvio que pueda ser, no deja de resultar a la vez significativo y pertinente para el problema que nos ocupa. Las evaluaciones de los resultados educativos –en particular las de aquellos países donde el sistema de educación pública es francamente mayoritario y simultáneamente portador de una cierta tradición de prestigio– tienen una estrecha relación con fuertes intereses corporativos de distinto tipo. Mencionemos al menos dos de los más relevantes.

de la tendencia que las cifras indicaban, subrayó la relatividad de los resultados absolutos alcanzados ya que entendió que el 66.3% de suficiencia en Lengua y el 48,3% en Matemática debían ser considerados, sin más argumentación, como un resultado claramente insatisfactorio.

Por un lado, muchas veces la educación privada requiere –al menos en países cuyos sistemas educativos tienen las características mencionadas– de la permanente afirmación de que sus resultados resultan ser superiores a los de la educación pública. En muchos casos su estrategia de desarrollo se basa en una reafirmación discursiva –más o menos consistente con la realidad– de su capacidad de obtener resultados educativos superiores a los de la educación pública. De lo contrario, de no reafirmar esta supuesta superioridad en la calidad educativa de la educación privada, la razón de ser de ésta queda reducida casi exclusivamente a la oferta de determinadas opciones confesionales o a la de algunas peculiaridades lingüísticas y culturales. En ese sentido, el discurso de la educación privada tiende a contribuir, de manera más o menos explícita, a fortalecer toda tendencia que favorezca una lectura crítica de los resultados de la educación pública.

Por el otro, las evaluaciones educativas también coliden parcialmente con los intereses de una muy poderosa corporación docente que, al menos en algunos de nuestros países, no está habituada a que los resultados de su actividad profesional sean evaluados y públicamente expuestos. Ante todo, las evaluaciones chocan con las tradiciones y con los intereses de la corporación docente porque una gran parte o todo el proceso de evaluación educativa, de la manera que se ha ido instrumentando en los nuevos procesos de reforma, escapa política y administrativamente al control de los centros de poder de la corporación.

Pero, además, las evaluaciones que terminan haciendo públicos los resultados educativos, de alguna manera ponen “en cuestión” –y, lo que es peor, lo hacen de manera sistemática, puesto que los sistemas de evaluación están concebidos

para funcionar de manera regular y permanente– su desempeño como docentes profesionales por más que la publicación de los resultados se lleve a cabo de la manera más general, promediada y anónima **imaginable**⁷ y no permita la identificación de docentes, centros educativos o unidad específica alguna dentro del sistema educativo.

Y creo que, en este sentido, cabe volver aquí a la comparación con la situación y las reacciones que se producen en la educación frente a las que se suscitan en otros ámbitos de la actividad pública, como la salud pública, cuando de medir resultados se trata.

En primer lugar, tanto o más poderosa que la corporación docente resulta ser la corporación médica de cualquiera de nuestros países. No he sabido, sin embargo, de que dispongamos –al menos en el Uruguay estoy seguro de que no es así– de algún mecanismo de evaluación de la eficacia y eficiencia de los hospitales públicos (o privados), y eso marca, desde el vamos, una diferencia importante.

En segundo lugar, y como síntoma mucho más importante de esas diferencias, a pesar de que en muchos países de América Latina, en los últimos años han hecho su aparición enfermedades previamente inexistentes o erradicadas como el cólera, el dengue hemorrágico –o la misma aftosa a nivel animal–, no tengo información de que el aparato de salud pública –o el de sanidad animal puesto que, para al caso, la circunstancia es similar–, haya sido radicalmente cuestio-

⁷ En países como el que hoy tan amablemente nos recibe, donde se ha establecido que las evaluaciones están autorizadas a “rankear” los resultados de los distintos establecimientos educativos, imagino que los cuerpos docentes han de percibirlos, además, como una fuerte amenaza a la unidad de la corporación de maestros y profesores.

nado en sus políticas y responsabilizado directamente por los medios y la opinión pública de estas calamidades.

Todo parece entonces acontecer como si la opinión pública concibiese que hay procesos que, por más que exhiban indicadores que muestran falencias o resultados de poca calidad, su actividad pertenece a un ámbito considerado de carácter “natural” –frente a los que poco pueden hacer al respecto los tomadores de decisiones y los responsables políticos– y, por lo tanto, el público tiende a eximir a los servicios estatales involucrados de toda responsabilidad de dichos acontecimientos o resultados.

Sin embargo, la misma opinión pública no parece admitir que pueda resultar también “natural” que exista una suerte de complejo pero explicable “trade-off” entre la expansión abrupta de la matrícula de determinado nivel del sistema educativo y la aparición de problemas de calidad en ese mismo nivel de educación, y se muestra particularmente sensible ante la constatación estadística de este tipo de problemas.

II.B.2. Una segunda explicación plausible para esta gruesa dificultad que se genera con la transmisión de los resultados de las evaluaciones educativas puede tener que ver con la imagen tradicional que la opinión pública tiene de la educación pública y, fundamentalmente, de la escuela pública.

Yo no estoy seguro de que esta argumentación sea aplicable a todos los países y no sea ésta una visión fuertemente “uruguaya” de un problema que, en realidad, es más general. Pero, en el caso del Uruguay, es evidente que la instauración de estos nuevos procesos de evaluación de la educación se dan de bruces con la creencia fuertemente arraigada en la opinión de que la escuela pública uruguaya es una

institución sistemáticamente exitosa⁸ y que se encuentra, en última instancia, por encima de cualquier evaluación.

En ese sentido, en el Uruguay, pero estoy seguro que hay casos parecidos en América Latina, la absoluta certeza del público de que la escuela pública –y generalmente esta certeza se refiere a la “imagen” de una escuela tradicional que hasta hace una década no hacía este tipo de evaluaciones– era y es un dechado de virtudes, conspira directamente contra la buena recepción de las políticas de evaluación sistemática en materia de resultados de la educación.

En algún sentido la opinión pública se ofusca ante la emergencia en la última década de estos “reformadores evaluadores”, que parecen llegar sólo para anunciar malas noticias. El público recibe un mensaje del tipo: “...*la enseñanza que creíamos tan buena ahora resulta que no lo es...*”. Como se comprende fácilmente, se trata de un discurso de difícil asimilación y la explicable tendencia a responsabilizar al mensajero por las malas noticias de las cuales él es portador, seguramente opere en muchos casos.

I.B.3. Una tercera hipótesis que, entendemos nosotros, debe tenerse en cuenta a la hora de explorar los desencuentros entre las evaluaciones educativas y la opinión pública de nuestros países tiene que ver con lo que podríamos llamar “*un problema de legitimidad*”. O, si ustedes lo prefieren, un problema entre el tipo de legitimidad sobre el que se asientan las

⁸ Exitosa en múltiples aspectos: exitosa a nivel individual como mecanismo socialmente igualador y hasta de ascenso social personal, exitosa a nivel cultural como institución generadora de una sociedad que se autoconcebe como “cultura”, y exitosa a nivel político como institución constituyente de una tradición democrática que forma, legítimamente, parte del núcleo fuerte de la identidad nacional.

instituciones educativas tradicionales y la emergencia de este nuevo tipo de evaluación que se ha ido generalizando.

En efecto, conviene recordar que educación y evaluación son dos actividades que están inextricablemente vinculadas entre sí desde hace ya mucho tiempo, si no es que desde el fondo mismo de la historia de la actividad educativa. Bajo las más diversas formas, maestros y profesores, dentro del tipo de institución educativa que fuere –escuelas carolingias o catedralicias, escuelas de “métiers” dependientes de las corporaciones medioevales, universidades de las más diversas órdenes eclesiásticas, tecnológicos republicanos, etc.–, han aplicado desde siempre a sus alumnos muy variadas modalidades de evaluación para confirmar la adquisición de saberes por parte de aquellos. Estas modalidades de intentar registrar los aprendizajes pueden haber sido más o menos racionales, más o menos adecuadas o, incluso, más o menos arbitrarias, pero no hay ninguna novedad en cuanto a que la educación evalúe los desempeños de los profesores o los saberes y los niveles de aprendizajes obtenidos por sus **alumnos**⁹.

Pero conviene subrayar que este tipo de evaluación, cuya legitimidad está asentada probablemente desde hace siglos, se lleva adelante esencialmente en el *espacio del aula* o de su equivalente y como una actividad estrictamente interna al proceso educativo mismo. O si se quiere, más precisamente, un mecanismo de evaluación que se despliega exclu-

⁹ Es más, desde que emergieron, hacia finales del siglo XIX, los sistemas públicos de enseñanza masiva, se generalizó lentamente el uso del “examen” –y a veces del examen general a nivel nacional–, para aquilatar los conocimientos adquiridos por los alumnos y, casi simultáneamente, se puso en marcha la construcción de cuerpos inspectivos institucionalmente destinados a supervisar el funcionamiento del proceso educativo de acuerdo a criterios más o menos atinados.

sivamente en el espacio de las relaciones entre maestro o profesor y alumnos y entre supervisor o director y maestros o profesores.

En este tipo de evaluación, la noción de “*resultado*” es algo que emerge del seno mismo de la relación maestro-alumno y, por lo tanto, los parámetros que definen un resultado como “exitoso” o “deficiente” son parámetros que surgen como parámetros *internos* a la actividad del aula, al centro educativo, y, eventualmente, al régimen de supervisión del sistema educativo. En suma, son *evaluaciones cuyos resultados se construyen sin que intervengan criterios ni agentes exteriores al sistema educativo tradicional*. Dicho en otros términos: las evaluaciones tradicionales en la educación poseen una amplia legitimidad, porque forman parte integrante del proceso de enseñanza-aprendizaje desde hace siglos, pero, sobre todo, porque en esas evaluaciones son los propios agentes del sistema educativo que se evalúan a sí mismos¹⁰.

Pero, en realidad, no es éste el tipo de evaluación que nos congrega aquí y no son estos mecanismos de evaluación los que nos preocupan por sus efectos en la opinión pública y, por ende, en la política. La evaluación y los resultados que hacen problema son aquellos que, precisamente, no se procesan “*in vitro*”, ni en el ámbito prerregulado del aula, ni en los

¹⁰ Si pensamos un instante en lo que esto significaría en otros ámbitos del quehacer de un servicio público, sus terribles limitaciones quedan inmediatamente al desnudo. Es como si eliminásemos el concepto de “oposición de intereses” de los sistemas de contralor contable o de auditoría de gestión. Por ello, conviene tener en cuenta no solamente el carácter “endógeno” de este tipo de evaluación: es necesario advertir que, así concebida, la evaluación educativa tradicional evita, de manera casi sistemática, la opinión sobre el funcionamiento y los resultados del sistema educativo que puedan tener los alumnos, los padres, la comunidad y hasta los representantes electos del poder político.

espacios jerárquicos del centro educativo o de la estructura corporativa de los sistemas de enseñanza.

Los resultados que producen “desinteligencias” con los medios y con la opinión pública son aquellos provenientes de censos, encuestas, muestreos, “panels”, etc., obtenidos y dirigidos por personal –docente o no, eso es secundario–, pero que tienen como objetivo medir el desempeño del sistema educativo *desde fuera* de las rutinas, las jerarquías y, en última instancia, los intereses de la corporación educativa.

Es entonces por ello que las evaluaciones de la actividad educativa así construidas, desde un punto de vista *explícitamente externo a la lógica tradicional del sistema educativo mismo*, plantean un problema de “legitimidad” y, en consecuencia, plantean en muchos casos problemas políticos significativos.

En particular en el seno del cuerpo docente y sus jerarquías, las preguntas obvias que se disparan ante la instauración de esta nueva mirada sobre el desempeño de la educación, son los siguientes: ¿Quién mejor que yo, profesor o maestro, puede evaluar los conocimientos que transmito? ¿Quién mejor que yo, docente, puede juzgar a mis alumnos? ¿Quién mejor que yo, director, puede juzgar el funcionamiento del centro educativo? ¿Quién mejor que yo, inspector o supervisor, puede medir efectivamente el desempeño de los **profesores**¹¹?

I.B.4. Por último, existe una cuarta hipótesis que conviene considerar en la búsqueda de las razones por las cuales las

¹¹ Y una vez más creo que resulta pertinente insistir con el paralelismo entre este sistema de “autoevaluación” tradicional que ha practicado la educación, con algunas falencias de control democrático sobre el accionar de la corporación médica y los resultados de éstas.

evaluaciones de la calidad educativa presentan tantas dificultades a nivel de la opinión pública. Se trata de una hipótesis que es necesario explorar en un nivel más conceptual o, si se quiere, más abstracto.

Cabe preguntarse hasta qué punto esta política de evaluación sistemática de los resultados educativos no introduce una incongruencia importante con algunos de los conceptos fundamentales de la educación pública tradicional de nuestros países. Si nosotros seguimos manteniendo acertadamente, como uno de los principios rectores de nuestra educación pública, el principio de *educar para la equidad*, no resulta sorprendente que la opinión pública imagine al sistema educativo como un servicio que imparte enseñanza de calidad homogénea y que los aprendizajes que en ella se obtienen deberían ser razonablemente homogéneos.

La publicación sistemática de resultados que exhiben la existencia de *diferencias* entre los aprendizajes de los alumnos, entre los aprendizajes conseguidos en los centros, entre los resultados educativos en las diferentes regiones de un país, etc., introduce una perspectiva que contradice la lectura ingenua del principio de *educar para la equidad*.

Que nuestros sistemas educativos eduquen para la equidad y se esfuercen sistemáticamente en trabajar en ese sentido, no significa –ni nunca significó– que obtengan ni puedan obtener resultados homogéneos, porque ningún sistema educativo –y, menos aún, sistemas educativos de masas como lo son los sistemas educativos contemporáneos–, puede compensar totalmente diferencias económicas, sociales y culturales que trascienden la capacidad democratizadora de los sistemas de enseñanza.

Otra cosa sería si nuestros sistemas educativos estuviesen concebidos –como efectivamente en ciertos países

pueden estarlo– y fundados en otros principios. Si decidiésemos fundar y hacer descansar la razón última de la educación pública sobre la búsqueda de *la eficiencia* o sobre el principio de *educar para la competitividad*, seguramente que, en ese escenario, la incongruencia a la que nos referimos más arriba sería menor o menos perceptible. De hecho dejaría de ser una incongruencia, ya que las políticas de evaluación deberían de ser reclamadas como una parte *necesaria y fundamental* de la orientación básica de una educación que tiene como objetivos centrales la eficiencia y/o la competitividad.

Entonces yo agregaría, como último elemento que intenta explicar las desinteligencias y las incomprensiones que constatamos entre el sistema educativo y la opinión pública en materia de evaluaciones de la calidad de la educación, este aspecto que tiene dos componentes. Un componente meramente informativo –la creencia del público de que *educar para la equidad* consiste en generar resultados homogéneos que hagan desaparecer las diferencias de origen social y cultural de los alumnos–, y un problema, de trasfondo filosófico, que hace a la concepción última que rige al sistema educativo de nuestros países.

Evidentemente, aunque nos aferremos como nos aferremos al principio cardinal de que es necesario *educar para la equidad*, no podemos por ello renunciar a las evaluaciones educativas aunque sus resultados puedan aparecer como demostrativos de que una equidad estricta no reina en la educación. En primer lugar, porque el principio de *educar para la equidad* se basa en la idea de brindar *igualdad de oportunidades* mediante la educación y no en la de la *igualdad de resultados* de los aprendizajes y, en segundo lugar, como veremos inmediatamente, porque la herramienta de las evaluaciones educativas constituye un componente fundamental

para el desarrollo de cualquier sistema educativo moderno. Particularmente para aquellos sistemas educativos que, orientados filosóficamente hacia *la equidad*, han de estar permanentemente monitoreando los resultados obtenidos en la materia.

III. Las razones de la evaluación educativa

En uno de los más recientes diccionarios de Pedagogía, editado en Francia¹², encontré la siguiente definición del término “evaluación”: *Dispositivo que tuvo nacimiento hacia mediados del siglo XX en los países anglosajones y que se genera por la transposición, hacia el ámbito de la enseñanza, de conceptos y modelos ya aplicados desde el siglo XIX en temas relativos a la economía y, particularmente, a la producción. En educación, este dispositivo se sitúa en el “marco de la puesta en funcionamiento de una pedagogía por objetivos. La evaluación apunta, por medio de métodos científicos, a apreciar con objetividad los efectos de una acción pedagógica...”*¹³.

Admitamos que, aunque aquejada de cierto esquematismo y fuertemente marcada por algunas querellas ancestrales que separan a la tradición intelectual de nuestros amigos franceses del pensamiento anglosajón, la definición es ciertamente portadora de algo que podríamos llamar una “certera ingenuidad”.

¹² *Dictionnaire de Pédagogie*; Arenilla, Louis; Gossot, Bernard; Rolland, Marie-Claire et Roussel, Marie-Pierre. Ed Bordas-Pédagogie, Paris, 2000.

¹³ Op. cit., pp. 125-126. (Trad. JBS).

En efecto: ¿de cuándo aquí?, y ¿por qué nos ha dado por “medir” en el terreno de la educación? Y medir qué: ¿la educación?, ¿los resultados de ésta?, ¿los aprendizajes?, ¿el desempeño de los docentes?, ¿las políticas educativas?, ¿los costos de la educación? Y podríamos seguir construyendo una larga lista de temas susceptibles de medición porque, en realidad, una vez adoptada esta postura epistemológica que instituye una mirada “sobre” –y no ya “desde”– el sistema educativo que se pregunta sobre el ¿cuánto?, el ¿cómo?, el ¿dónde?, etc., debemos admitir que todos los aspectos del quehacer educativo pueden ingresar en el terreno de la evaluación.

Creo que deberíamos ser conscientes de que el ingreso de la cultura de la evaluación sistemática, tanto en el discurso pedagógico como en las prácticas de los sistemas educativos, constituye un cambio de trascendencia porque establece de manera implícita y no dicha, una concepción sobre cómo debe ser analizada la educación que, lo menos que podemos hacer, es tornarla *explícita*.

En efecto, creemos que el asunto es problemático porque una vez que aceptamos ese cambio en la mirada pedagógica al que nos referíamos más arriba, es necesario reconocer que estamos admitiendo que el complejo conjunto de mecanismos de formación intelectual, cognitiva, ética, actitudinal, etc., del niño o del joven que constituyen eso que llamamos “educación” están compuestos por procesos de naturaleza “cuantificable”. Y esta aceptación no es una aceptación menor porque quien admite que un proceso es “cuantificable” admite, simultáneamente, que ese proceso puede ser reducido –creo que este es el mejor término que se puede emplear– a algún tipo de unidades homogéneas que son la condición de posibilidad de toda “cuantificación”.

Y una vez que estamos en el terreno de la cuantifica-

ción, una vez que –al menos una gran parte de– la compleja trama de fenómenos que integran lo que genéricamente llamamos el proceso de enseñanza-aprendizaje es declarado reductible a unidades homogéneas que permiten la mensuración, el conteo, el tratamiento estadístico, etc., el desarrollo de la cultura de la evaluación ya es sólo un problema de construir herramientas metodológicas cada vez más afinadas.

Decíamos que la definición que nos sirvió de punto de partida podía ser considerada como portadora de una “certera ingenuidad” porque, más allá de que reenvía atinadamente el origen de la evaluación educativa a la evolución e influencia del pensamiento anglosajón –en realidad, el tema adquiere una fuerza particular en el sistema educativo de los EE.UU.¹⁴–, al mismo tiempo imagina una suerte de “transposición mecánica” de modelos de investigación derivados de “la economía” y de “la producción”. En sentido estricto, la definición no es errada cuando se refiere al origen de la idea, pero no resulta feliz en su relacionamiento de la evaluación educativa con las metodologías de investigación derivadas de “la producción” o “la economía”.

Y la definición no es errada porque lo que subyace efectivamente en el trasfondo de la tentativa de desarrollar la evaluación educativa de manera sistemática, es la idea de introducir, en el ámbito del discurso pedagógico, la idea de la vigencia posible de una *racionalidad técnica*¹⁵ radicalmente

¹⁴ Es en los Estados Unidos, y concretamente entre las últimas décadas del siglo XIX y las primeras del XX, donde se pone en marcha el complejo proceso de concepción de lo que en el futuro sería la evaluación educativa en el sentido contemporáneo. Es, sin embargo, correcto afirmar que la implantación sistemática de métodos de evaluación en la educación norteamericana data de los años '50 y, particularmente, de los '60.

¹⁵ Todos somos conscientes, como vimos, de que el concepto de evaluación

ordenadora del quehacer educativo. Si leemos atentamente la definición convocada, ella misma revela explícitamente el fondo del problema: “*La evaluación apunta, por medio de métodos científicos, a apreciar con objetividad los efectos de una acción pedagógica...*”.¹⁶

Es efectivamente aquí donde yace el nudo central de la cuestión de esa recurrente propuesta de introducir una racionalidad técnica en el proceso de evaluación de la actividad educativa. El recurso a la “cuantificación” para evaluar, el echar mano a un instrumental técnico, esencialmente estadístico, altamente formalizado, etc., tiene como trasfondo la certeza, no siempre abiertamente explicitada de que, en materia educativa, nos estamos moviendo en un terreno “científico”. En otros términos, estamos reconociendo la existencia de eso que se ha dado en llamar “*ciencias de la educación*”.

Por los tiempos que corren, en los que el estallido del paradigma científico tradicional en una multiplicidad de enfoques epistemológicos que reinvidican el carácter de “científicos” con la más sorprendente ligereza, no consideramos apropiado poner radicalmente en cuestión la existencia de esas “*ciencias de la educación*” que, además de haber adquirido carta de ciudadanía en los sistemas educativos más modernos, la han obtenido, también, hacia los años ‘70, en muchos ámbitos universitarios e, incluso, en los institutos de investigación de dichas instituciones.

Pero una vez establecido esto, no es menos claro que

siempre estuvo presente en el pensamiento pedagógico, pero lo que resulta novedoso es la irrupción, en la noción de evaluación educativa sistemática, de la idea de una evaluación basada en un tipo de racionalidad exclusivamente técnica que se plantea “desde fuera” de la práctica cotidiana de los agentes del sistema educativo.

¹⁶ Ibidem.

no sería serio dejar de recordar que el estatuto científico de las llamadas “*ciencias de la educación*” constituye, en el mejor de los casos, un estatuto de definición **problemática**.¹⁷

El primer elemento que devela esta dificultad emerge, casi a nivel de síntoma, a través del uso sistemático del plural para intentar designar ese campo del **conocimiento**¹⁸: prácticamente nadie habla de “ciencia de la educación”. En el caso de querer utilizar el singular, se echa mano a una disciplina que posee otra historia y, sobre todo, otros títulos: la pedagogía.

Estamos, entonces, ante la evidencia de que, detrás de

¹⁷ Para una excelente aproximación a la cuestión de la problemática constitución de las “ciencias de la educación” y el desarrollo de la investigación y evaluación educativa, véase: **Condcliffe Lagemann, Ellen**: *An Elusive Science: The troubling History of Education Research*, University of Chicago Press, 2002.

¹⁸ El plural en la designación de los distintos campos del conocimiento científico consolidado aparece, en general, solamente en dos casos. Un primer caso es cuando intentamos referirnos, específicamente, a una efectiva pluralidad de disciplinas científicas. Así podemos hablar de “ciencias formales”, de “ciencias naturales”, “fácticas”, “sociales”, etc. Pero nadie ignora que, fundamentalmente en el terreno de las disciplinas científicas “duras” o de mayor tradición, la física es *La* física, la química es *La* química e, incluso más allá de las ciencias “duras”, la historia es *La* historia. El segundo caso en el que el recurso al plural para designar un campo del conocimiento razonablemente formalizado es plausible, aparece cuando, en el seno de ese campo, coexisten paradigmas explicativos diferentes –incluso contradictorios entre sí–. Sin embargo, la existencia de una geometría eucliana, de un álgebra booleana, de una física cuántica, de una biología molecular o de una psicología freudiana o conductista pueden convocar la necesidad de hablar de geometrías, álgebras, físicas, biología o psicologías, pero evidentemente la significación de estos plurales no es la misma. En realidad, para el caso que nos interesa, sólo se nos ocurre el ejemplo de las llamadas “*ciencias políticas*” que, al igual que las “*ciencias de la educación*”, parecen tener que recurrir al plural fundamentalmente por un problema de indefinición de su campo de estudio y, por ende, por un problema de identidad como disciplina autónoma.

la emergencia de estas nuevas y problemáticas técnicas de evaluación e investigación educativa –*versus* los métodos tradicionales de evaluación que se han dado los educadores desde siempre– hay una transformación significativa, hay un cambio de trascendencia porque en el fondo de la cuestión están en juego dos concepciones sobre lo que *significa* el estudio de la educación y, como decíamos anteriormente, estas dos concepciones corresponde tornarlas *explícitas*.

Ya vimos que, tradicionalmente, cuando se quiso hablar de la disciplina que se ocupaba de la educación, se echaba mano al término “pedagogía” y esta pedagogía funcionó durante siglos como una disciplina anexa a la filosofía o, si se desea más precisión, a la filosofía de la educación. Pero vimos también que, simultáneamente, hoy estamos ante el desarrollo de las llamadas “ciencias de la educación” que, más allá de las dudas que pueda suscitar su calificación como “científicas”, han hecho acto de presencia y se han instalado en el discurso educativo para quedarse.

Lo que conviene entender es que, en el tránsito de “la pedagogía” a “las ciencias de la educación”, no estamos simplemente ante un simple cambio terminológico o frente a una mera moda académica. Estamos ante dos perspectivas de evaluación y de investigación muy *diferentes*.¹⁹

Así, ante una demanda de evaluación “pedagógica”, lo que encontramos generalmente es la necesidad de hallar mecanismos aptos para mejorar o adecuar una amplia serie de problemas prácticos que el docente observa en su actividad dentro del proceso de enseñanza-aprendizaje: ¿Cómo disponer de las herramientas para mejorar la enseñanza de las

¹⁹ Que se reflejan, con relativa claridad, en los dos tipos de evaluación educativa a los que nos hemos referido anteriormente.

ciencias? ¿Cómo hacer para que los alumnos superen sus problemas de ortografía?

Quizás, echando mano a la experiencia de nuestro país y sus problemas actuales en materia educativa, sea interesante mostrar un ejemplo real y contemporáneo. En el Uruguay estamos convencidos de que tenemos un problema significativo en la enseñanza de la Matemática. Pero este convencimiento nace de una serie de datos e informaciones que no han sido sistemáticamente recogidos y analizados. Pero aun en el caso de que ya tuviésemos ese paso concluido: ¿Cuál habrá de ser el procedimiento que permita definir la noción de “una enseñanza correcta de la Matemática”?

En primer lugar, por un lado está el camino de salir a la búsqueda de mejores métodos, herramientas e instrumentos, para mejorar el proceso de enseñanza-aprendizaje de la Matemática. Ese camino nos conduce a evaluar y a trabajar en el terreno de la pedagogía. Pero, en segundo lugar, evidentemente surge la pregunta sobre cuáles habrán de ser los criterios que fundarán esa “enseñanza correcta” o “satisfactoria” de la Matemática. Y estos criterios difícilmente podrán ser criterios definidos sin echar mano a modelos sociales y culturales referidos a los conocimientos matemáticos –literarios, lingüísticos, filosóficos, musicales, etc.–, sin imaginar cómo ha de ser una integración del individuo a determinada cultura matemática –o del tipo que sea–, que resultan imposibles de ser pensadas desde el interior mismo del mero acto educativo.²⁰

²⁰ Al respecto conviene recordar el importante texto de **Mialaret** en el que se define precisamente el tránsito del “hacer pedagógico”, característico de toda pedagogía, hacia “...la búsqueda de una solución mediante la utilización de la experimentación científica”. (Véase **Mialaret, G et Debesse, M:** *Traité des Sciences Pédagogiques*, P.U.F. Paris, 1969). Se trata, entonces, de dar paso a una aproximación técnica de la práctica pedagógica mediante la medición, la

Por ello, las primeras son las preguntas de la investigación y de la evaluación pedagógica y, como lo señalásemos anteriormente, son preguntas “internas” al proceso de enseñanza-aprendizaje, “internas” a la relación docente-alumno y que se plantean desde “dentro” del hecho educativo y en la perspectiva de contestarse también desde “dentro” de la **pedagogía**.²¹ En ese sentido es que la investigación pedagógica se plantea, en el fondo, como una investigación cerrada sobre sí misma, en donde el pedagogo construye sus preguntas a partir de su práctica y sale en búsqueda de las respuestas con la esperanza de encontrarlas en esa práctica misma.

En el segundo camino esbozado en el ejemplo, necesitamos una perspectiva de investigación y evaluación concebida desde las “ciencias de la educación”. Allí de lo que se trata es, precisamente, de analizar las prácticas pedagógicas desde “**fuera**”²² de la pedagogía y, por lo tanto, de plantear las preguntas que ponen en cuestión parte –o el conjunto– de las prácticas pedagógicas que se desarrollan en el seno de un

experimentación, la evaluación de los procedimientos del proceso de enseñanza-aprendizaje trabajando sobre muestras, cuantificaciones y mecanismos estadísticos –o, a veces, más cualitativos– que permitan controlar con relativa precisión los efectos de la actividad educativa.

²¹ En el mencionado texto de **Ellen Condliffe Lagemann**, se señala atinadamente que, a partir de fines del siglo XIX, los psicólogos se inclinaron sobre los problemas de la pedagogía en la intención de aportar respuestas “externas” al universo estrictamente educativo. Los resultados fueron dudosos y ello se refleja en el título que la autora le otorga al capítulo sobre los aportes de la psicología: “**Reluctant Allies: Psychologists turn to Education**”, Op. cit. pp. 23 a 40.

²² Este desde “fuera” significa varias cosas: desde “fuera” del proceso de enseñanza-aprendizaje, desde “fuera” de las rutinas institucionales del sistema y, sobretodo, desde “fuera” de la reflexión estrictamente pedagógica lo que implica la puesta al servicio de la investigación y de la evaluación educativas de la filosofía, de la psicología, de la antropología, de la sociología, etc.

sistema educativo. En nuestro caso particular, las preguntas que nos permitan definir qué entenderemos por una integración adecuada del educando a una cultura matemática determinada por una serie de variables que son, como vimos, en el fondo, estrictamente extra-pedagógicas.

Con un ejemplo clásico, particularmente gráfico, que se inscribe en la perspectiva de la psicología educativa y revela muy claramente la radical novedad de la investigación concebida desde las “ciencias de la educación”, Jean Claude Filloux²³ trae a colación el conocido trabajo de R. Rosenthal y L. Jacobson²⁴, donde se demuestra el efecto causado por el hecho de que las expectativas del docente están ligadas a las representaciones que éste se hace sobre los niveles intelectuales de los alumnos. Este efecto –el conocido “efecto Pygmalion”– repercute directamente sobre los resultados obtenidos por los alumnos independientemente de cualquier otra variable. Una investigación de este tipo no nació de un problema estrictamente pedagógico, de un trabajo anclado en la práctica pedagógica: en realidad, presumiblemente ningún pedagogo habría concebido su realización.

Esta investigación dejó en claro, en su momento, que la acción pedagógica es portadora de una multiplicidad de fenómenos que sólo pueden ser aprehendidos desde una perspectiva de investigación y evaluación más amplia que la que nos ofrece la pedagogía puesto que pone en juego elementos psicológicos, sociológicos, económicos, culturales, etc. que desbordan inexorablemente el campo de toda práctica pedagógica.

²³ Filloux, Jean-Claude: *Epistémologie, Ethique et Sciences de l'Education*; Ed. L'Harmattan, Paris-Budapest-Torino, 2001; pp. 14-15.

²⁴ Rosenthal, R y Jacobson, L.: *Pygmalion à l'école*; Ed. Castermann, 1970.

La idea de la investigación y la evaluación vinculadas a las “ciencias de la educación” supone entonces una perspectiva que, en el fondo, resulta todavía novedosa en América Latina –y por ello, a veces, un tanto revulsiva del discurso institucional de nuestros sistemas educativos–: abordar tanto el análisis de los procesos de enseñanza-aprendizaje como el estudio de las grandes tendencias de desarrollo de nuestros sistemas educativos *desde* las ciencias de la educación, que no es otra cosa que intentar organizar una mirada específica, desde la filosofía y desde las ciencias humanas más diversas, sobre el fenómeno educativo en su conjunto.

IV. A modo de conclusión

Resulta evidente que la temática que nos hemos planteado en esta instancia de trabajo sobre *“La dimensión política de la evaluación de la calidad educativa en América Latina”* plantea una serie de dificultades que deben ser tratadas, al menos en nuestra opinión, en diferentes niveles de análisis ya que, en sentido estricto, pertenecen a distintas temáticas. Concluamos sintetizando al menos dos de estas temáticas.

Por un lado advertimos lo que podríamos tipificar como una “dificultad de comunicación” que tiene que ver, fundamentalmente, con la irrupción de un nuevo lenguaje técnico en el ámbito de la educación que no resulta familiar dentro del tejido comunicacional que toda sociedad democrática posee y debe cultivar. Este nuevo lenguaje técnico referido a la educación es poco conocido por los medios, es casi desconocido por la opinión pública y aún no está claramente integrado dentro de la terminología que utilizan los partidos políticos. En el terreno de las evaluaciones de los resultados

educativos esta dificultad de comunicación es particularmente notable ya que la mayor parte de los agentes sociales que reciben dicha información no poseen los parámetros interpretativos adecuados para poder evaluar con precisión qué significan exactamente esos resultados.

Como problema de comunicación su persistencia no debería ser más que un problema de tiempo. Si, como vimos, el lenguaje técnico de la teoría del desarrollo recién se hizo familiar en América Latina hacia mediados de los años 70 y el de la macroeconomía y las finanzas internacionales recién encontró su traducción mediática hace solo una década en nuestra región, no hay por qué pensar que, en su debido momento, el mismo proceso de vulgarización no ha de cumplirse con el lenguaje técnico relativo a la educación. Si en el día de hoy la opinión pública medianamente informada de nuestros países puede razonablemente dimensionar lo que significa una tasa de mortalidad infantil alta o baja, una hiperinflación o un proceso de recesión económica, ello se debe a que hubo un paulatino proceso de aprendizaje y de democratización de ese tipo de informaciones. Cabe consecuentemente suponer que el mismo proceso habrá de cumplirse, más o menos rápidamente, con el lenguaje técnico relativo a la educación. Por lento que resulte este proceso, en la actualidad resulta impensable que los sistemas de educación pública del continente renuncien a hacer uso de las herramientas de la evaluación de resultados educativos, y a la utilización de éstos para orientar, en lo posible, la toma de decisiones.

Por otro lado, y esto nos remite a un cambio conceptual más trascendente, es necesario advertir que la irrupción de este nuevo lenguaje está estrechamente vinculada a la consolidación de una nueva “mirada” o perspectiva para posicionarse ante los sistemas educativos y a la educación en

general. Su presencia cada vez más sistemática en el análisis de los procesos educativos tiene que ver con la consolidación de lo que se ha dado en llamar las “ciencias de la educación”.

Independientemente de lo que éstas puedan efectivamente significar como campo específico del conocimiento “científico”, no es menos cierto que su principal mérito consiste en haber permitido la irrupción de una óptica externa que contempla los procesos educativos “desde fuera” del espacio tradicional del proceso de enseñanza-aprendizaje y “más allá” de lo que fue el discurso tradicional de la pedagogía.

Las ciencias de la educación convocaron a la filosofía, a la psicología, a la sociología, a la antropología cultural, a la estadística, etc., a dar su opinión en un ámbito en el que la pedagogía reinaba como único discurso, a la vez “legítimo” y “legitimador”. Hoy las evaluaciones concebidas desde una perspectiva exclusivamente pedagógica conservan toda su validez como herramientas de orientación del proceso de enseñanza-aprendizaje, pero no resultan suficientes como instrumentos que permitan iluminar el proceso de orientación y gestión de las políticas que llevan adelante gigantescos sistemas educativos que involucran a porcentajes muy altos de la población de nuestros países. Por dificultosas que resulten de transmitir a la opinión pública y por problemáticas que, en determinadas circunstancias, puedan tornarse para los responsables de la gestión de las políticas educativas, las nuevas evaluaciones sobre la calidad de la educación han llegado para quedarse. De su desarrollo y de su consolidación depende, en gran medida, que los sistemas educativos de América Latina logren avanzar en el camino de la profesionalización que aún les queda por recorrer.

V. Bibliografía

Bobbio, Norberto: *El futuro de la democracia*, Ed. Fondo de Cultura Económica, México, 1984.

Dunn, John et al.: *Democracy: The unfinished journey*, Ed. Oxford University Press, 1992.

Eccleshall, Robert et. al.: *Ideologías políticas*, Ed. Tecnos, Madrid, 1993.

Dictionnaire de Pédagogie, **Arenilla, Louis; Gossot, Bernard; Rolland, Marie-Claire et Roussel, Marie-Pierre**, Ed. Bordas-Pédagogie, Paris, 2000.

Condliffe Lagemann, Ellen: *An Elusive Science; The troubling History of Education Research*; University of Chicago Press, 2002.

Mialaret, G. et Debesse, M.: *Traité des Sciences Pédagogiques*, Presses Universitaires de France, Paris, 1969.

Filloux, Jean-Claude: *Epistémologie, Ethique et Sciences de l'Éducation*; Ed. L'Harmattan, Paris-Budapest-Torino, 2001.

Rosenthal, R. y Jacobson, L.: *Pygmalion à l'école*; Ed. Castermann, 1970.

Tedesco, Juan Carlos, Morduchowicz, Roxana: *El papel de los medios de comunicación en la formación de la opinión pública en Educación*, Ed. IPE - UNESCO, Buenos Aires. 2003.

Límites de la lectura periodística de resultados educacionales

José Joaquín Brunner

|

A comienzos del mes de julio del año 2003 se dieron a conocer, también en Chile, los resultados de la Prueba PISA (Programa Internacional de Evaluación de Estudiantes) administrada a alumnos de 15 años de 43 países. En el caso de Chile, los titulares de prensa que informaban sobre dichos resultados fueron lapidarios:

- *“Chile obtiene bajo resultado en educación”*
- *“Estudio internacional muestra graves fallas en comprensión de lectura de escolares chilenos”*
- *“Preocupa mala nota de Chile en educación”*
- *“El nuevo analfabetismo”*
- *“Remezón por mala nota de Chile en la prueba educacional PISA”*
- *“Descalabro en rendimiento escolar: soluciones al «hoyo negro» en educación chilena”.*

Sin excepción, entonces, la noticia recibió el tratamiento de un hecho dramático, adverso, calamitoso, catastrófico. Lo interesante es que la catástrofe, así entendida, viene anunciándose, con previsible regularidad, cada año, ya bien a propósito de las mediciones nacionales de la calidad educacional o

de los exámenes internacionales en que Chile ha participado: el Internacional Adult Literacy Test, el Third Internacional Mathematic and Science Study (TIMSS-R), el Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados para Alumnos del Tercer y Cuarto Grado de la Educación Básica de la UNESCO y el Program for International Student Assessment (PISA). Se trata, por tanto, desde el punto de vista periodístico, de una rutina que, a esta altura, se encuentra bien establecida.

A su turno, la lectura catastrofista da paso a un escenario donde, habitualmente, la prensa destaca tres “hechos”. Primero, que a nivel mundial Chile se hallaría ubicado en el furgón de cola de la educación; segundo, que existiría una grave crisis en la educación chilena; tercero, que ésta afectaría particularmente a la educación pública.

Así titulado y definido el escenario educativo nacional, la prensa se encarga a continuación –en los días siguientes a la publicación de la noticia– de interpretar las causas del mal diagnosticado.

En primer lugar, la prensa aprovecha la propia publicación de la noticia para incluir elementos interpretativos, ya sea en el párrafo que sigue al título o en el cuerpo de aquella. Un buen ejemplo de subtítulo interpretativo es éste: *“Especialistas dicen que es la hora de revertir las políticas aplicadas hasta ahora, enfatizando el trabajo en la sala de clases”*. Un ejemplo llamativo de desarrollo interpretativo de la noticia en su cuerpo es el siguiente: *“Por lo menos las autoridades reconocen que la educación chilena, junto a otras de América Latina, está en pésimas condiciones de competencia académica, lo que se ratifica con la prueba PISA. Para Chile, la situación es más patética debido a la triplicación de los recursos a partir de 1990, lo que hoy implica más del 7% del*

PIB para este sector social, en el cual el Estado coloca dos billones de pesos. Pero hay nulo avance en calidad en 12 años de esfuerzo gubernamental y 8 de reforma educacional’.

En segundo lugar, la prensa editorializa sobre los resultados y ofrece, por esta vía, su propio análisis de la situación. Por ejemplo, bajo el encabezamiento “*Educación de nivel latinoamericano*”, un diario de la capital señala: “*La situación de nuestra economía y la solidez de la democracia chilena nos han llevado muchas veces a mirar con distancia a América Latina. Y aunque hay razones para sentir orgullo, éste, a veces, raya en la soberbia. El estudio PISA («Programme for International Student Assessment»), recientemente divulgado, nos devuelve bruscamente a la realidad. Él evalúa las habilidades lectoras de los estudiantes de 15 años en 43 países, cinco de ellos latinoamericanos. Chile tiene un logro que equivale a 410 puntos, esto es, 90 por debajo del promedio general del informe. Su rendimiento general, estadísticamente, no difiere del que alcanzan México, Argentina y Brasil; estos cuatro países superan a Perú, el quinto latinoamericano, que, a su vez, obtiene el rendimiento más deficiente de todos los países participantes’.*

En tercer lugar, la prensa entrevista a expertos cuyas opiniones deben interpretar los negativos resultados. Regularmente, el Ministro de Educación es entrevistado para que reaccione frente a la noticia. Inevitablemente, su respuesta adopta la forma de una “reacción defensiva”, como cuando frente a los resultados del PISA la autoridad chilena expresa: “*Entre los países desarrollados nos sacamos un 3; entre los países latinoamericanos, un 6’.* Por su lado, los expertos tienden, ya a confirmar la construcción negativa de la noticia ofrecida por la prensa o, por el contrario, la suavizan y explican. En el primer caso, por ejemplo, se halla la experta que

tras señalar “*que en Chile la calidad de la educación no ha obtenido los resultados esperados*”, luego agrega “*que es posible concluir que medidas como la jornada escolar completa obligatoria, cambios curriculares, mejoramiento de las condiciones de los profesores y otros programas centralizados que han llevado a un gasto de más de 2 billones de pesos no están teniendo los resultados que se esperaban*”. En el otro caso se halla el experto cuya opinión es que “*falta poco para revertir los resultados de la prueba PISA*”, argumentando que “*la experiencia en 66 escuelas críticas en 2001 demostró que, cuando se aplican métodos adecuados de enseñanza, los resultados cambian rápidamente*”.

En suma, el tratamiento periodístico de los resultados de las pruebas de logro escolar apenas alcanza a revelar que nos encontramos frente a un ámbito saturado de interpretaciones contradictorias, donde se enfrentan dialécticamente distintas visiones de la educación, de las políticas educacionales y de sus resultados. Tampoco dicho tratamiento se preocupa de transparentar que, en las interpretaciones ofrecidas, se conjugan elementos de realidad, preferencias ideológicas y las agendas de los diversos actores que acceden, por medio de la prensa, a la esfera pública. En fin, un ingenuo pacto implícito de objetivismo hace aparecer como que en realidad las cosas son como se presentan en la noticia o en las interpretaciones congregadas en torno a ésta.

La situación no es distinta en otras latitudes del mundo, con excepción de aquellos pocos países cuyos resultados en las pruebas internacionales son excepcionalmente altos y, por ende, aparecen habitualmente en los primeros lugares del ranking **mundial**.²⁵

²⁵ Para una visión global de reacciones de prensa a propósito de pruebas na-

Quizá de modo inevitable, entonces, la prensa, impulsada por su propia visión de mundo como sucesión de acontecimientos o eventos noticiables y su lógica adversaria, se vea llevada a producir un tratamiento simplista y sesgado de hechos complejos como son los resultados de las pruebas que miden el desempeño de los sistemas educacionales.

¿Cuáles son las reglas que determinan dicho tratamiento? Sin pretender un grado de exhaustividad, pueden señalarse al menos las siguientes tres reglas que, en conjunto, parecen explicar una alta proporción de los efectos anteriormente descritos.

1. *La regla de la reducción llamativa.* Todo acontecimiento complejo debe ser reducido a su mínima expresión y convertido en un signo de exclamación que capte el interés del lector. En el caso que nos preocupa, como han señalado Ferrer y Arregui (2003): “sólo importa la posición relativa de logro en el escenario internacional y no la abundante información sobre factores sociales, escolares y personales asociados al rendimiento [...] La prensa, particularmente, tiende a desconocer los aspectos más complejos y auténticamente informativos de los resultados y sólo destaca las posiciones en el ranking, a menudo con títulos dramáticos y escandalosos”. No sólo los titulares de la noticia, sino su tratamiento completo, se subordinan a esta regla. Conforme a ella, el valor comunicacional de la noticia debe encontrarse en los síntomas del hecho reportado; no en sus causas, declaradas siempre “demasiado intrincadas” como para ser incluidas o como para despertar el interés del público.

cionales e internacionales, ver Grupo de Trabajo sobre Estándares y Evaluación del PREAL y GRADE, sección “Recortes: Las evaluaciones en la prensa” en <http://www.grade.org.pe/gtee-preal/gteepreal.htm>

2. *La regla de la descontextualización facilitadora.* Cualquier hecho social, para que se lo pueda tratar como noticia, debe ser arrancado de la enmarañada red dentro del cual se produce. Luego, no sólo se reducen los resultados a meras posiciones de ranking sino que, simultáneamente, se los aísla, cortándolos de los factores de contexto que los generan y que, por lo mismo, deben ser parte de su explicación. En efecto, para partir por lo más simple, no da lo mismo comparar países con distinto nivel de desarrollo; digamos, Finlandia, con un ingreso per capita de 24 mil dólares, y Chile, con uno de 4,6 mil dólares. Para la educación este es un antecedente fundamental. Así, por ejemplo, un 60% de la variación de los puntajes promedio obtenidos por los países participantes en PISA se puede predecir a partir de su ingreso por habitante. Asimismo, debe considerarse el grado de desigualdad existente en cada una de las sociedades. Las diferencias socio-familiares entre hogares son, efectivamente, el mayor obstáculo para una educación con resultados homogéneos de calidad. Como señala el propio informe PISA, un entorno socio-familiar desfavorable “parece ser uno de los factores más poderosos que influyen en el desempeño de las escalas de aptitudes para la lectura, matemática y ciencias”. Los niveles de desigualdad medidos por el Coeficiente de Gini explican un 26% de la variación en el desempeño promedio de los países. En seguida debe considerarse el gasto por alumno en que incurren los países, que es un reflejo de su nivel de desarrollo. Efectivamente, el gasto educacional acumulado hasta los 15 años explica un 54% de la variación entre países medida en su desempeño medio. En América Latina, el gasto así medido apenas alcanza a un tercio o menos del gasto promedio de los países de la OECD.

3. *La regla de la homogeneidad cultural.* Los resultados de una actividad social cualquiera serían independientes de la cultura en que se producen. En otras palabras: ¡la cultura no importa! En este caso específico, como bien señalan Ferrer y Arregui en el artículo antes citado, la prensa –al efectuar sus análisis e interpretaciones–, no considera “las grandes diferencias culturales entre los alumnos de diferentes países”; ni al interior de una misma sociedad, según estratos socioculturales (que un reciente estudio afirma explicarían alrededor de un 50% de la varianza en los resultados escolares que se manifiesta entre escuelas en Chile²⁶), ni entre sociedades con diferentes tradiciones y constelaciones culturales, como pueden ser las culturas de los países del sudeste asiático, de los países nórdico-europeos o de Europa Central y del Este previamente sujetos a la dominación de la U.R.S.S. Contrariamente pues a todo lo que sostienen las ciencias sociales contemporáneas –esto es, de que la cultura sí **importa**²⁷–, paradójicamente en el ámbito de la educación y el aprendizaje, donde más se podría esperar que ella incida, su influencia es borrada de un plumazo, para dar paso a una suerte de universalidad global y abstracta de los efectos y los procesos escolares.

Bajo estas reglas, no debiera llamar la atención que la prensa produzca el tipo de noticias e interpretaciones sobre resultados escolares caracterizado al inicio de esta presentación, sobre todo en el caso de pruebas internacionales. Tales noticias e interpretaciones tenderán por necesidad a ser descontextualizadas, a reducir por tanto el peso de la sociedad y sus desigualdades; a apuntar sólo a los síntomas, reduciendo la complejidad del logro escolar a una tabla de ganadores

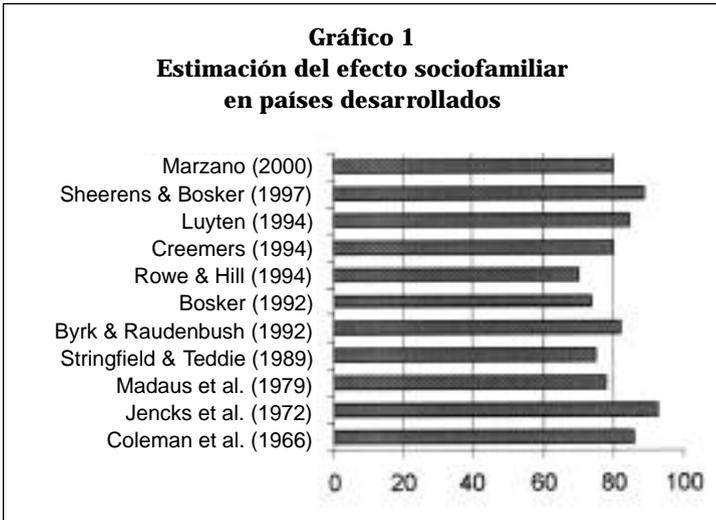
²⁶ Ver Cassasus (2003).

²⁷ Ver Harrison, L. y S. Huntington (2000).

y perdedores, en una aritmética social de escaso valor; y a presentar la carrera de rendimientos entre los alumnos del mundo (aunque nunca el universo real de comparación incluya a más de un 20% de los países y territorios del mundo) como si dicha competencia se llevara a cabo sobre una tabla rasa, donde todos compiten en igualdad de condiciones y sin que intervengan factores socioeconómicos y culturales de ninguna especie.

II

Nada se halla más alejado de la realidad, sin embargo, como mostraremos en la segunda parte de este artículo. En efecto, durante los últimos 40 años los investigadores educacionales alrededor del mundo han producido grandes avances en la tarea de desentrañar los factores que explican los resultados escolares, medidos por el logro de aprendizaje de sus alumnos. A esta altura existe un amplio consenso entre los especialistas. Los factores explicativos son dos: el entorno sociofamiliar de los alumnos y la efectividad de la escuela. ¿Cuánto pesan, respectivamente, estos factores en la varianza de los resultados observados entre establecimientos? En el caso de los países desarrollados, los estudios atribuyen un peso de 80% y de 20%, respectivamente, a la familia y la escuela (Gráfico 1). En América Latina, los pesos asignados son 60% y 40%, respectivamente (Cohen 2002, Banco Mundial 1995; Gerstenfeld 1995), subrayándose con ello el hecho de que allí donde hay más pobreza la escuela adquiere, asimismo, mayor importancia.



Suele pensarse que el factor socio-familiar estaría determinado exclusiva o principalmente por el nivel de ingresos del hogar. En verdad, este factor es mucho más complejo que eso. Tiene que ver con el “mundo de vida” en que nace y se desarrolla el niño. Ciertamente, el status socioeconómico de los padres es importante. Pero aún más decisivo parece ser, en el caso de niños provenientes de hogares de escasos recursos, la organización de la familia, su clima afectivo, la socialización lingüística o la adquisición temprana de actitudes y motivaciones, como muestra el siguiente Cuadro.

Variables de entorno familiar que inciden en los logros de aprendizaje

Ocupación, ingreso y nivel educacional de los padres.
Infraestructura física del hogar y grado de hacinamiento.
Recursos del hogar (libros, diccionarios, escritorio, computadora).
Organización familiar y clima afectivo del hogar.
Capital social de la familia, vecindario, comunidad y pares.
Alimentación y salud durante los primeros años de vida del niño.
Prácticas de socialización temprana.
Desarrollo lingüístico y régimen de conversaciones en el hogar.
Rutinas diarias, desarrollo de actitudes y motivación.
Acceso a, y calidad de, la enseñanza preescolar.
Elección de escuela.
Armonía entre códigos culturales de la familia y la escuela.
Estrategias de aprendizaje y conocimiento previo adquiridos.
Comunicación familia-escuela .
Involucramiento familiar en las tareas escolares.
Uso del tiempo en el hogar y durante las vacaciones.

Levin y Belfield (2002) Marzano (2000) Sheerens (2000).

Y en el caso de las escuelas, ¿cuáles insumos afectan el rendimiento de los alumnos? ¿Hay algunos que importan más que otros? Una revisión efectuada por Fuller y Clarke (1994), sobre la base de más de 300 estudios realizados sobre sistemas escolares de países en desarrollo, permite apreciar qué porcentaje de dichos estudios muestran una asociación positiva entre el respectivo insumo y los resultados obtenidos por

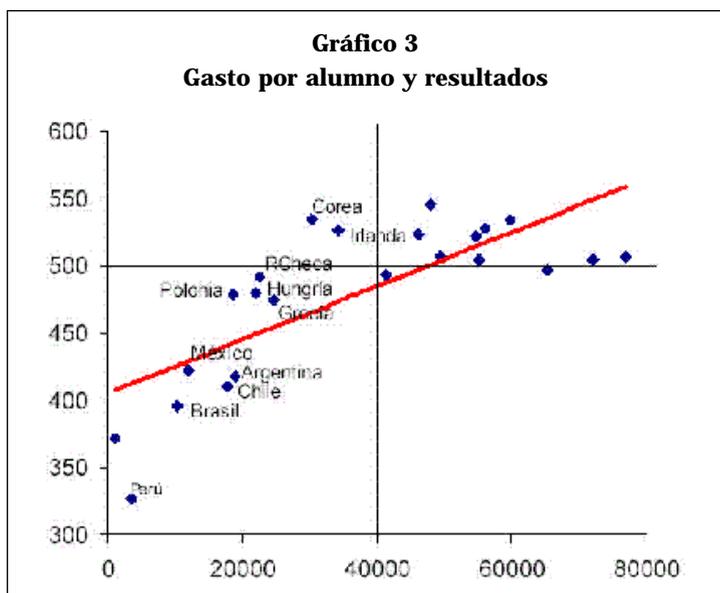
los alumnos, una vez que se controla por su nivel socioeconómico (Gráfico 2).



Sin embargo, si se adopta una perspectiva metodológica más restringida y exigente, la efectividad de la escuela aparece determinada, principalmente, por la efectividad del profesor en la sala de clase. En general, se estima que el efecto escuela depende, en alrededor de 2/3 partes, de la calidad de la docencia impartida. A su turno, ésta se halla determinada, en orden decreciente de importancia, por las siguientes variables (Wenglinsky 2002): las prácticas de enseñanza del docente en la sala de clase; el desarrollo profesional docente (dominio de la materia que enseña y capacidad de enseñar a

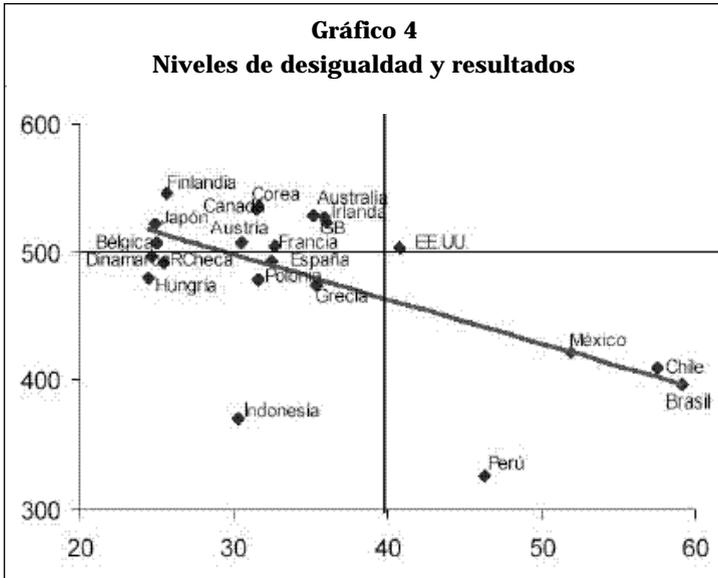
alumnos de diverso origen sociofamiliar) y los insumos (tamaño del curso, educación inicial y experiencia del profesor).

Particular importancia adquiere en el contexto de los procesos de reforma educacional el impacto que podría tener el gasto por alumno sobre los resultados o logros escolares. La evidencia internacional muestra lo siguiente (Gráfico 3): si se combina el puntaje promedio obtenido por los alumnos de varios países, participantes en PISA, con el gasto acumulado por alumno hasta los 15 años, expresado en dólares de valor equivalente, se observa en primer lugar que sí existe una asociación entre ambas variables, aunque con diferencias como permite ver la distribución de los países incluidos en torno a la línea de tendencia. En seguida, se constata que sólo unos pocos países –Corea e Irlanda aparecen allí– obtienen altos resultados (por encima de 500 puntos, que es el puntaje promedio de los países de la OECD participantes) con un gasto que en la escala se sitúa por debajo de la mitad del rango entre 0 y 80 mil dólares (PPC). Estos países, ubicados en el cuadrante noroeste del [Gráfico 3](#), se hallan naturalmente bien arriba de la línea de tendencia. Hay un segundo grupo de países que se aproxima al cuadrante anterior de gasto austero y alto rendimiento: República Checa, Polonia, Hungría y Grecia. Por otro lado, ningún país de gasto superior a 40 mil dólares se ubica en el cuadrante suroeste, por ende, de alto gasto y resultado inferior al puntaje promedio de la OECD. Por el contrario, los países de alto gasto se concentran en el cuadrante noreste, aunque algunos están por encima y otros por debajo de la línea de tendencia. Finalmente, los países de América Latina se agrupan todos en la posición suroeste –de bajo gasto y bajos resultados– aunque con diferencias entre ellos. En efecto, a pesar de estar todos por debajo de la línea de tendencia, México está en términos rela-



tivos menos mal que Argentina, ésta que Chile y éste que Brasil.

¿Existe algún elemento adicional que pudiera ayudar a explicar el diferente desempeño de los países? Si nos atenemos solamente a los factores externos o de contexto, el mejor candidato parece ser el nivel de desigualdad de las distintas sociedades. En efecto, como puede observarse en el [Gráfico 4](#), los países con mayor desigualdad –medida por el Coeficiente de Gini– tienen más dificultad para producir buenos resultados escolares, incluso si son ricos, como sucede en el caso de los EE.UU., justamente porque deben compensar un mayor déficit de origen sociofamiliar. Por el contrario, sólo obtienen buenos resultados países con un nivel bajo de



desigualdad, con la excepción de Indonesia, que apenas acumula un gasto de 1.164 dólares (calculados con PPC) en sus alumnos de 15 años, 50 veces menos que Australia, por ejemplo, o 1/15 del gasto de Chile. Es interesante observar, asimismo, que de los países de la muestra, Chile ostenta el más alto nivel de desigualdad, ubicándose en general todos los países latinoamericanos participantes en PISA en el cuadrante sureste, el de alta desigualdad con bajo rendimiento.

Si se considera el conjunto de antecedentes aquí presentados –cuyo desarrollo *in extenso* puede encontrarse en Brunner y Elacqua (2003)– emerge con nitidez la enorme complejidad del proceso de producción de resultados educacionales. En este proceso influyen diversos factores y múltiples variables. Entre aquéllos, como se señaló, los más im-

portantes son el entorno sociofamiliar de los alumnos y la escuela. Entre las variables son decisivas aquéllas que caracterizan el hogar como agencia de transmisión de capital cultural, por un lado, y las que inciden sobre la efectividad de los establecimientos escolares, por el otro. Por último, destaca el hecho de que allí donde los hogares son grandemente desiguales, la tarea educativa enfrenta también mayores problemas y desafíos. En estos casos la cultura nacional –y su elaboración “micro” a nivel de la familia, los pares y la comunidad– juegan un papel importante, como parece indicar el caso de los países del sudeste asiático (McBeath y Mortimore 2001). En efecto, la cultura nacional y sus elaboraciones “micro” pueden, ya bien contribuir a facilitar y reforzar la labor de la escuela o, al contrario, ser neutrales frente a ella u obstaculizarla.

En cuanto a las posibilidades de compensar educacionalmente las desigualdades de origen sociofamiliar de los alumnos, ellas sólo pueden materializarse allí donde la mayoría de las escuelas que atienden a niños y jóvenes provenientes de hogares en situación vulnerable son efectivas. Como muestra ampliamente la literatura especializada, sólo éstas logran mejores resultados para sus alumnos que aquellos que podrían anticiparse considerando su origen sociofamiliar (Woessman 2001, Sheerens 2000, 1992, Cotton 1995, Sammons, Hillam y Mortimore 1995, Levine y Lezotte 1990, Purkey y Smith 1983).

Ninguna de estas complejidades y dificultades forma parte, sin embargo, de los habituales análisis periodísticos sobre resultados educacionales. Como vimos, éstos se limitan a resaltar unidimensionalmente el producto final de estos procesos –es decir, resultados como puntajes– y a ubicarlos en un esquemático ranking de países o tipos de estableci-

mientos, suprimiendo toda referencia al contexto en que dichos resultados se obtienen. Con esto se banaliza la información que aportan los estudios nacionales e internacionales de medición del desempeño escolar y, lo que es más grave, se distorsiona la deliberación pública y se limitan sus efectos, debilitándose de esta forma uno de los mecanismos más poderosos que poseen las democracias para abordar y resolver problemas colectivos.

Bibliografía

Banco Mundial (1995), *Priorities and strategies for education.*; Washington, D.C.

Bosker, R. J. (1992), *The stability and consistency of school effects in primary education.* Enschede: University of Twente.

Brunner, J.J. y **G. Elacqua** (2003), *Chile: Informe sobre capital humano.* Santiago de Chile: La Araucana.

Disponible en <http://www.uai.cl>

Byrk, A.S. y **S.W. Raudenbush** (1992), *Hierarchical linear models.* New York: Sage.

Cassassus, J. (2003), *La escuela y la (des)igualdad.* Santiago de Chile: LOM.

Cohen, E. (2002), "Educación, eficiencia y equidad: Una difícil convivencia"; *Revista Iberoamericana de Educación*, N° 30.

Coleman, J. et al. (1966), *Equality of educational opportunity.* Washington, D.C.: U.S. Government Printing Office, Washington D.C.

Cotton, K. (1995), "Effective schooling practices: A research synthesis". Northwest Regional Educational Laboratory.

- Creemers, B.P.M.** (1994), *The effective classroom*. London: Cassell.
- Ferrer, J. G. y P. Arregui** (2003), "Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación"; PREAL, Documento de Trabajo N° 26, 2003 <http://www.preal.cl/publicdt2003.php>
- Fuller, B. y P. Clarke** (1994), "Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy". *Review of Educational Research*, Vol. 64, No. 1.
- Gerstenfeld, P.** (1995): "Comparación regional del impacto de las características del hogar en el logro escolar ", Santiago de Chile: CEPAL, Serie Políticas Sociales, n.º 9, (LC/L.924).
- Harrison, L. y S. Huntington** (2000), *Culture Matters. How Values Shape Human Progress*; Basic Books.
- Jencks, C. et al.** (1972), *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.
- Levin, H.M. y C.R. Belfield** (2002), "Families as Contractual Partners in Education". Teachers College, Columbia University, National Center for the Study of Privatization in Education, Occasional Paper No. 44.
Disponible en http://ncspe.org/keepout/papers/00044/965_OP44.pdf
- Levine, D.U. y L.W. Lesote** (1990), *Unusually effective schools: A review of analysis and research and practice*. Madison, Wisconsin: National Center for Effective Schools Research and Development.
- Luyten, H.** (1994), *School effects: Stability and malleability*. Enschede: University of Twente.
- Madaus, G.F. et al.** (1979), "The sensitivity of measures of school effectiveness". *Harvard Educational Review*, 49(2).
- Marzano, R.J.** (2000), "A New Era of School Reform: Going Where the Research Takes US"; Aurora, CO: McREL.
Disponible en <http://www.mcrel.org/products/school-improve/new-era.pdf>http://www.mcrel.org/products/school-improve/new_era.pdf
- McBeath, J. y P. Mortimore** (eds.) (2001), *Improving school effectiveness*. Buckingham: Open University Press.

OECD-UNESCO (2003), *Literacy skills for the world of tomorrow. Further Results from PISA 2000*. Paris: Organization for Economic Co-operation and Development.

Rowe, K.J. y P.W. Hill (1994), "Multilevel modeling in school effectiveness research: How many levels". Paper presented at the International Congress of School Effectiveness and Improvement, Melbourne, Australia.

Purkey, S. C., & Smith, M. S. (1983), "Effective Schools: A Review." *The Elementary School Journal*, 83(4), 427-452.

Sammons, P., J. Hillman, P. Mortimore (1995), *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Scheerens, J. (2000), *Improving School Effectiveness*. Paris: International Institute for Educational Planning.

Scheerens, J. (1992), *Effective Schooling: Research, Theory and Practice*. London: Cassell.

Scheerens, J. y R.J. Bosker (1997), *The foundations of educational effectiveness*. New York: Elsevier.

Stringfield, S. y C. Teddlie (1989), The three first phases of the Louisiana school effectiveness study. En B.P.M. Vreemers, T. Peters, and D. Reynolds (eds.) *School effectiveness and school improvement: Proceedings of the Second International Congress*, Rotterdam. Lisse: Swets & Zeitlinger.

Wenglinsky, H. (2002), "How Schools Matter: The Link Between Teacher Classroom Practices and Student Academic Performance"; *Education Policy Analysis Archives*, 10 (12), December Disponible en <http://epaa.asu.edu/epaa/v10n12/>

Woessmann, L. (2001), "Why Students in Some Countries Do Better". *Education Next*, 2001 / 2.

Disponible en <http://www.educationnext.org/20012/67.html>

Indicadores y sus usos en educación: una tensión no resuelta

Pablo Halpern Britz

Tensión entre transparencia y rigor del debate

La investigación en materia de políticas públicas es cada día más indispensable. La producción de indicadores en esta área es clave para una correcta toma de decisiones, para evaluar la gestión, para identificar y proyectar tendencias, para focalizar, corregir y reorientar iniciativas, para anticipar problemas y soluciones.

De la correcta orientación que provean estos indicadores depende en parte que el diseño y aplicación de políticas públicas sea adecuado y resuelva efectivamente los problemas y desafíos que se propone.

La difusión de indicadores para medir resultados de políticas públicas también provee transparencia. La autoridad asume el compromiso de difundir los indicadores de su gestión en diversas áreas, sometiéndose implícitamente al escrutinio público de su acción: acepta la evaluación de sus planes, metas y objetivos, aportando los instrumentos para ello: los indicadores.

Esto contribuye, por una parte, a que la gestión de recursos públicos comprometidos sea crecientemente eficiente y, por otra, a que otros actores –proveedores, beneficiarios,

usuarios, etc.– reciban señales correctas y oportunas que estimulen y orienten su toma de decisiones.

Sin embargo, la búsqueda de transparentar la gestión pública mediante la difusión de este tipo de indicadores conlleva una tensión implícita entre el necesario rigor técnico del análisis y debate respecto de los resultados y la inevitable dimensión política de la discusión, donde no necesariamente es el rigor técnico el que guía el debate.

Más aún, la creciente masificación de la divulgación de resultados a través de los medios de comunicación sitúa el debate en otro nivel, absolutamente legítimo por lo demás, que dice de una relación con la interpretación de resultados desde el punto de vista de los intereses de la opinión pública.

En el caso de indicadores del área educación como los resultados de la prueba SIMCE, la integración de la audiencia medial al debate necesariamente ubica a éste en un plano que no es necesariamente aquél para el que está diseñado el SIMCE. El debate medial se centra en preguntas acerca del mejoramiento o no de la educación, la superación o manutención de niveles de inequidad en el sistema, la calificación de los maestros, etc.

Una tensión canalizada

Hay ámbitos del quehacer público donde la tensión antes mencionada está prácticamente resuelta. En el plano económico, por ejemplo, ésta se ha reducido paulatinamente, el debate público de los indicadores económicos se produce en un contexto técnico adecuado.

Las interpretaciones de las cifras del IPC, IMACEC o empleo se enmarcan en un nivel adecuado de rigor técnico.

Existen, en este sentido, ciertos consensos básicos mayoritarios entre quienes participan en el debate sobre la calidad técnica y los significados específicos de cada uno de los indicadores.

Podrá haber discrepancias respecto de qué influye en uno u otro resultado, pero prácticamente nadie utilizará un indicador económico para indicar algo que éste no mide o refleja, o cuestionará la calidad de los mismos.

Otro factor que contribuye a que en esta esfera predomine el rigor técnico del debate, reduciéndose con ello la tensión con la interpretación medial de los datos, habla de una relación con la calidad de los participantes en el debate público. Estos son, en general, académicos, economistas de prestigio o políticos legitimados en este ámbito.

Por otra parte, la profusión de indicadores en este ámbito y la periodicidad de su entrega enriquece el debate permitiendo análisis e interpretaciones de mayor complejidad y alcance.

Una situación precaria

No ocurre lo mismo en otras esferas del quehacer público, donde la situación es claramente más precaria. Áreas como educación, vivienda y salud han producido históricamente indicadores que muestran la situación de cada uno de estos sectores:

- En salud, están los tradicionales indicadores de mortalidad infantil, del estado nutricional de la población, de la evolución (reducción) de determinados tipos de enfermedades.
- En educación están los índices de alfabetización de la po-

blación, los indicadores de escolaridad, de matrícula universitaria, etc.

- En vivienda, en tanto, se dispone de indicadores de déficit habitacional (decreciente), de dotación de servicios de las viviendas, de pavimentación, etc.

Sin embargo, a medida que en estas áreas se avanza en la resolución de los problemas básicos (acceso), cambia el foco del objeto de medición. En estas esferas adquiere relevancia creciente medir la calidad de los servicios, la eficiencia de la gestión, la satisfacción de los beneficiarios. Ahí, la situación es claramente más precaria.

Este tipo de indicadores son de data más reciente, con metodologías menos conocidas y, lo que es más importante, menos consensuadas; por ejemplo:

- En salud se mide la evolución de la cobertura (número de atenciones) en relación al incremento de recursos comprometidos en la gestión, pero no se considera necesariamente la evolución en los niveles de calidad de la atención ni en los niveles de complejidad de la misma.
- En vivienda se mide el incremento en la cobertura, pero está pendiente la medición de los niveles de calidad de las viviendas sociales, por ejemplo.

Algunos ejemplos de esta precariedad y los efectos que genera son evidentes:

- En salud, por ejemplo:
 - Los primarios intentos por establecer alguna medición de eficiencia resultaron rápidamente abortados tras su di-

fusión en los medios. La publicación de un índice de eficiencia (Informe de Caldera) en los servicios de salud en 1996 trajo serios conflictos entre el entonces ministro Massad y el Colegio Médico.

– Este año un informe elaborado por la Contraloría sobre el uso de recursos humanos y materiales del servicio público de salud para fines particulares terminó con el Ministro del ramo pidiendo excusas a los médicos.

En otros ámbitos, la situación es aún más compleja. En seguridad ciudadana, por ejemplo, un tema de altísimo interés público, el único indicador disponible son las tasas de denuncias de delitos y de detenidos.

Un instrumento que adolece de deficiencias conocidas, que están en pleno proceso de perfeccionamiento –se avanza en la creación de un indicador de victimización, se busca crecientemente homogeneizar los datos de las policías, etc.–, en el debate público de los datos ha terminado siendo la vara con que se mide la gestión gubernamental en esta materia, en un ambiente altamente politizado:

- Los indicadores de denuncias, por ejemplo, son frecuentemente homologados por los medios con número de delitos, cuando la relación no es necesariamente directa. Y las cifras sirven para debatir sobre eventuales éxitos o fracasos en el control de la delincuencia y exigir enmiendas en la política de seguridad ciudadana, así como medidas específicas para enfrentar el delito en nuestra sociedad.

La demanda sobre los indicadores en educación

La reforma educacional ha sido uno de los procesos más complejos, de mayor visibilidad y envergadura en los 90. El énfasis puesto por los gobiernos de la Concertación en esta área, así como la sensibilidad pública del tema (la educación sigue siendo vista por la población como el principal medio de ascenso social) han hecho de la reforma objeto de preocupación y permanente escrutinio.

Ello genera una fuerte demanda por contar con indicadores que den cuenta de los avances o eventuales retrocesos de la reforma en marcha. Así, diferentes pruebas de distinta procedencia y nivel (Simce, Pisa [OCDE], TIMS, PAA) sirven en los medios de comunicación para *evaluar* la reforma.

Nuevamente, datos que tienen claramente otra finalidad se usan como parámetros para medir un objeto distinto:

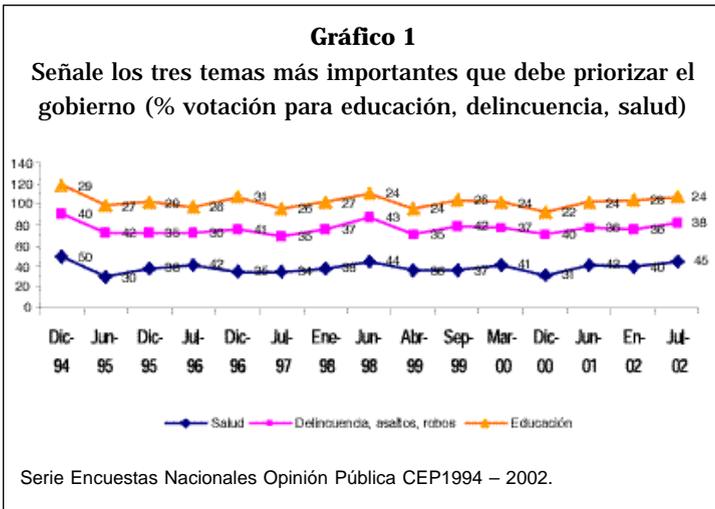
- La PAA, destinada a seleccionar alumnos para la educación superior, se usa como instrumento para medir variaciones en la equidad del sistema educativo.
- Pruebas que comparan niveles de aprendizaje y rendimiento entre distintos países, son utilizadas indirectamente para evaluar los resultados de la reforma.

Las dimensiones del debate en educación

En el caso de la educación se verifica una fuerte tensión en los alcances de los debates técnicos y mediales o públicos. Las discusiones que se originan en torno a los indicadores en esta área se proyectan en tres dimensiones:

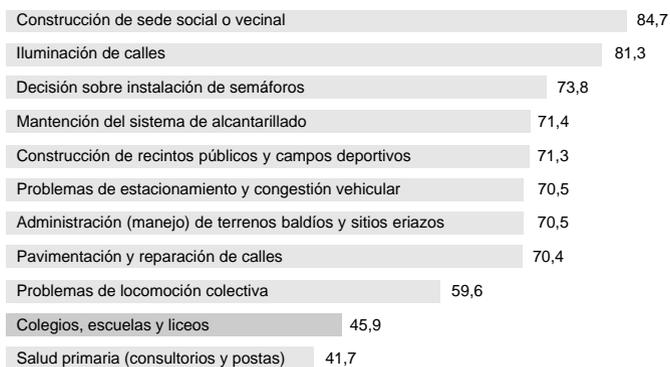
- Política
- Ideológica cultural
- Ciudadana

Desde el punto de vista político, es natural que la oposición analice políticamente estos indicadores para evaluar la gestión concertacionista, en un área de sostenido interés ciudadano –pocas áreas como educación tienen incidencia en la cotidianidad de la mayoría de las familias chilenas (gráfico 1)–, que supone puede tener alta rentabilidad política. Además, un área que los propios gobiernos de la Concertación han puesto en el eje de las miradas (siempre entre las primeras prioridades) y donde los indicadores –aunque no hayan sido construidos para ello– mostrarían resultados deficientes.



Desde otro punto de vista, indicadores como la prueba SIMCE pueden potencialmente ser objeto de nuevos debates políticos. Por ejemplo, en la medida que los municipios también son responsables de la gestión educacional y son crecientemente evaluados al respecto (**gráfico 2**), esta evaluación puede alcanzar también a la gestión de la derecha en esta materia, en la medida que controla actualmente el 42% de los municipios del país.

Gráfico 2



Centro de Estudios Públicos.

Desde otro punto de vista, el tema de la educación se ha transformado en un campo donde emerge una renovada discusión ideológico-cultural; recientes debates así lo demuestran:

- Interpretación respecto de lo que es la libertad de enseñanza v/s los alcances de la gestión estatal para me-

jorar la calidad de educación.

- El rol del Estado en áreas que se sitúan en el límite de lo público y lo privado; por ejemplo, educación sexual (oposición a proyectos como JOCAS y otros).

El intenso debate en torno al SIES dejó entrever diferencias ideológicas no siempre reconocidas como tales:

- La férrea defensa de la libertad de enseñanza supuestamente amenazada por esta prueba: *El Sies es un atentado contra la libertad de enseñanza...va contra lo propuesto por la Reforma Educacional*, declaraba el director del CEP, Arturo Fontaine.
- Las diferencias en los rendimientos en PAA y específicas entre colegios municipalizados y particulares, son finalmente atribuidos por los detractores de la prueba SIES a la libertad curricular, relativizando el peso que tiene el contexto socioeconómico.

Por último, desde la mirada ciudadana, el valor de estos indicadores se traduce en la posibilidad de contar con un instrumento para orientar sus decisiones en materia de educación de sus hijos. Por ejemplo, los resultados de la prueba SIMCE son usados para decidir qué colegio elegir (o evitar).

La traducción de la prensa muchas veces se orienta en tal sentido:

- *80% de alumnos de mejores colegios del Simce ingresa a la universidad.* (La Tercera 28/09/02)

Sin embargo, muchos de estos indicadores no están diseñados para este tipo de demandas ciudadanas. El SIMCE no es-

tá concebido para evaluar cada colegio en particular, sino el sistema en su conjunto. De acuerdo a lo planteado por Sergio Martinic, director del CIDE (La Tercera, 22 de julio de 2002), *el Simce o los sistemas internacionales de medición como el Timss y Pisa informan de los aprendizajes que tienen los estudiantes en un año escolar determinado en función de estándares o de logros esperados. Estas pruebas son muy útiles para dar una visión macro del país o de un conjunto de países, pero no son suficientes para buscar soluciones y definir líneas de trabajo en escalas intermedias y locales, como lo requieren los municipios y las escuelas.*

Aun así es usado para diseñar rankings de calidad de los colegios, lo demuestra la lectura que de este indicador hace la prensa:

- *Simce: mejores colegios de Chile y resultados del gran Santiago.* (La Tercera, 27/09/02)

Una mirada más bien pesimista sobre esta tensión entre debates técnico y público en educación, señala que la situación se hará crecientemente compleja. La masificación de los indicadores impedirá circunscribir su uso al ámbito para el cual fueron diseñados: cada vez más será usado como un ranking para tomar decisiones inmediatas de la familia; cada vez tendrá más sentido transformarlo en base de un debate político.

Esta desnaturalización de los alcances de los indicadores finalmente puede llevar al inmovilismo: ¿estarán dispuestas las autoridades a seguir transparentando su gestión si los instrumentos se vuelven en su contra? ¿Estarán dispuestos los medios de comunicación a no disponer de estos datos?

Una mirada más pragmática debe considerar esta ten-

sión como un hecho de la causa, y orientarse a la búsqueda de fórmulas que, sin mermar los niveles de transparencia, permitan equilibrar el debate, no sólo para que éste no se politice y termine siendo un flanco de castigo para la gestión del gobierno, sino para elevar la calidad del debate y su utilidad para el mejoramiento de la educación. Tal como ocurrió en el plano económico, es posible proyectar un mayor contrapeso entre el debate técnico y el público, reduciendo su explotación política.

Desde el punto de vista político-ideológico, es posible desplazar la evaluación desde el sistema educacional a la gestión de los colegios, que es donde el proceso se lleva a cabo.

Es necesario aumentar y diversificar los indicadores: no sólo hay que difundir resultados sobre niveles de aprendizaje; también hay que hacerlo, por ejemplo, respecto de la capacidad de retención de alumnos en el sistema educacional; sobre la relación entre infraestructura y rendimiento; etc.

Es indispensable incorporar a todos los sectores a la discusión: en la medida en que las políticas educativas generen más altos niveles de consenso entre todos los actores interesados en el debate –maestros, políticos, empresarios, técnicos, las familias, etc.–, al igual que en el ámbito económico, es posible esperar un menor grado de politización.

La evaluación como fetiche

Martín Granovsky

La evaluación educativa se convirtió en un método insuficiente, exagerado, fetichista y formalista. Pero a la vez es una herramienta utilísima para gobernantes y ciudadanos, y no hay por qué arrojarla al rincón de los trastos inservibles. Puede seguir. Debe seguir. ¿O acaso la política prescindirá, por ejemplo, de las encuestas de intención de voto, a pesar de que registren solamente fotos de lo que pasa y su instantaneidad no tenga la riqueza de una película?

La evaluación incomoda a los ministros de Educación porque se convirtió en una trampa de la cual ellos son las víctimas. Primero la sobredimensionaron y ahora no saben cómo escaparse. Y no importa si las cosas van mal o bien. La evaluación, para los ministros, representa la fatalidad de que las cosas aparecen públicamente mal aunque vayan bien. Más aún: como los ciudadanos carecen de instrumentos científicos de medición sobre la calidad de la educación –otros, claro, además de su relación concreta y familiar con el sistema educativo–, terminan esperando el momento de la difusión de resultados con una ansiedad que en realidad no experimentan realmente.

Un ejemplo son las famosas listas de los colegios mejor evaluados. Una vez por año se publican y entonces se pone en marcha el circuito informativo de los medios de comunicación.

- Los diarios informan sobre la lista.
- A la mañana siguiente, un programa de radio llama a la directora para preguntarle por qué en un pueblo perdido del interior argentino, sin recursos económicos ni asistencia pedagógica, ella y su comunidad educativa consiguieron el milagro.
- Por la tarde viaja el equipo de un canal de televisión. Esta vez la nota son los alumnos. ¿Qué tendrán ellos de raros, pobrecitos?
- Y a la noche todo terminó. El circuito devoró su alimento y sólo habrá que esperar a que el apetito vuelva a abrirse, seguramente con informaciones que no provienen de la educación.

A veces, la evaluación se refiere a la calidad de los docentes. O a cómo son los maestros cuando no se exponen en clase: qué leen, cómo viven, qué comen, cuánto ganan, qué valores sustentan, hasta qué punto superaron la discriminación, cómo están formados, qué sentido tienen la patria y la democracia para los encargados de formar a los chicos en la educación sistemática. En este caso, el circuito será distinto, porque además de la información juega con fuerza la relación entre los docentes como colectivo gremial y el Estado como patrón empleador. El circuito anterior quedará “contaminado” por una doble sensación. Los docentes dirán que la evaluación no es intelectualmente honesta porque el fin del Estado-patrón es el recorte de derechos adquiridos. El Estado argumentará que los docentes como colectivo de empleados están haciendo una utilización gremial de resultados que sólo tenían como fin saber el estado real de la educación en un determinado momento.

Si la evaluación queda limitada al primer circuito o a

constituirse en un elemento más de la puja entre los maestros y el gobierno pero todos tienen conciencia de que ahí no se acaba el mundo, no habrá problema alguno. Sin dramas extraordinarios, el show continuará, porque siempre debe seguir.

El ruido se produce cuando la evaluación cumple su papel de profecía autocumplida y asciende en el ranking de la agenda pública hasta parangonarse con la mismísima política educativa.

Esa es la fetichización tan temida. La imagen termina reemplazando al objeto y ya nadie puede saber de qué se habla cuando las cosas no eran tan difíciles.

¿Por qué la evaluación es un producto de consumo tan digerible para los medios? Muy sencillo. Hay tres razones que tienen que ver con la mañana del periodismo y con la tensión natural entre periodismo y poder político.

Razón número uno: la evaluación permite ver cómo el gobierno se mide a sí mismo.

Número dos: si hay fracaso evidente, permite fastidiar al gobierno con un arma creada por la propia administración del Estado.

Número tres: si el fracaso no es evidente, es posible interpretarlo como tal.

Pero también hay razones que no tienen que ver con las mañas y la picardía. Razones que dan la pauta de posibles alternativas al sobredimensionamiento de la evaluación educativa.

Razón número uno: las listas de colegios corporizan la educación en instituciones de carne y hueso.

Número dos: acercan la educación a la verdad de cómo es percibida, es decir, como un elemento central de la vida cotidiana.

Número tres: ofrecen una variante a la cobertura educativa tradicional del periodismo, a menudo atada a los problemas presupuestarios, el pago de los docentes y la permanencia o no del ministro de turno.

Número cuatro: brindan historias, individuales y colectivas, sugieren rarezas, plantean misterios e incógnitas.

El riesgo de la evaluación como fetiche es que, lo mismo que sucedió con la economía, termine vaciando la política. Que sea un sucedáneo educativo de los fundamentals económicos y quede cristalizada como una ideología vacía que funciona como espejo deformado de la realidad.

¿Quién podría negar la utilidad de evaluar los números macro de la economía? Incluso antes de Maastricht, pero sobre todo desde el tratado de Maastricht de 1991, por el que Europa se autoimpuso una serie de exigencias económicas, el análisis económico tiene en cuenta, entre otras variables, la relación entre deuda y producto bruto interno, tasa de inflación y déficit fiscal. Pero a partir de ese punto el problema es doble.

Por un lado, la descripción suele terminar en una forma de prescripción.

Por otro, esa descripción es forzosamente limitada. Un ejemplo es el déficit fiscal argentino. En diferentes momentos de los últimos años la Argentina hubiera podido ser miembro de la Unión Europea. Pero la Argentina no queda en Europa, y además los organismos multilaterales de crédito cuestionaron los grandes números a causa del déficit fiscal, que el economista Salvador Treber calcula a moneda constante, tomado a pesos del 2001, con una media anual de 66.657 millones de pesos para el período 1998-2000.

La cifra sola no dice nada. No tiene en cuenta, por ejemplo, el endeudamiento externo, fruto a su vez de la en-

trada de capitales en su mayoría especulativos, atraídos con altas tasas de interés y transformados muy pronto en una impagable deuda en dólares. Una visión más profunda, como la del ex vicepresidente del Banco Mundial Joseph Stiglitz, indica que los fundamentals entendidos con esa limitación no permiten comprender la realidad. En el caso argentino, según Stiglitz resulta imposible comprender el déficit fiscal sin tener en cuenta que el bache no se debe tanto al gasto público como a la privatización repentina y sin salvaguardias del sistema de jubilación de reparto o público. La discusión, naturalmente, excede el plano académico. Dado el origen distinto del déficit, una cosa es diagramar políticas pensando que el único problema es el gasto provincial, y otra diseñarlas teniendo en cuenta la inconveniente relación, que en la Argentina llega hasta la identidad absoluta, entre bancos y administradoras de fondos de pensión.

Al margen de este hecho, el rubro “Remuneraciones” se mantuvo en un 33 por ciento, mientras que el rubro “Intereses de la deuda” pasaba del 12 al 16 por ciento.

O sea: el número permite sólo sobrevolar la realidad y compararla en grueso. Es como un juego. Inútil, como todo juego, además, pero sin una pizca de lúdico. El análisis más profundo, en cambio, tiene la desgracia de los matices, que suelen ser engorrosos, pero la utilidad de la sintonía fina.

La educación es, por definición, un fenómeno de largo plazo. Es imposible mejorarla de un día para otro. Por fortuna, claro, también es imposible empeorarla de un día para otro. La evaluación de la calidad educativa, pero sobre todo su difusión, su conversión en instrumento público, se encuentran con ese dato inamovible de la realidad, pero a la vez están tironeadas por la urgencia fabricada en los últimos años en todo el mundo.

Ante este panorama negro, ¿la evaluación difundida igual que en los últimos años tiene alguna utilidad? Es evidente que sí. Resulta interesante tomar, en el caso argentino, no los resultados de la evaluación sino los títulos de las coberturas de prensa para ver qué datos se convirtieron en historia o en noticia y llegaron así a los ciudadanos.

- Los adolescentes tienen más dificultad para leer un texto periodístico que uno literario. Por ejemplo, no saben distinguir la argumentación de la noticia y carecen de los recursos para interrogar al texto.
- Los colegios más caros no son los primeros en el ranking de Lengua y Matemática.
- Entre los primeros hay más privados que estatales.
- El nivel socioeconómico de los alumnos es determinante tanto en Matemática como en Lengua. A nivel nacional, los hijos de padres con educación secundaria o terciaria obtuvieron en 2001 un 69 por ciento de respuestas correctas, mientras que los hijos de padres con escolarización primaria llegaron al 51 por ciento. A nivel de la Capital Federal, las cifras un dan 74,5 contra un 56,6 por ciento.
- Aun así, la escuela no reproduce exactamente ni en términos regionales ni en términos sociales la cristalización de la Argentina en una sociedad cada vez menos plebeya. La diferencia promedio en 2001 llegaba al 14 por ciento, en una sociedad más polarizada que lo que esa cifra puede indicar, como lo revela la existencia, en ese momento, de más de un 40 por ciento de pobres dentro del total de población. El secretario de Educación de la ciudad de Buenos Aires lo resumió así: “Entre un pobre que se recibió de abogado y un rico que también se recibió hay menos diferencia que entre un pobre que dejó de estudiar y un

rico que estudió”. Conviene recordar que en 2001, a diferencia de años anteriores, el Ministerio de Educación no puso el énfasis en los rankings de escuelas sino en los factores sociales que inciden en la calidad educativa. En palabras del subsecretario de Educación de ese entonces, Gustavo Iaies, “lo importante es que las escuelas sepan cuáles son los factores que más inciden en la obtención de buenos resultados para que los pongan en marcha”.

- A veces, también, la evaluación educativa fue interpretada por algunos medios como si los editores fueran controladores de un organismo de crédito y surgieron títulos como “La escuela no cumple con los objetivos mínimos”.
- En la Capital Federal, nuevamente, el 88 por ciento de los chicos que llegan a tercer año terminan la secundaria.
- Los alumnos rinden mejor en Matemática que en Lengua.
- La Argentina no está peor que el resto de América latina, pero América latina está al final del pelotón mundial.
- Es ilustrativo ver cómo interpretan los directivos de los colegios los resultados de cada evaluación:
- Se quejan de los sueldos docentes.
- Explican que los sueldos bajos influyen sobre la concentración, que resulta insuficiente, porque genera lo que en la Argentina se conocía como “profesor-taxi”, aunque obviamente se trata de un nombre-paradoja.
- Los directivos de escuelas públicas envidian a sus colegas de las privadas porque suponen que el pago a los docentes es sensiblemente superior.

En otras ocasiones, el reflejo periodístico de las evaluaciones apunta a lo particular particularísimo. Al caso. A la historia. Al costado humano.

Con la evaluación de 1998 en la Argentina proliferaron

los artículos de admiración por los resultados de las escuelas del interior del país.

La mejor primaria pública fue la escuela número 49 de Dorila, en La Pampa, en la zona centro-sur del país. Los alumnos del pueblo de 244 habitantes habían alcanzado un promedio de 91,43 puntos en Lengua y Matemática. Una profesora explicó los resultados notables por el trabajo personalizado: no había más de 16 chicos por aula. La directora contó que todos los días un ómnibus recorre 205 kilómetros para traer y llevar a 25 chicos que viven en los campos de los alrededores. También dijo que no había deserción porque cuando un chico faltaba las maestras iban hasta la casa a preguntar qué estaba ocurriendo en la familia.

También en 1998, la mejor escuela secundaria fue una técnica de San Nicolás, en el cinturón industrial que une Buenos Aires con Rosario, al norte de la Capital Federal. Los diarios reflejaron que los alumnos, en su mayoría de barrios periféricos, triunfaron en las pruebas, según el director, por su trabajo y por el empeño de los docentes.

Es fácil imaginar el reparo ante estos casos. ¿No se está, acaso, en presencia de algo en principio tan poco científico como el ejemplo? ¿La rareza, insumo básico de cualquier trabajo periodístico, no oculta la regularidad, objetivo máximo de una medición imparcial y objetiva? ¿No existe el riesgo de que las peculiaridades tapen una visión más profunda sobre la educación?

Naturalmente, los peligros están al alcance de la mano. Pero conviene apuntar que son los mismos de cualquier evaluación, e incluso de cualquier descripción de la realidad. En buena medida por razones ideológicas, la caída del Muro de Berlín y el famoso final de los grandes relatos supuso, para muchos científicos sociales, la desconfianza estructural

hacia todo tipo de análisis cualitativo y una absolutización del valor de lo cuantitativo, lo internacionalmente comparable a través de índices, otra vez los grandes números.

Para que quede claro: no es que lo cuantitativo sea “de derecha” y lo cualitativo “de izquierda”. Es que, por un momento, el sacudón del mundo fue interpretado por muchos intelectuales como un modo de finalización de la política. Si para el viejo comunismo el final de la política llegaría con la desaparición del Estado y la extinción de la sociedad clasista, para el enfoque inverso la desaparición de ese horizonte casi providencialista equivalía al vaciamiento de toda discusión de fondo, a la renuncia por anticipado a todo análisis que no se apoyara en un índice.

Para que quede claro también esto: sería tonto, inútil y fetichista al revés despreciar la cuantificación de metas y resultados. Sonaría a un antiintelectualismo que remata siempre en una política de primates, emparentada finalmente con el poder en estado puro y descarnado y una visión pragmática de las cosas en el peor sentido de la palabra, el del utilitarismo sin objetivos sociales ni planes desplegados desde el Estado.

Lejos de cualquier maniqueísmo, los ejemplos anteriores muestran que las evaluaciones educativas sirvieron para acercar elementos del debate educativo, de los especialistas, a los ciudadanos. Sin embargo, quien dice acercar no dice profundizar ni continuar. Ni el Estado, ni los planificadores educativos, ni los medios, ni los periodistas especializados aprovecharon esa formidable cantera de temas para seguir una discusión pública apasionante. Es posible, por ejemplo, que el éxito pampeano sea una rareza. Pero, ¿lo es también el debate sobre la proximidad entre docentes y alumnos? Y con el caso de San Nicolás, ¿no se discute también en público, con ab-

solita transparencia y sencillez, que no hay éxito posible sin una comunidad educativa fuerte, motivada y con mística? Es verdad, como se vio, que los diarios titularon muchas veces con un tono que disgustaba al gobierno por los buenos motivos (la superficialidad después de un trabajo tan serio como la evaluación) y también por los malos (a los funcionarios suele disgustarles la crítica). Pero, nuevamente, en cada artículo del ramo de los interesantes había material para el seguimiento y el debate. En el caso argentino, quizás la discusión pública sobre la educación sea pobre y poco específica, pero tal vez haya aportado su cuota de información y polémica al debate general sobre la pobreza, la exclusión y la desigualdad que forman parte de la agenda pública central del país.

¿Y los docentes? El reflejo de los maestros en los medios acostumbra ser simpático y cariñoso. Son, por ejemplo, el único gremio capaz de resolver huelgas con gran frecuencia y ser “perdonados” por la sociedad. Pero la evaluación es una pelea permanente que los medios reflejan marcando la desconfianza mutua entre el Estado y los docentes. Reflejan otra cosa también: que no hay acuerdo, acercamiento y discusión pública intelectualmente honesta entre los dos sectores. Los docentes desconfían de las evaluaciones oficiales y aun de las públicas no estatales. Y el Estado recela de lo que califica como falta de colaboración de los maestros.

En la Argentina tuvo mucho impacto una encuesta encargada por el gremio docente. Demostraba, entre otras cosas, estas cosas:

- El 40 por ciento de las maestras es jefe de hogar y tiene a su cargo entre dos y cinco personas.
- El 88 por ciento de los docentes trabaja en escuelas sin gimnasio.

- El 41 por ciento se desempeña en establecimientos sin biblioteca.
- El 18 por ciento de los maestros trabaja en escuelas con más de mil alumnos.
- El 90 por ciento de las escuelas no tienen servidor de Internet ni e mail.
- La mitad de los docentes está por debajo de la línea de pobreza.

Pero, a la vez, la encuesta fue reacia a descubrir y describir el mundo de los valores de los maestros, en rigor no muy distintos del mundo de valores del resto de la sociedad y por lo tanto no siempre noble, altruista y modélico para los alumnos.

Si la evaluación se convierte en una trampa, si al final parece que no midiera nada útil, si sólo genera problemas con los docentes y a primera vista no es útil para intensificar la discusión educativa dentro de la agenda pública, ¿no convalida su suprimirla?

Una tentación a mano de los funcionarios del Estado es que la evaluación sea secreta. No estoy de acuerdo por razones de principio y por motivos prácticos. Razones de principio: en una democracia moderna, y armas atómicas al margen, no hay secreto válido que obture la transparencia. Motivos prácticos: si es difícil conservar un secreto cuando el protagonista es una sola persona, es imposible mantenerlo si ya son como mínimo dos. Los seres humanos no suelen comportarse como conspiradores, y aun si lo lograsen no sería conveniente encarar la evaluación educativa como una tarea de topes dignos de John Le Carré.

La otra tentación es que la evaluación no sea secreta pero que, a la vez, no dependa de la difusión activa por parte de las autoridades oficiales. Mi opinión es que también éste es

un atajo sin sentido. La evaluación ya es una herramienta aceptada por la sociedad, que la espera. Tal vez la espere con una ansiedad desproporcionada, pero así es la vida. Quitarla ahora de la difusión habitual equivaldría a instalar una sospecha: por algo se oculta hoy lo que antes era público.

No quiero entrar en cómo el Estado evalúa la educación. No es mi especialidad. Pero desde un punto de vista más general –social, político, incluso de comunicación entre gobernantes y ciudadanos– parece claro que si la intención es contar con un Estado fuerte que opere activamente para reducir la desigualdad, la comunicación no puede limitarse a la administración de índices. Más allá de la evaluación, y sin sugerir de ningún modo la manipulación de los datos o su tergiversación, es evidente que el Estado puede orientar el debate educativo. Igual que en periodismo, al Estado debe exigírsele la mayor objetividad posible respecto de los datos y la descripción de la realidad. Pero, más aún que en periodismo, no hay gobierno sin encarnar en objetivos políticos y de liderazgo social las políticas públicas de la educación. Los datos son los datos: sagrados. Sería deseable, en mi opinión, que las conclusiones reflejaran cada vez más la intención de recoger el consenso social sobre educación y de alimentar ese consenso con nuevos temas, nuevas provocaciones, nuevos debates, nuevos objetivos.

Se dirá que no hay garantías de que el periodismo acompañe. Que incluso las mejores intenciones oficiales chocan con la ignorancia, los intereses creados y la indiferencia de los medios. Pero, otra vez, así es la vida. Ninguna iniciativa política tiene asegurada su repercusión, aunque al mismo tiempo conviene recordar que el impacto siempre es mayor cuando el Estado informa de una manera honesta, interesante, motivadora y poco burocrática. Y, del mismo modo

que ocurre con la libertad de expresión, no se trata sobre todo de la libertad para informar sino de la respuesta al derecho de los ciudadanos a saber. Ese derecho determina que informar sea, entonces, no una facultad sino un deber social, equivalente a un servicio público.

Tengo algunas conclusiones provisionarias:

Una, que la evaluación no es la panacea, como no lo es ninguna herramienta puramente cuantitativa.

La segunda, que si no es la fórmula de la felicidad universal no debe ser presentada como tal.

La tercera, que el Estado no debe suprimirla.

La cuarta, que nada reemplaza al ejercicio de la política.

La quinta, que la política sigue incluyendo la instalación de temas, de conclusiones, de valoraciones, válidos en tanto no surjan de la manipulación de los datos sino de su interpretación transparente y honrada.

La sexta, que así como no hay periodismo sin convertir la información en tema no hay evaluación bien comunicada y socialmente útil sin transformar el dato en eje de debate público.

La séptima, que el futuro de la evaluación como herramienta política está ligado a la participación de los docentes. Participación difícil y en apariencia imposible, pero insustituible.

La octava, que la sociedad, con una visión en este tema más cualitativa y sabia, no espera la evaluación de cada año para tomar posición frente a la realidad educativa.

La novena, que reproducir a escala internacional la lógica de los grandes números es vaciar la política como herramienta de ciudadanía y transformación de los habitantes en sujetos.

La décima, que siempre habrá tensión entre el Estado

y los docentes, y entre el Estado y los medios. Por eso, diseñar políticas o actuar como si esa tensión pudiera suprimirse sólo llevará a la creación de nuevos dogmas y nuevos ámbitos de rigidez. Una parte de la política es la imprevisibilidad, la contingencia, el azar, la pelea, el conflicto, el acuerdo, el consenso y el juego de las veleidades. Pensar que en sociedades culturalmente ricas y complejas como las nuestras toda realidad, toda evaluación y todo proceso comunicativo es controlable del principio al fin suena a una ilusión de chicos.

¿Qué pretendemos evaluar, qué evaluamos y qué conclusiones podemos extraer de la evaluación?

Alejandro Tiana Ferrer

Los años '90 se caracterizaron por la puesta en marcha y el desarrollo de un buen número de reformas educativas en los países latinoamericanos. Una vez superada la desconfianza hacia la educación que caracterizó a los años '70 y '80, los gobiernos de la región se lanzaron a una intensa actividad reformadora, que se dejó sentir en muchos países y que abarcó diversas esferas de la realidad educativa (Gajardo, 1999).

Las reformas emprendidas en esos años persiguieron varios objetivos, entre los que cabe destacar la extensión de la escolarización (especialmente en la enseñanza secundaria), la mejora de la calidad de la educación impartida y el aumento de la equidad en la distribución de los servicios educativos. Aunque no sea éste el lugar para hacer un balance de la década, no cabe duda de que los objetivos previstos se alcanzaron de manera muy desigual, consiguiéndose algunos logros destacables en el primer sentido de los mencionados, pero logrando resultados mucho más modestos en los otros dos. La situación es tal que un reciente informe internacional sobre el estado de la educación en los países de América Latina adoptaba como título el lema *Quedándonos atrás*, expresando de ese modo las carencias que aún aque-

jan a los sistemas educativos de la región (Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe, 2001).

El desarrollo de políticas nacionales de evaluación

Las reformas de los años '90 afectaron a diversos aspectos de la realidad educativa, que conviene detallar. Entre las orientaciones adoptadas por dichos procesos de reforma, y aun sin ánimo de exhaustividad, pueden distinguirse varias líneas de actuación que, si bien no se dieron siempre de manera simultánea, estuvieron presentes de uno u otro modo en las diversas iniciativas nacionales:

- Un primer aspecto en que se avanzó considerablemente, aunque no sin problemas, fue en la redistribución de las competencias y responsabilidades en materia de educación, lo que supuso la puesta en marcha de políticas de descentralización y de autonomía escolar, acompañadas en ocasiones de procesos de privatización, cuyos efectos fueron ambivalentes (Winkler y Gershberg, 2000).
- Una segunda orientación de las reformas, especialmente influyente en muchos países de la región, consistió en la introducción de cambios en la organización curricular, que llevaron generalmente aparejada una revisión de los planteamientos fundamentales en que se inspira la definición y la construcción del currículo.
- Un tercer dominio de acción de las reformas tuvo que ver con los procesos de formación inicial y capacitación del profesorado, considerado habitualmente pieza clave de

las reformas, pero no siempre tan atendido como debiera en su papel de agente transformador.

- Un cuarto ámbito de reforma estuvo concretamente relacionado con la educación secundaria o media, que se ha convertido en la pieza clave y más conflictiva de los sistemas educativos y cuya expansión se ha ido también produciendo en América Latina en las últimas décadas (Braslavsky, 2001).
- Por fin, una última orientación tuvo que ver con el desarrollo de mecanismos y modelos de evaluación; dicho de otro modo, con la consideración de la evaluación, como un poderoso instrumento al servicio de la gestión de la educación y de la mejora de su calidad (Tiana, 1996).

Esta última orientación de las reformas de los '90 cobra especial relevancia desde la perspectiva que se adopta en este trabajo, pues es precisamente el impacto de esos mecanismos de evaluación lo que está sometido a crítica y discusión. Por este motivo, merece la pena realizar un par de comentarios adicionales sobre sus rasgos más significativos.

Una primera observación que cabe destacar desde este punto de vista es que la mayoría de los países de la región pusieron en marcha sus propios sistemas nacionales de evaluación a lo largo de los años '90, generalmente en el marco de procesos más amplios de reforma educativa. Con configuraciones institucionales muy diferentes entre sí y con ámbitos de actuación diversos, fueron varios los gobiernos que se sumaron a la experiencia pionera que otros países como Chile venían desarrollando ya con anterioridad. Argentina, México, Colombia, Bolivia o Brasil, por no citar sino algunos casos destacados, dieron pasos decididos en esa dirección, de manera que al final de la década eran mayoría los países la-

tinoamericanos que contaban con algún tipo de sistema nacional de evaluación (Ravela, 2001).

Sin embargo, el indudable interés por la evaluación de la educación que se despertó en muchos países no marchó asociado sino mucho más tardíamente con un interés paralelo por participar en estudios comparativos internacionales de evaluación del rendimiento educativo. Aparte de la valiosa experiencia regional del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, en el que participaron un total de trece países de América Latina y el Caribe, los estudios promovidos por organizaciones tan reconocidas como la IEA o la OCDE no han alcanzado todavía en la actualidad cotas altas de participación (Tiana, 2000). Durante los años '90, fueron más bien los programas de cooperación orientados hacia el reforzamiento de los sistemas nacionales de evaluación, como el impulsado en esa dirección por la OEI, los que lograron una participación más amplia.

Así pues, y hablando siempre en términos generales, la situación de la evaluación educativa en la región latinoamericana se caracteriza hasta el momento por la existencia de un contraste llamativo entre el desarrollo de ambiciosos programas nacionales de evaluación y una limitada participación en estudios comparativos internacionales. El desarrollo de esos programas nacionales está basado en la puesta en marcha de un buen número de instituciones y agencias de evaluación, cuya experiencia comienza ya a ser importante.

Una mirada a los sistemas nacionales de evaluación

Dando un paso más en el análisis, conviene que nos pre-

guntemos por las principales características que presentan esos sistemas nacionales de evaluación que tanto auge adquirieron durante la última década. Tres son las preguntas clave que debemos hacernos, que están ligadas respectivamente a otras tantas dimensiones de la evaluación: ¿para qué se evalúa?, ¿qué se evalúa? y ¿qué uso se hace de la información?

a) *Una multiplicidad de propósitos*

La primera pregunta tiene que ver con los propósitos que orientan la evaluación de la educación. Y anticipando la respuesta, hay que reconocer que los sistemas nacionales de evaluación persiguen una multiplicidad de propósitos, al orientarse hacia la consecución de uno o varios de los siguientes objetivos:

- Ayudar a las escuelas y a los profesores a mejorar la calidad de la educación que imparten, proporcionándoles una información relevante y significativa que les permita valorar su situación real y sus logros.
- Conocer y valorar en qué medida se están alcanzando los objetivos educativos o los estándares propuestos para los distintos niveles o parcelas del sistema educativo, contribuyendo al mismo tiempo a establecer metas y niveles de referencia.
- Forzar a las instituciones escolares y a los agentes educativos a generar dinámicas de cambio, por medio de la comparación o la competencia.
- Rendir cuentas a la sociedad o a las autoridades educativas acerca de los logros conseguidos, eliminando así la opacidad del sistema educativo y de las instituciones escolares.

- Establecer sistemas de incentivos para las instituciones y para los profesionales de la educación.
- Certificar, acreditar o seleccionar instituciones, programas de estudio, profesores o estudiantes.
- Conocer y valorar la situación del sistema educativo, así como el impacto producido por las políticas adoptadas.
- Introducir elementos de racionalidad en el debate público y en la construcción de políticas en el ámbito de la educación.

Es cierto que estos propósitos no son en todos los casos compatibles entre sí, pero no lo es menos que las políticas nacionales de evaluación seleccionan algunos frente a otros, les conceden distinto énfasis, establecen prioridades entre ellos o los interpretan de diferente manera. Así, por ejemplo, un propósito compartido por muchos, como es el caso de la incentivación del cambio a través de la comparación, no deja de ser objeto de polémica. Mientras que en algunos países ha llevado, por ejemplo, a la publicación de listas de clasificación de escuelas o universidades, con el ánimo de forzar una competencia abierta por conseguir más estudiantes o más recursos, en otros se ha rechazado frontalmente esa posibilidad y se han establecido mecanismos de confidencialidad de los datos que impiden ese tipo de uso, aunque permitiendo beneficiarse de la comparación. Como puede apreciarse, un mismo propósito puede inspirar políticas muy diferentes.

b) *Una diversidad de actuaciones*

La segunda pregunta tiene que ver con la actividad desarrollada por los sistemas nacionales de evaluación. Y también en este caso hay que anticipar la existencia de una considerable diversidad de modalidades de actuación, referidas a su vez a distintos ámbitos educativos. Entre los principales as-

pectos que son objeto de evaluación en la actualidad, hay que destacar cuatro campos prioritarios:

- Un primer campo que es objeto de considerable atención se refiere a la evaluación de los logros conseguidos por los alumnos. Muchas de las actividades desarrolladas en la actualidad por los sistemas nacionales de evaluación, y sin duda las más importantes de entre ellas, están orientadas a la valoración de los resultados de la educación, medidos a través de los logros que consiguen los estudiantes. En la mayoría de los casos, esas actuaciones se centran en algunas áreas curriculares seleccionadas, especialmente las que se consideran fundamentales para la formación de los jóvenes (Lenguaje, Matemáticas, Ciencias, Historia). A veces se complementan con la evaluación de los logros conseguidos en otros ámbitos de la formación escolar, como la educación cívica, el dominio de las tecnologías de la información y la comunicación o las estrategias de aprendizaje. Las evaluaciones suelen estar referidas a etapas o momentos clave del proceso educativo, tales como el tránsito de una etapa a otra o el final de la escolarización obligatoria o de un determinado nivel de estudios. En ocasiones se evalúa a la totalidad de los estudiantes de los grados seleccionados, mientras que otras veces se trabaja solamente con muestras representativas. En cualquier caso, es ésta una línea de actuación prioritaria para los sistemas nacionales de evaluación.
- Un segundo campo que también atrae considerable atención se refiere al rendimiento que logran las instituciones educativas. El interés que ha generado la apertura de la “caja negra” que constituyen las instituciones ha estimulado el desarrollo de diversos programas orientados a su

evaluación. Tanto las escuelas primarias y secundarias como las universidades han sido el objetivo prioritario de muchos planes de evaluación y de acreditación puestos en marcha en diversos países latinoamericanos. Esos planes de evaluación tienen características heterogéneas, pero suelen coincidir en conceder una atención especial a los resultados conseguidos por los estudiantes, aunque sin dejar generalmente de lado el análisis de la organización y el funcionamiento de la institución de que se trata.

- Un tercer campo que ha ido desarrollándose recientemente, si bien con mayores dificultades que los dos anteriores, es el relativo al desempeño profesional de los docentes. La convicción de que la actuación de los profesores está en la base de los resultados conseguidos por los estudiantes, aunque no sea su único determinante, ha generado un conjunto de actuaciones de diverso tipo, entre las que también se cuentan algunos programas de evaluación. Se trata de acciones que no han dejado de generar cierta polémica, tanto por sus características técnicas como por el tipo de uso que se hace de la información obtenida. En muchos casos, quizás en la mayoría, se ha optado por utilizar la evaluación docente en el marco de planes de incentivos profesionales, eludiendo sus aspectos más punitivos, pero en otros casos la evaluación se ha utilizado como instrumento de selección o de promoción profesional, o incluso se ha asociado con mejoras salariales o laborales.
- Un cuarto campo que también ha atraído el interés de los sistemas de evaluación se refiere al impacto de las políticas educativas adoptadas. En un contexto de reformas educativas, como las impulsadas en la región durante los '90, no es extraño que se haya planteado abiertamente la

pregunta acerca de los efectos y el impacto de los procesos de cambio iniciados y de las medidas adoptadas. Desde esta perspectiva, han coexistido dos tendencias que deben diferenciarse, aunque ninguna de ellas haya llegado a atraer tanta atención como las arriba mencionadas. La primera se refiere a la evaluación de los programas más emblemáticos de los procesos de reforma, los denominados “programas estrella”, aquellos en que las autoridades educativas han realizado una mayor inversión y depositado mayores esperanzas. La segunda se refiere a la evaluación de los propios procesos de reforma, con la pretensión de establecer un mecanismo de seguimiento sistemático de sus efectos. Pese a la limitada experiencia desarrollada en ambas direcciones, no cabe duda de que se trata de un campo llamado a experimentar un mayor desarrollo en el futuro próximo.

Aunque los cuatro campos mencionados no agotan la totalidad de los ámbitos abarcados por los sistemas nacionales de evaluación, no cabe duda de que se trata de los más destacados y los que han experimentado un mayor empuje. Nuevamente hay que insistir en que no todos ellos han sido desarrollados en todos los países, ni del mismo modo, ni con la misma energía, pero sí puede afirmarse que son objeto de atención especial por parte de los sistemas nacionales de evaluación.

c) *Una pluralidad de usos*

La tercera pregunta tiene que ver con el tipo de uso que se hace de la información obtenida por medio de la evaluación. Y al igual que en los dos casos anteriores, también hay que reconocer la existencia de una pluralidad de situaciones. In-

tentando analizar los usos más habituales de los resultados de la evaluación, podemos distinguir cinco tendencias predominantes:

- Una primera tendencia consiste, por paradójico que pueda resultar, en la renuncia a hacer uso de la información obtenida. Aunque no resulte muy justificable, ni desde el punto de vista de la eficacia administrativa, ni desde la perspectiva de la transparencia, ni menos aún desde las exigencias del comportamiento político democrático, hay que reconocer que existen casos en que los resultados de la evaluación no son publicados, ni difundidos, ni apenas utilizados.
- Una segunda tendencia, que es la más habitual, consiste en la elaboración y difusión de informes de evaluación. Los informes pueden ser de diverso tipo, tener una cobertura variable (nacional, regional, sectorial) y estar dirigidos a audiencias distintas (familias, profesores, autoridades educativas, investigadores y técnicos, etc.), llegando incluso a adoptar formatos diferentes. Lo importante en este caso no son las características concretas de los informes, que pueden ser de hecho muy diferentes, sino la confianza subyacente en que la difusión pública de los resultados puede satisfacer las demandas de información planteadas por los diversos agentes educativos. En algunos casos se han diseñado y aplicado estrategias bastante completas de difusión, logrando un impacto social y mediático importante.
- Una tercera tendencia consiste en la devolución confidencial de resultados a las instituciones participantes en la evaluación, con el propósito de que puedan utilizar los datos recibidos para elaborar sus propios planes de mejo-

ra, sin verse necesariamente expuestas al escrutinio público. Los modos concretos que puede adoptar esa devolución son muy variables, pero lo que caracteriza esa modalidad de actuación es su confianza en el valor intrínseco que encierra la comparación y en la capacidad de reflexión y de mejora que tienen las instituciones y los agentes de la educación cuando se enfrentan con su propia realidad.

- Una cuarta tendencia, no excesivamente habitual y desde luego controvertida, consiste en la publicación de los resultados obtenidos, en forma de tablas de clasificación. En algunos casos la publicación se realiza por los responsables de la evaluación, mientras que en otras ocasiones se permite o incluso se favorece que sean otros agentes (prensa, investigadores) quienes publiquen la información. Este tipo de uso resulta más amenazador para las instituciones educativas que los anteriores, motivo por el cual ha recibido críticas y generado resistencias.
- Una quinta tendencia consiste en la utilización de los resultados de la evaluación para proporcionar orientación pedagógica a instituciones y profesores. Este tipo de uso de la información obtenida a través de la evaluación adopta dos modalidades distintas. En primer lugar, son varios los sistemas nacionales que han elaborado instrumentos tales como cuadernillos de asesoramiento a los profesores, en los que se analizan las respuestas correctas de las pruebas, así como los errores más habituales que suelen producir los estudiantes, y se finaliza ofreciendo sugerencias didácticas más o menos concretas. En segundo lugar, aunque sea de forma más esporádica, también se han desarrollado acciones de orientación para instituciones determinadas, ayudándolas a elaborar sus

propios planes de mejora, y se han incluido análisis de los resultados de la evaluación en programas de formación docente. Este tipo de uso, que es menos tradicional que la difusión de informes, está recibiendo una atención creciente por parte de diversas administraciones educativas.

Igual que sucedía en los epígrafes anteriores, también hay que decir aquí que los usos mencionados no son necesariamente excluyentes, sino que son cada vez más las administraciones educativas que los combinan en distintos modos. Es cierto que algunos de ellos resultan contrapuestos, como puede ser la alternativa existente entre la publicación de los resultados de las escuelas y la entrega confidencial de los datos, pero no siempre es el caso. De hecho, la combinación de la publicación de informes, la devolución de resultados a las instituciones participantes en la evaluación y la elaboración de instrumentos de orientación didáctica constituye una realidad bastante extendida en los países de la región.

A la vista de la situación que se acaba de exponer cabe extraer una primera conclusión, consistente en que los sistemas nacionales de evaluación de la región latinoamericana no recurren a estrategias de acción uniformes ni adoptan enfoques de evaluación idénticos. Y esta conclusión no es sólo una confirmación de los hechos que se han presentado, sino que se debe incluso considerar lógico que exista tal variedad. En efecto, si aceptamos la existencia de una multiplicidad de propósitos, una diversidad de actuaciones y una pluralidad de usos, debemos aceptar la diversidad interna de los sistemas de evaluación. Serán precisamente los propósitos concretos que se persigan, los estilos de actuación que se adopten y los usos que se prefieran, los que determinarán las características del sistema de evaluación.

Una mirada a las pruebas nacionales de logros

Como se indicaba en el apartado anterior, los sistemas nacionales de evaluación que se han desarrollado en los países latinoamericanos desde los años '90 vienen prestando una atención especial a la medición de los logros conseguidos por los alumnos. Los instrumentos que se aplican con esa finalidad son las genéricamente denominadas *pruebas nacionales*, que merecen una mirada especial por la importancia que han llegado a adquirir.

Las pruebas nacionales presentan bastantes diferencias de unos países a otros, aunque también ofrecen algunas similitudes. Entre sus elementos comunes, quizás el más destacado sea la preocupación prácticamente universal que demuestran por evaluar el rendimiento alcanzado en Lenguaje y Matemáticas. No obstante, aunque se trata de dos áreas especialmente atendidas por los sistemas de evaluación, no hay que pensar que sean los únicos dominios evaluados. En efecto, la medición de los logros en ambas áreas suele ir acompañada por la evaluación de otras áreas curriculares complementarias (entre las que sobresalen las Ciencias, la Historia o las Lenguas extranjeras) o de otras capacidades de tipo transversal, no ligadas necesariamente a áreas específicas (como el autoconcepto, algunas estrategias de aprendizaje o ciertas actitudes).

Las pruebas nacionales suelen aplicarse preferentemente en los grados terminales de las principales etapas educativas (sobre todo, al final de la enseñanza primaria y de la secundaria). No obstante, también se aplican a veces en otros grados que tienen un especial interés desde el punto de vista diagnóstico (por ejemplo, el final de ciertos ciclos), aun-

que la identificación de tales puntos focales varíe de unos sistemas educativos a otros. El reciente desarrollo de algunos estudios internacionales ligados a la edad de los evaluados y no al grado que cursan, como es el caso del proyecto PISA de la OCDE, aún no ha encontrado traducción directa en los sistemas nacionales de evaluación latinoamericanos, ni es previsible que lo haga a corto plazo. El interés que despierta la valoración de los logros conseguidos al final de determinadas etapas supera hoy en la región al que suscita la valoración del nivel de formación a ciertas edades. La participación en aquellos proyectos internacionales parece satisfacer suficientemente las preguntas que puedan plantearse en este último sentido.

Dependiendo del propósito que guía a la evaluación, las pruebas se aplican en ocasiones a poblaciones o cohortes estudiantiles completas, mientras que en otros casos se limitan a muestras representativas de estudiantes. La primera decisión permite hacer un uso más singularizado de los datos obtenidos, tal como comparar el rendimiento de diversas escuelas o estudiantes, mientras que la segunda es más adecuada para usos diagnósticos. Por otra parte, hay países en que las pruebas se aplican cada año, aunque es más frecuente que se adopten ciclos plurianuales, de distinta duración.

Las pruebas nacionales parecen satisfacer ampliamente a las autoridades educativas, a la vista del uso tan extenso que se hace de ellas. No obstante, en un análisis más profundo también presentan insuficiencias y problemas, que requieren una consideración más atenta.

Tres son las principales insuficiencias de las pruebas nacionales de logros que han subrayado los especialistas (Ravela, 2000). La primera consiste en la existencia de algunas debilidades técnicas en los procesos de desarrollo y vali-

dación de los instrumentos de medición. La segunda se refiere a la insuficiente calidad y capacidad que las pruebas ofrecen para evaluar aprendizajes complejos. La tercera tiene que ver con el insuficiente aprovechamiento que suele hacerse de la información obtenida.

Los problemas a los que se ha hecho alusión son de distinto tipo. Para comenzar, habría que señalar algunos relativos al diseño y confección de los instrumentos de evaluación:

- Un primer problema tiene que ver con el contenido de las pruebas, asunto que plantea dos tipos de dificultades. La primera se refiere a la relación que debe existir entre las pruebas y los objetivos educativos. Lo deseable es que las pruebas estén referidas a los objetivos establecidos en el sistema educativo, de manera que permitan valorar el grado en que se alcanzan las metas previstas. Aunque este planteamiento resulta lógico y plausible, plantea dificultades cuando no existen objetivos explícitos (lo que sucede en muchos casos), o cuando los objetivos de cada curso o etapa se formulan de manera poco concreta. Cuando no se pueden determinar los objetivos de forma precisa, resulta sin duda más difícil elaborar las pruebas. La segunda dificultad se refiere al tratamiento de la diversidad curricular existente entre diferentes regiones, distritos o escuelas. La mayoría de los modelos curriculares aplicados en América Latina conceden cierto grado de autonomía a las autoridades regionales o distritales o a las propias escuelas para definir el currículo. En consecuencia, no resulta sencillo hacer pruebas que abarquen todas las enseñanzas efectivamente impartidas en todo el sistema educativo. Las soluciones posibles son dos: elaborar

pruebas de mínimos, lo que implica reducir el campo de evaluación, o elaborar pruebas comunes a partir de un consenso profesional o técnico. Ambas soluciones son utilizadas en la región, sin que quepa conceder prioridad a una de ellas sobre la otra.

- Un segundo problema tiene que ver con algunas disyuntivas que se plantean en relación con el diseño y la confección de las pruebas. La primera disyuntiva se refiere al enfoque de las pruebas, que pueden estar centradas en los conocimientos adquiridos por los alumnos o abarcar también el desarrollo de actitudes. Las pruebas que miden unos y otras no son iguales, ni siquiera plantean las mismas exigencias técnicas, lo que obliga a decidir su enfoque. Una segunda disyuntiva se refiere al carácter de las pruebas, pudiéndose optar por la elaboración de pruebas referidas a criterios de logro o por la confección de pruebas de tipo normativo, adaptadas a una distribución de resultados estadísticamente normal. La tercera disyuntiva se refiere a la delimitación de la población sujeta a evaluación, que permite optar por pruebas de aplicación censal (dirigidas a toda la población estudiantil de un determinado grado o edad) o de tipo muestral (dirigidas a muestras representativas de población).

Otro grupo importante de problemas están relacionados con la explicación de los resultados conseguidos. La mayoría de los sistemas nacionales de evaluación no suelen contentarse con medir los logros alcanzados por los estudiantes, sino que también pretenden explicar por qué se producen tales resultados. Sin embargo, dicho intento de explicación tropieza con varios problemas y dificultades:

- Un primer problema tiene que ver con el grado de coherencia realmente existente entre la evaluación y el currículo. Como se señalaba al comienzo del trabajo, el desarrollo de los sistemas nacionales de evaluación ha ido generalmente asociado a otros procesos de reforma curricular. No obstante, esa simultaneidad no ha asegurado siempre la existencia de una adecuada coherencia entre ambos aspectos. En los casos más extremos, puede incluso decirse que ambos procesos se han desarrollado en direcciones divergentes. Cuanto más importancia se ha concedido a las pruebas, mayor ha sido su impacto sobre el desarrollo curricular.
- Un segundo problema, menos complejo pero no menos influyente, está relacionado con el tipo de interpretación que permiten los diversos modelos de pruebas. Como se señalaba más arriba, dos son los modelos de pruebas utilizados más habitualmente: criteriosiales y normativas. Las pruebas de tipo criterial proporcionan una base sólida para valorar la suficiencia de los logros conseguidos, ya que se apoyan en una definición operativa de los objetivos que deben alcanzarse. Por el contrario, las pruebas de tipo normativo exigen alguna operación adicional para poder valorar la suficiencia de los logros alcanzados, dado que se refieren a situaciones de normalidad estadística. La existencia de estos dos modelos alternativos obliga a considerar cuidadosamente qué usos se quieren hacer de las pruebas, antes de decidir el que conviene aplicar.
- Un tercer problema se refiere a las dificultades conceptuales y técnicas que se plantean a la hora de buscar explicaciones. Por una parte, hay que reconocer la debilidad teórica de los modelos habituales de producción educativa, que en buena parte deriva del problema que plantea la

causalidad en educación. Esa debilidad influye en la selección de los factores que se consideran asociados al rendimiento, pocos de los cuales cuentan con confirmación empírica sólida (Scheerens, 1996). Además, la explicación de los resultados se ha basado tradicionalmente en la distinción de dos tipos de variables, unas extrínsecas y otras intrínsecas. Mientras que las primeras (nivel socio-económico y cultural, recursos destinados a la educación) han sido objeto de análisis abundantes, las segundas (organización del sistema educativo, procesos institucionales, procesos de aula) vienen resultando menos concluyentes y aún requieren investigaciones adicionales.

- Un cuarto problema tiene que ver con la voluntad que a veces existe de buscar muchas vías de explicación de los resultados obtenidos, que puede traducirse en un exceso de datos recogidos. Cuando ese exceso de datos va unido, como ocurre en ocasiones, con una insuficiente delimitación conceptual de las variables seleccionadas y con limitaciones en el análisis estadístico de los datos, la situación puede llegar a resultar inmanejable. La consecuencia suele ser una evidente infrautilización de los datos y una ineficiencia en el uso de los recursos disponibles.
- Un último grupo de problemas que plantean las pruebas de logros están relacionados con el uso que se hace de las mismas y con las interpretaciones a que dan pie:
- Un primer problema tiene que ver con los distintos tipos de uso que pueden hacerse de la información. Los resultados de la evaluación pueden utilizarse con una finalidad formativa, orientada al desarrollo institucional y profesional de los agentes implicados, para fomentar la competición, e incluso para impulsar políticas de mercado en el ámbito educativo. Obviamente, las interpretaciones que

permiten tales posibilidades son diferentes y persiguen distintos efectos. Es necesario señalar que, si bien esos son los extremos de un continuo de posibilidades de uso, existen otras opciones intermedias en que dichos propósitos pueden combinarse en diferente proporción. La decisión sobre el tipo de uso que se pretende hacer de los datos de la evaluación condiciona en buena medida el diseño de la evaluación.

- Un segundo problema tiene que ver con la existencia de diversas posibilidades de comparación (con una norma o criterio, con otras realidades semejantes, consigo mismo a lo largo del tiempo), cada una de las cuales plantea sus propias exigencias y ofrece diferentes posibilidades. El uso de la comparación plantea un problema complicado, que tiene que ver con la justicia de la comparación. El intento de dar respuesta a esa cuestión ha impulsado el desarrollo de técnicas de cálculo del denominado *valor añá - dido*.
- Un tercer problema consiste en el excesivo reduccionismo en que a veces se cae cuando se pretende interpretar los resultados de la evaluación. Ese defecto puede venir originado por una excesiva limitación del contenido de las pruebas, que no permite llegar a conclusiones razonables sobre el logro de los objetivos propuestos, o por una interpretación poco rigurosa de los datos, que puede incluso llegar a distorsionar las interpretaciones.

A la vista del análisis que acaba de hacerse de las pruebas nacionales de logros, se llega a la conclusión de que constituyen la práctica de evaluación más habitual en la región, encontrándose bastante asentadas en la actualidad. No obstante, como se ha puesto de manifiesto, presentan algunas defi-

ciencias y problemas que deberán solventarse en el futuro, si se quiere asegurar su credibilidad y se pretende que contribuyan a mejorar la educación.

Algunas reflexiones finales

Si bien es cierto que las pruebas nacionales de logros constituyen la práctica más habitual de evaluación educativa en América Latina, no es la única que se desarrolla en la actualidad. Como se indicaba en las primeras páginas del trabajo, existen otros tres ámbitos que son también objeto de evaluación. El primero corresponde a los programas y las políticas de educación, que despiertan especial interés en el actual contexto de revisión de las reformas educativas puestas en marcha durante los años '90. En sus formas más elaboradas ha servido para diseñar mecanismos de seguimiento de los procesos de cambio, aunque se trata de una práctica escasamente implantada. El segundo ámbito corresponde a las instituciones educativas, para cuya evaluación se han diseñado y aplicado algunos modelos, tradicionales o innovadores, aunque sin que pueda decirse que esta tendencia haya alcanzado todavía suficiente desarrollo. El avance más importante que se ha logrado en este sentido consiste en la expansión de una conciencia nueva acerca del valor que tiene la evaluación institucional con vistas a la mejora de la calidad de la educación. El tercer ámbito, y al mismo tiempo el más controvertido, corresponde al desempeño docente. Aunque en algunos países se han desarrollado mecanismos de valoración de la tarea docente que están conectados con la medición del rendimiento de los estudiantes, no existen todavía experiencias suficientes ni de larga tradición en este sentido.

Las experiencias que han explorado otras direcciones son aún más escasas.

En conjunto, puede afirmarse que estos otros tres ámbitos abarcados por la evaluación educativa están actualmente infradesarrollados en relación con las pruebas de rendimiento. Hay que subrayar, no obstante, que la evaluación de tales aspectos puede contribuir notablemente a la mejora de la calidad de la educación. Precisamente de esa potencialidad procede el interés que despiertan en muchos países de la región.

Como conclusión final, cabe afirmar que la evaluación es una realidad integrada que cuenta con varios componentes: estudios internacionales de evaluación, estudios diagnósticos nacionales del rendimiento, evaluación de las instituciones educativas y otras evaluaciones complementarias de diversos aspectos de la actividad educativa (preparación y desempeño de los docentes, impacto de las políticas aplicadas, funcionamiento de algunos programas relevantes). Aunque no todos ellos se han desarrollado con la misma energía, los sistemas nacionales de evaluación no deberían infravalorar la importancia de concebirllos de forma integrada, de manera que las acciones emprendidas se refuercen mutuamente.

El colofón a estas reflexiones que provoca la mirada a los sistemas nacionales de evaluación educativa consistiría en que la evaluación debe considerarse como un instrumento para la mejora de la calidad de la educación y no como una panacea capaz de solucionar todos los problemas que la educación plantea. Otra consideración diferente de ésta no puede sino conducir a algunos errores de cierta importancia.

Bibliografía:

Braslavsky, C. (2001): *Educación secundaria: ¿cambio o inmutabilidad?*, Buenos Aires, Santillana.

Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe (2001): *Quedándonos atrás. Un informe del progreso educativo en América Latina*, Santiago de Chile, PREAL.

Gajardo, M. (1999): *Reformas educativas en América Latina. Balance de una década*, Santiago de Chile, PREAL (documento de trabajo nº 15).

Ravela, P., ed. (2000): *Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?*, Santiago de Chile, PREAL.

Ravela, P. (2001): *¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?*, Santiago de Chile, PREAL.

Scheerens, J. (1996): "Can the School Effectiveness Knowledge Base Guide School Management?", en *II International Conference on School Management. Participative Management and School Evaluation*, Bilbao, Universidad de Deusto, pp. 98-119.

Tiana, A. (1996): "La evaluación de los sistemas educativos", *Revista Iberoamericana de Educación*, nº 10, 1996, pp. 37-61.

Tiana, A. (2000): *Cooperación internacional en evaluación de la educación en América Latina y el Caribe. Análisis de la situación y propuestas de actuación*, Washington, BID - Departamento de Desarrollo Sostenible.

Winkler, D. y Gershberg, A.I. (2000): *Los efectos de la descentralización del sistema educacional sobre la calidad de la educación en América Latina*, Santiago de Chile, PREAL (documento de trabajo, nº 17).

Una mirada técnico-pedagógica acerca de las evaluaciones de calidad educativa

Felipe Martínez Rizo

Introducción

Las preguntas sugeridas como punto de partida para la reflexión de esta mesa fueron tres: ¿Qué estamos evaluando? ¿Qué de lo evaluado resulta significativo? ¿Por qué las reformas educativas realizadas a lo largo de la década del 90 no se ven traducidas en los resultados de las evaluaciones?

Para aventurar la respuesta que considero debe darse a esas tres preguntas considero necesario presentar antes la que considero apropiada para una pregunta más: ¿qué deberíamos evaluar de modo que la evaluación sirva para mejorar la calidad educativa?

Organizaré mi presentación, pues, en cuatro apartados, cada uno de los cuales pretende responder una de las cuatro preguntas mencionadas.

¿Qué deberíamos evaluar para que la evaluación sirva para mejorar la calidad?

Ante un auditorio interesado en la educación no es necesario abundar en lo relativo a la importancia de su calidad y su

evaluación. Pero teniendo en cuenta la diversidad de concepciones de estas dos nociones (calidad y evaluación) juzgo necesario explicitar mis propias ideas al respecto, de modo que las respuestas que ofreceré para las tres preguntas siguientes cobren sentido pleno.

Tratando, pues, de responder a esta primera pregunta, afirmo que, idealmente, una evaluación adecuada para que sirva realmente para mejorar la calidad educativa sería una muy amplia, que incluyera todas las dimensiones de la calidad; todas las áreas del currículo; los aspectos cognitivos pero también los actitudinales y valorales; no sólo niveles de dominio elementales, sino también los más elevados; insu- mos, pero también procesos y productos; los diversos actores y niveles organizacionales del sistema educativo; y que, además, incluyera el análisis de los factores que inciden en la calidad, de manera que dé bases para el diseño y la implementación de políticas.

Desarrollo el primero de los elementos anteriores: una buena evaluación deberá comprender todas las dimensiones de la calidad.

Sabemos que las políticas educativas han puesto el acento sucesivamente en la atención de la cobertura, luego en la eficiencia terminal y, más recientemente, en el nivel de aprendizaje alcanzado por los alumnos y en la equidad del servicio educativo. La coincidencia de la secuencia anterior de tales políticas en muchos países de la región latinoamericana y otras partes del mundo no es casual: responde, desde luego, a la similar evolución del contexto demográfico, económico, social y cultural.

Es frecuente que el término calidad se defina en forma restringida, limitándolo a lo que se refiere a los niveles de aprendizaje. En este sentido se dice que, tras centrar la aten-

ción en la cobertura y la eficiencia, ahora las políticas educativas ponen el énfasis en la calidad. Otra manera de usar la palabra, que juzgo preferible, le da un sentido más amplio, que se aplica también a las políticas de cobertura y eficiencia terminal.

Un concepto amplio de calidad, en efecto, no puede dejar fuera las dimensiones de cobertura y eficiencia, aunque no pueda limitarse a ellas y deba incluir además el nivel de aprendizaje, entre otras cosas.

Un concepto amplio de calidad debe incluir, en mi opinión, varias dimensiones que, en forma sintética, pueden expresarse diciendo que un sistema educativo de calidad es aquél que:

- Establece un currículo adecuado a las necesidades de la sociedad, incluyendo las de una mayor productividad económica, pero también otras necesidades básicas en una perspectiva de desarrollo integral, como las que tienen que ver con la democracia política, el respeto de los derechos humanos, el desarrollo de la ciencia, el cuidado del medio ambiente y la preservación y enriquecimiento de la diversidad cultural. Esta dimensión puede definirse con la expresión *relevancia de los objetivos curriculares*.
- Logra que la más alta proporción posible de destinatarios acceda a la escuela, que permanezca en ella hasta el final del trayecto previsto y que egrese alcanzando los objetivos de aprendizaje establecidos. Esta dimensión incluye, pues, la cobertura y la eficiencia terminal, así como el nivel de aprendizaje, y en términos sistémicos coincide con la *eficacia interna* del sistema.
- Consigue que los aprendizajes logrados por los alumnos sean asimilados por éstos en forma duradera y deriven en

comportamientos sociales sustentados en los valores de libertad, equidad, solidaridad, tolerancia y respeto a las personas, que son fructíferos para la sociedad y para el propio individuo, quien podrá así alcanzar un desarrollo pleno en los diversos roles que habrá de desempeñar como trabajador, productor, consumidor, padre de familia, elector, servidor público, lector y telespectador, entre otros; en pocas palabras, como ciudadano cabal. En términos sistémicos esta dimensión es la de la *eficacia externa o impacto* del sistema.

- Cuenta, para lograr lo anterior, con recursos humanos y materiales suficientes, y los aprovecha de la mejor manera posible, evitando despilfarros y derroches. Es fácil apreciar que esta es la dimensión *eficiencia* del sistema.
- Tiene en cuenta la desigual situación de alumnos y familias, de las comunidades en que viven y las escuelas mismas, y ofrece apoyos especiales a quienes lo requieren, para que los objetivos educativos sean alcanzados por el mayor número posible. Esta última dimensión es la de *equidad*.

Conceptualizada así, multidimensionalmente, la educación será de calidad si es *relevante*, si tiene *eficacia interna y externa*, si tiene un *impacto* positivo en el largo plazo, si es *eficiente* en el uso de los recursos y si busca la *equidad*.

Considero que es mejor este concepto amplio de calidad que uno restringido, porque creo que nadie creará que un sistema educativo es de calidad si atiende sólo a una fracción de la población en edad de acceder a cierto nivel, o si pierde en el camino a muchos de los alumnos que lo comienzan, aunque los que terminan lo hagan con altos niveles de aprendizaje.

Los otros componentes de una buena evaluación son claros: todas las áreas del currículo, comenzando por supuesto con Lectoescritura y Matemáticas, pero sin limitarse a ellas, sino contemplando, al menos en el mediano plazo, las ciencias naturales y las del hombre, la educación física y artística, etc.

Deberán incluirse los aspectos cognitivos pero también los actitudinales y valorales, ya que la educación no sólo pretende desarrollar los primeros, sino también, y con creciente énfasis, los segundos, cuya importancia para la vida en una sociedad democrática y culta se reconoce cada vez más. Obviamente, no sólo deberán valorarse niveles de aprendizaje elementales, como memorización, sino también los más elevados, de razonamiento propio, juicio crítico y creatividad.

Una buena evaluación no podrá limitarse a los insumos del sistema, si bien no puede olvidarlos, ya que sin ellos no es posible medir la eficiencia del sistema, pero habrá de considerar también los productos, para valorar la eficacia, y los procesos, como parte de una aproximación explicativa que indague en lo relativo a las causas de la situación descrita, con lo que se dará sustento sólido al diseño y la implementación de acciones de mejoramiento.

Con un enfoque integral como el que se propone, la evaluación tomará en cuenta no sólo a los alumnos, sino también a los maestros, los directores y supervisores y las autoridades educativas, así como los diferentes niveles de organización de un sistema complejo como el educativo, en cada uno de los cuales hay una problemática específica, y oportunidades particulares de intervención: el aula, la escuela singular, la zona o distrito escolar, el municipio, la región, provincia o entidad federativa, y el país.

Una buena evaluación, además, deberá caracterizarse

por las cualidades técnicas de cualquier medición buena, que se resumen en la validez y la confiabilidad, en sus diversas facetas: que se mida realmente lo que se pretende medir, y que se haga de manera que los resultados sean comparables en el espacio y en el tiempo.

Subrayo que la evaluación no puede reducirse a la aplicación de pruebas de aprendizaje. Estas son, sin duda, necesarias para la medición de los resultados de la enseñanza, pero la valoración de otros aspectos implica la construcción de indicadores con otro tipo de datos, tomados de otras fuentes, como los sistemas de información estadística o estudios ad hoc.

Es claro que esta es una visión ideal, que en sentido estricto es imposible llevar a la práctica, pues exigiría una cantidad de trabajo y un volumen de recursos enorme para recolectar la información necesaria.

Pero si no se tiene este punto de referencia es fácil caer en el error opuesto, que en la realidad se ha dado en muchos lugares: el de reducir el sistema de evaluación a los elementos más fácilmente disponibles, con lo que sólo se dispone de datos gruesos y burdos sobre el número de alumnos inscritos en el sistema al inicio y al fin de un ciclo, el de maestros y escuelas, el presupuesto asignado a las escuelas y poco más.

Por ello creo que es importante manejar una visión ideal de lo que debería ser la evaluación como referente para el desarrollo de un sistema concreto; con tal punto de referencia, y con ayuda de las metodologías de investigación y evaluación educativa desarrolladas en las últimas décadas, es factible diseñar sistemas mucho más finos que los habituales, con requerimientos de recursos al alcance de los sistemas educativos reales.

No está de más precisar que diversos tipos de evaluación son competencia de diversas instancias. En particular conviene distinguir la evaluación de individuos de la de los sistemas como tales. La evaluación de los alumnos en lo individual, por ejemplo, es y sin duda seguirá correspondiendo a los maestros y las escuelas en que están los alumnos; la evaluación de los maestros en lo individual corresponde y corresponderá seguramente a los directores y supervisores.

La evaluación a la que me refiero en este escrito es, precisamente, la del sistema educativo en cuanto tal. Se trata, pues, de una evaluación de nivel macro, que no sustituye a las evaluaciones individuales, sino que las complementa, iluminando un ángulo diferente de la realidad educativa.

Tras lo anterior, las respuestas a las siguientes preguntas serán más breves.

¿Qué estamos evaluando?

Resumo mi respuesta, en forma algo simplista y cruda, diciendo que evaluamos algunas de las dimensiones de la calidad educativa, pero dejando fuera varias de las más importantes, que lo hacemos de manera imprecisa, y que no llegamos a las causas.

Hasta hace muy pocos años, la situación de la mayor parte de los sistemas de evaluación educativa de los países de la región era justamente la que se ha mencionado antes: solamente se contaba con datos sobre alumnos, maestros, escuelas y presupuestos, con lo que se podían estimar indicadores de cobertura, eficiencia terminal y costo alumno en forma gruesa, ya que por lo general no se controlaba la calidad de los procesos de recolección de información, no se manejaban datos por edad, con lo que sólo podían construir-

se tasas brutas y no netas y se manejaban sólo cohortes aparentes.

Los esfuerzos que han hecho muchos de nuestros países en la última década han incluido, por una parte, mejoras en los sistemas de información y, por otra, el desarrollo de sistemas de pruebas nacionales para valorar los aprendizajes de los alumnos. Estos esfuerzos son importantes, sin duda, pero aún están por consolidarse y adolecen de fallas claras, explicables por la falta de una tradición en el campo y la consecuente escasez de recursos humanos especializados.

Pecando tal vez de simplismo, creo que en la región atendemos la cobertura y la eficiencia terminal en forma imperfecta y el nivel de aprendizaje de los alumnos en áreas curriculares y grados clave, pero por lo general limitándose a niveles cognitivos bajos. La contraparte obvia, lo que no estamos evaluando, incluye áreas curriculares adicionales, niveles cognitivos superiores, aspectos actitudinales y valorales, procesos tanto pedagógicos como de gestión y lo relativo a equidad.

En las escuelas se evalúan muchas de estas cosas, desde luego, pero no parece haber procedimientos para valorar tales dimensiones o aspectos de la calidad de manera permanente y sistemática en el nivel nacional. Debe añadirse que muchas veces las evaluaciones que hacemos no tienen, en un grado suficiente, las características técnicas mencionadas de validez y confiabilidad.

Por lo que se refiere a validez, la ausencia de controles de las variables relevantes del contexto pueden hacer que midamos en realidad pobreza, cuando creemos medir calidad de las escuelas. En este sentido en la región es muy raro todavía el uso de las poderosas técnicas actuales, como los modelos jerárquicos lineales, que permiten identificar la influen-

cia de los factores del entorno distinguiéndolos, por ejemplo, de los de la escuela, el aula y el propio alumno.

En cuanto a confiabilidad, es también raro el uso de metodologías psicométricas avanzadas, como los modelos de la teoría de la respuesta al reactivo, que permiten contar con escalas más precisas, comparables en el tiempo, que cuidan la unidimensionalidad y evitan el sesgo de manera más adecuada que las técnicas de la teoría clásica de las pruebas. Debe añadirse que, hasta hace poco tiempo, no era raro que se manejaran pruebas que no empleaban en forma rigurosa siquiera los principios de la teoría clásica. Las técnicas de valor agregado son también muy poco conocidas.

¿Qué de lo evaluado resulta significativo?

En principio los mecanismos de evaluación existentes son todos significativos y, aun en ausencia de los que se sugiere añadir, contienen potencialmente elementos útiles para el mejoramiento de los sistemas educativos. Sin embargo, es frecuente que aún esos elementos se desaprovechen y sólo sirvan para acumular datos en anuarios estadísticos destinados a empolvarse en los estantes, sin que los responsables de tomar decisiones, y mucho menos los maestros, los conozcan y utilicen para retroalimentar su quehacer.

Esto quiere decir que, además de mejorar los actuales mecanismos de evaluación y de enriquecerlos con nuevos elementos, es necesario un importante esfuerzo de difusión de resultados y de capacitación de los usuarios para que estén en condiciones de entender y aprovechar los resultados de las evaluaciones.

¿Por qué las reformas realizadas en los 90 no se ven traducidas en los resultados?

Esta pregunta parte de un supuesto que no me parece incuestionable: que las reformas educativas implementadas en América Latina en los '90 no han dado resultados.

Los niveles educativos de nuestros países son el resultado de un enorme conjunto de factores, incluyendo unos tan complejos como el explosivo incremento demográfico que hizo del subcontinente la región de mayor crecimiento poblacional en toda la historia hasta mediados del siglo XX, la pobreza que aflige a gran parte de la población y la diversidad étnica y lingüística. Otros factores igualmente complejos incluyen el carácter autoritario de muchos regímenes políticos, las estructuras sindicales y burocráticas, y la desigualdad ancestral de nuestras sociedades.

Los cambios educativos, por otra parte, son complejos también y, en el mejor de los casos, llevan mucho tiempo; no pueden ser el resultado rápido de esfuerzos simples.

Por ello, aun si en verdad las reformas de los años 90 no han dado resultados, podrían encontrarse diversas explicaciones. Las políticas educativas de la región han adolecido, en muchos casos, de serias deficiencias en diversos aspectos: en el diagnóstico en que se sustentan, muchas veces incompleto e impreciso; en el diseño de los programas, muchas veces simplista, apostando demasiado a pocos elementos de los que se espera más de lo que pueden dar; y, tal vez, sobre todo, en la implementación.

No es fácil, en efecto, que las reformas lleguen hasta el nivel más importante, el de la escuela y el aula. Para ello se requiere de tiempo y, además, de estrategias muy finas que aseguren la participación de los maestros. Desgraciadamente, es más frecuente encontrar estrategias simplistas e implementaciones apresuradas que procesos cuidadosos y respetuosos de los actores de base y de la complejidad de la cuestión.

No debería sorprender, por consiguiente, que los resultados obtenidos disten mucho de los esperados.

Pero el punto que quiero destacar es que no tenemos evidencia suficiente para afirmar de manera contundente que las reformas de los '90 no han tenido resultados. Hay, sin duda, indicios importantes en tal sentido, pero los sistemas de evaluación de que disponemos no sustentan afirmaciones tajantes.

En particular, creo que en México hay elementos para sostener que el nivel de aprendizaje ha mejorado ligeramente en los últimos años; que el impacto del gasto público en educación básica está siendo mayor para los tres deciles más pobres de la población, contra lo que ocurría antes; y que los programas compensatorios emprendidos a principios de los 90 han logrado reducir la deserción en medio rural y mejorar los niveles educativos en las zonas más desfavorecidas.

Por lo demás, la gran desigualdad que caracteriza a nuestras sociedades hace que necesitemos evaluaciones más precisas, que tengan en cuenta el peso de los factores del entorno, tanto en razón del principio de equidad, como para valorar en su justa medida los avances que puedan darse en contextos diferentes, en los que la mejora es más o menos difícil.

Conclusión

La última pregunta es, a mi juicio, la más importante de las cuatro que he tratado de responder, tanto por lo que expresa como por lo que sugiere de manera implícita. Ampliando lo dicho hasta ahora trataré de aportar algunas ideas al respecto.

La experiencia de las reformas educativas implementadas en los países de mayor desarrollo económico ha dejado lecciones importantes, en la dirección ya apuntada: no deben esperarse cambios espectaculares; las mejoras educativas se consiguen en lapsos de tiempo largos, y mediante esfuerzos sistemáticos y complejos, no gracias a recetas mágicas simples o panaceas.

Hoy sabemos también que no debe exagerarse la importancia de los recursos materiales; estos son importantes, sin duda, sobre todo hasta ciertos umbrales mínimos, y tal vez más en cuanto a la regularidad o continuidad de su existencia que en cuanto a su monto absoluto. Pero parece claro que tanto y más importan los procesos a través de los que se utilizan los recursos: procesos pedagógicos especialmente en el interior del aula, procesos más amplios de gestión en el nivel nacional y regional y, sobre todo, en cada escuela, donde el trabajo del equipo docente, el liderazgo del director y la participación de los padres de familia han mostrado ser fundamentales.

Para poder responder con bases sólidas a la pregunta sobre el éxito o el fracaso de las reformas educativas necesitamos, pues, en mi opinión, contar con sistemas de evaluación bastante más sólidos, aunque no mucho más costosos que los existentes, que nos den la información clave, de que hoy carecemos, sobre las diferentes dimensiones de la calidad educativa de nuestros países en forma continuada y permanente.

Es posible que nos llevemos algunas sorpresas agradables, en el sentido de que los esfuerzos realizados hasta ahora han dado algunos resultados, no espectaculares, pero sí razonables, en una perspectiva histórica y comparativa en el plano internacional.

También es posible que se confirme la idea de que no ha habido resultados positivos. En este caso, si tenemos buenos sistemas de evaluación tendremos también elementos para saber por qué ha ocurrido tal cosa, y para diseñar programas de mejora más adecuados.

La evaluación y las actitudes de los docentes frente a ella: dificultades y alternativas de política

Juan Carlos Navarro²⁸

Introducción

Se ha convertido en un tópico de los planteamientos de reforma educativa en toda América Latina el reconocimiento de la importancia de contar con los docentes como pieza clave en la ejecución de programas innovadores y políticas sistémicas dirigidas al mejoramiento de la educación. La mayor parte de las reformas importantes de la década pasada hicieron un esfuerzo por “traer a bordo” a los docentes a la hora de emprender cambios significativos en la organización, los contenidos, el financiamiento o las metodologías para promover el aprendizaje. Mecanismos para la incorporación del punto de vista de los docentes y sus organizaciones gremiales en el diseño de políticas, recuperación de los salarios severamente disminuidos durante la década de los años ‘80, mejoramiento de los recursos con que cuentan los maestros en el aula, y un renovado énfasis en el reconocimiento público a la importancia del rol de maestro en la sociedad han

²⁸ Las opiniones expresadas en este trabajo son de la responsabilidad exclusiva del autor y no pretenden representar la posición oficial del Banco Interamericano de Desarrollo.

sido aplicados en diversas combinaciones en varios de los países de la región.

Evaluar sistemáticamente estos intentos por lograr que los docentes se apropien de las reformas va mucho más allá del alcance de este trabajo. Cabe simplemente apuntar que, aunque el énfasis en centrar los esfuerzos en mejorar el trabajo docente y su efectividad ha sido en general bienvenido como una parte esencial de las reformas, y de hecho parece encontrarse todavía en ascenso en el marco de las políticas de prácticamente todos los Ministerios de Educación, al mismo tiempo se constata cierta decepción en la capacidad que han tenido las reformas para llegar al aula y para afectar positivamente el trabajo cotidiano de los docentes y sus resultados. Tampoco puede dejarse de mencionar que frecuentemente importantes reformas se han topado con la desconfianza, cuando no resistencia organizada, por parte de los docentes y sus organizaciones, lo que puede verse, hasta cierto punto, como un elemento más del carácter participativo del avance de las reformas, pero al mismo tiempo como el síntoma de las limitaciones que parecen persistir en la capacidad de las políticas educativas para encontrar simpatía y apoyo activo entre aquellos sin cuya participación esas mismas políticas están condenadas a tener efectos menores o nulos. El tema abierto, si la caracterización anterior refleja la discusión sobre reformas educativas en la región, es así: qué puede hacerse para entender mejor las reacciones de los docentes a las reformas, y para conceptualizar éstas de forma que mejore su capacidad para encontrar reconocimiento, despertar entusiasmo y generar acciones constructivas en el cuerpo docente.

Pocos ámbitos más fértiles para el examen de este asunto que el de la evaluación de la calidad educativa. Du-

rante la década pasada varios países de la región introdujeron sistemas de exámenes estandarizados para la medición de los resultados de aprendizaje de sus escuelas, al punto que hoy casi ningún país carece de al menos alguna experiencia en este campo y varios han acumulado trayectorias prolongadas que los ponen en posesión de sistemas sofisticados y reconocidos desde el punto de vista técnico. Simultáneamente, el auge en tiempos recientes de pruebas internacionales comparativas ha contribuido, por un lado, a estimular el desarrollo de capacidades técnicas e institucionales en los países, y, por otro, a elevar el perfil de la discusión nacional y regional sobre los resultados de este tipo de pruebas. El hecho de que los resultados muestran consistentemente considerables problemas en términos de cómo se comparan los logros de aprendizaje en los sistemas de la región cuando se los compara con los de otras partes del mundo, ha agregado un ingrediente polémico a la discusión.

¿Dónde están los docentes en esta discusión? El propósito central de este trabajo es abordar el tema desde el punto de vista, sin duda limitado, que ofrecen algunos de los resultados de un proyecto de investigación comparativa sobre la carrera docente en América Latina promovido hace unos tres años por la Red de Centros de Investigación del Banco Interamericano de **Desarrollo**²⁹. El proyecto en cuestión se propuso observar y comparar las características, percepciones y conductas de los maestros en diversos ambientes institucionales, con el fin de investigar si los cambios

²⁹ Los resultados del estudio están disponibles en el volumen editado por este autor: *Qué piensan los maestros: Carreras e incentivos docentes en América Latina*. Red de Centros de Investigación en Economía Aplicada, Banco Interamericano de Desarrollo, 2002.

introducidos en los incentivos presentes en esos ambientes tenían un impacto en la efectividad de los docentes y las escuelas en las que **trabajan**³⁰. Para asegurar variabilidad en los ambientes institucionales, por un lado, se realizó el estudio en varios países –siete en total– y, por otro, dentro de cada país se hizo lo posible por captar la variabilidad institucional en lo interno del sistema educativo, distinguiendo entre escuelas públicas y privadas, y allí donde fue posible entre diferentes tipos de públicas –escuelas federales y municipales, por ejemplo– y diferentes tipos de privadas. Los países incluidos en el estudio fueron Argentina, Chile, Guatemala, Perú, República Dominicana, Uruguay y Venezuela.

Como parte de la investigación requerida para avanzar en el objetivo central de este proyecto, se recogió, por un lado, información acerca de las actitudes y opiniones de maestros en ejercicio de los países mencionados respecto de cuán apropiado consideraban que su labor docente fuera objeto de evaluaciones. Por otro lado, cada estudio nacional investigó hasta qué punto existían experiencias en materia de introducir sistemas de evaluación docente y qué tanto estos tenían vinculación con evaluaciones de la calidad de los aprendizajes. A estos resultados parciales nos referiremos a continuación, luego de exponer algunos elementos conceptuales necesarios para encuadrar la discusión en su justa dimensión.

³⁰ El término “incentivo” se usó en el estudio, como aquí, en un sentido amplio, para referirse a cualquier estímulo que los maestros experimenten para conducir su trabajo de una cierta forma o tomar decisiones profesionales. En lo absoluto el término se utiliza en el sentido estricto de pagos o recompensas pecuniarias a individuos. Las expresiones “institución” o “ambiente institucional”, por su parte, designan indistintamente conjuntos o “paquetes” de incentivos que caracterizan a un particular sistema escolar o grupo de escuelas que responden al mismo tipo de autoridad o principio de organización.

Evaluación de calidad y evaluación del desempeño docente

El avance en la implementación de sistemas de evaluación de la calidad de la educación en la mayor parte de los países de América Latina que se registraba al principio de este trabajo parece indicar que, o bien no ha sido activamente antagonizado por los docentes y sus organizaciones, o bien, si lo ha sido, este tipo de oposición resultó particularmente poco efectiva. En general podría decirse que, independientemente de las complejidades de la economía política de la educación en los diversos contextos nacionales, la actitud docente no fue un obstáculo de consideración para la instauración de exámenes estandarizados de medición de los resultados de aprendizaje.

Por otra parte, es ampliamente conocido que ha habido considerable desconfianza por parte de muchos docentes frente al hecho de la introducción de los sistemas de evaluación. También prevalece un sentido general de insatisfacción en cuanto a la capacidad que los sistemas educativos han tenido para utilizar los resultados de estas evaluaciones de manera efectiva en, por un lado, orientar políticas y reformas nacionales, pero, muy especialmente, para llegar a la escuela y el aula de forma constructiva y práctica.

Es necesario mencionar que sí se ha registrado oposición a la mayoría de los intentos por introducir sistemas de evaluación de los docentes o su trabajo, incluyendo aquellos que han propuesto una vinculación entre estos sistemas y los sistemas de evaluación de calidad ya existentes. Esto último es tanto más importante cuanto que parece del más puro sentido común el pretender que, una vez que se tienen en la mano registros sistemáticos y confiables de la calidad de la edu-

cación en un sistema –muchas veces desagregado a nivel de región, provincia, sector rural o urbano, sector público o privado, municipio, escuela o cualquier otra subdivisión relevante–, una forma de tomar decisiones a partir de ellos es responsabilizar a las escuelas y los docentes que han intervenido en la producción de los resultados de aprendizajes observados. Si a esto se suman las dificultades bien conocidas para observar la productividad del trabajo docente, dado lo problemático de observar las prácticas de aula y la escasez de otros indicadores de tal productividad, se entiende fácilmente que la tentación para dar el salto entre evaluaciones de calidad y evaluaciones de la docencia y de los docentes tiende a ser fuerte.

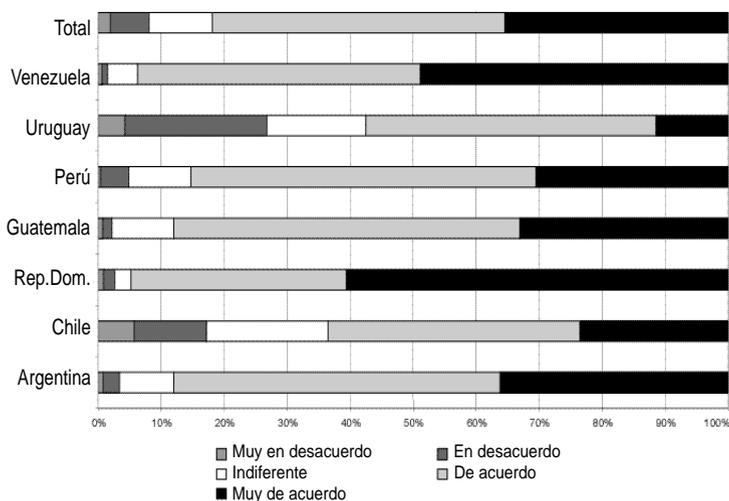
Este salto, por lo demás, no se produce solamente en las páginas de los análisis de políticas, sino, de hecho, tiene lugar en las políticas educativas que reaccionan a los resultados de las evaluaciones de calidad, y, de forma tanto o más importante, en la actitud del público en general respecto de la responsabilidad de los docentes por el lamentable estado de la educación nacional. Y es a cualquier indicación de que las autoridades educativas de un país están cediendo a una tentación como ésta, a la sola posibilidad de que cedan o de que la discusión pública se vuelva en su contra, que los maestros se han opuesto con fuerza.

Actitudes frente a la evaluación del desempeño docente

Un hecho sólido, documentado en el estudio de referencia, es que cualquiera que sea la intensidad de esta hostilidad, no parece que pueda atribuirse a una actitud de los docentes de

oposición a la idea misma de la evaluación. Así, el estudio contenía un ítem, en el marco de una escala actitudinal más amplia, que interrogaba especialmente a los maestros acerca de qué tan de acuerdo estaban con la proposición “Es justo que se evalúe mi desempeño”. Las respuestas a este ítem de la encuesta pasada a más de cinco mil docentes de educación básica incluidos en el estudio muestran un respaldo ampliamente mayoritario a la proposición en cuestión. En el [Gráfico 1](#) se encuentra un resumen de la información correspondiente, discriminada según país.

Gráfico 1
Actitudes respecto a la evaluación del desempeño docente



Fuente: El desempeño de los maestros en América Latina: Carreras e incentivos. BID, 2002.

Tal como puede observarse, en todos los países las respuestas aprobatorias de la proposición presentada son claramente mayoritarias, con las alternativas “Muy en desacuerdo” y “En desacuerdo” alcanzando alrededor del diez por ciento de la muestra global. Solamente en Uruguay y en Chile se encuentra que la suma de estas categorías alcanza el 30 por ciento de los maestros entrevistados. A juzgar por este elemento de juicio, difícilmente puede argumentarse que la noción de evaluación del desempeño docente es enérgicamente rechazada por los profesionales de la docencia. Muy al contrario, parece haber una considerable apertura a la idea o, si se quiere, al principio según el cual la evaluación del desempeño docente es algo justo y apropiado.

Más allá de este hecho sólido, la información recolectada en el proyecto sobre carreras docentes permite caracterizar de forma algo más específica la aceptación de los docentes a la idea de evaluación, proporcionando el fundamento para afirmaciones como las siguientes:

- La aceptación de la idea de evaluación del desempeño docente crece a medida que es más fuerte su acuerdo con las condiciones en las que el docente labora y también en la medida en que se eleva su satisfacción con la profesión docente misma como práctica profesional.
- Los maestros cuyas prácticas y actitudes se acercan más a aquellas normalmente identificadas en la literatura como conducentes a una enseñanza más efectiva, tienden a aceptar más la idea de que su desempeño debe ser evaluado. Esto es particularmente claro en el caso de la estrecha asociación entre aceptar la evaluación y la existencia de un clima normativo fuerte en la escuela en que trabajan los maestros, y también en el caso de la relación di-

recta entre maestros que registran actitudes fuertes de “ciudadanía **organizacional**”³¹ y maestros que aceptan la evaluación.

- La aceptación de la evaluación, aunque muy fuerte entre todas las cohortes, tiende a ser significativamente más fuerte entre los docentes más jóvenes, de 15 o menos años de experiencia. También tiende a ser más fuerte entre los docentes con mayor número de años de educación formal.

Hallazgos como estos apuntan a algunas lecciones de interés: en primer lugar, la actitud favorable hacia la evaluación puede ser fomentada a través de intervenciones que mejoren la satisfacción del docente con su ambiente de trabajo y con la profesión. Estas son claramente variables sujetas a la intervención de las autoridades educativas, y que además han evolucionado favorablemente en buena parte de los países de América Latina en tiempos recientes, lo que de por sí puede contribuir a explicar el importante grado de aceptación de la evaluación entre los maestros. En segundo lugar, la asociación entre prácticas educativas efectivas y aceptación de la evaluación del desempeño sugiere que en la medida en que las escuelas ganan en efectividad y competencia para producir mejores aprendizajes –algo que, por supuesto, es el desideratum de la política educativa, antes que un resultado de intervenciones específicas, como en el caso de la mejora del

³¹ El concepto de “ciudadanía organizacional”, adaptado de la psicología laboral, se refiere en este contexto a qué tanto los docentes tienen un sentido de la obligación y deber en el trabajo que va más allá de la definición explícita de sus tareas cotidianas. Se trata de hasta qué punto el docente está dispuesto a “poner un esfuerzo extra”, y es receptivo a pedidos de que así lo haga que venga, por ejemplo, de parte de colegas o del director de la escuela, con tal de apuntalar el éxito de la organización en que trabaja.

ambiente de trabajo del docente–, debería esperarse un campo más fértil para la introducción de mecanismos de evaluación del trabajo docente. Finalmente, la mayor apertura de las nuevas generaciones de maestros y su mayor nivel de logro educativo promedio permiten abrigar también cierto optimismo a la hora de sopesar las perspectivas de la evaluación en el futuro mediano.

Ahora bien, es bueno recordar al alcanzar este punto que hasta aquí la discusión ha tenido que ver con la aceptación por parte de los docentes del principio de la evaluación del desempeño, no de su práctica específica en el contexto de las instituciones educativas en las que les toca trabajar. Resultó prácticamente imposible en el contexto de los sistemas escolares en los que se realizó la encuesta explorar este tema de forma similar al punto general de la aceptación de la evaluación. La principal razón de esta imposibilidad, sin embargo, quedó registrada sistemáticamente en los estudios nacionales, y resulta ampliamente instructiva.

La problemática puesta en práctica del principio de evaluación del desempeño

Así, se desprenden del conjunto de los casos estudiados un par de hechos relevantes para esta discusión:

- El principio de evaluación del desempeño docente a través de programas o sistemas especializados para la producción de tal tipo de evaluaciones está incorporado en principio, si bien muy modestamente –esto es, no como un principio central de la evaluación docente, sino como uno de varios y con importancia secundaria–, en las políticas

educativas de al menos 4 de los 7 países estudiados. Téngase en cuenta que en 3 de estos casos se ha intentado de una forma u otra establecer una vinculación entre evaluación del desempeño docente y sistemas de medición de la calidad de los aprendizajes de los alumnos.

- La puesta en práctica de este principio ha sido muy problemática y nada exitosa.

En otras palabras, aunque existe un reconocimiento formal y teórico de la necesidad de evaluar docentes y de incluir el desempeño como un elemento menor de una serie de otros criterios necesarios para tal evaluación, ha sido muy difícil –por no decir imposible– poner a funcionar sistemas que representen una expresión competente y sostenible en el tiempo de tal reconocimiento.

Hay de hecho una lista de casos en los que esto puede documentarse con claridad:

- El estudio sobre el Perú muestra que una escala formal de criterios de evaluación para docentes, vigente aunque implementada en la práctica sólo en una oportunidad, otorga al desempeño un máximo de 60 puntos en un total de 200 posibles a la hora de evaluar a un docente. Ha sido imposible acordar los criterios técnicos y obtener el acuerdo político para poner a funcionar esta escala sobre una base regular.
- La “Tabla de Valoración de Méritos”, en Venezuela, incorpora un “certificado anual de eficiencia” para cada docente, que puede llegar a pesar entre 5 y 11 por ciento de la evaluación que se utilizaría para los ascensos, y durante largo tiempo ha permanecido sin efecto por ausencia de la necesaria reglamentación.

- En República Dominicana los aumentos del salario del docente podrían deberse hasta en un tercio a una evaluación del desempeño, pero tal aproximación a la remuneración docente no ha podido ser implementada en forma efectiva, dado el intento inicial de hacer depender la evaluación de los maestros de los directores, mientras al mismo tiempo se hacía depender la evaluación de los directores de la opinión de los maestros, lo que en la práctica restó al sistema su credibilidad y efectividad desde su primer intento de implantación.

En la práctica, como puede constatarse, el destino de iniciativas como ésta no ha sido muy feliz. Los estudios nacionales describen casos en los que se instituyen primas o bonificaciones para todos los docentes y se las llama incentivos, o pagos por desempeño, si bien forman en la práctica parte de la remuneración básica de todo docente. También hay varias experiencias en las que se instituyen pagos que inicialmente tienen la intención de actuar como estímulos o reconocimiento a evaluaciones del desempeño, pero terminan por convertirse en derechos adquiridos de todos los maestros e incorporarse a los pagos básicos. Todo esto ocurre como resultado de la baja capacidad de los sistemas educativos para aplicar de forma ordenada y profesional los principios evaluativos que formalmente profesan. En el primer caso se llama incentivo a algo que es una prima salarial general. En el segundo se ponen de manifiesto las limitaciones de los sistemas de administración de carreras docentes para manejar mecanismos de compensación más allá de los más simples e indiferenciados, desprovistos de una intervención de criterios de evaluación de desempeño docente.

Excepción hecha del SNED introducido hace pocos

años en Chile, y sobre el que se regresará en breve, se puede generalizar y decir que todos los casos nacionales incluidos en el proyecto sobre carreras docentes oscilan entre tener un sistema de evaluación docente débil y que da muy poca importancia al desempeño, y el no tener un sistema de evaluación en lo absoluto.

La contribución de esta revisión de la vida real de los sistemas de evaluación del desempeño docente a la discusión adelantada en este trabajo es clara: poner en práctica el principio de la evaluación es mucho más complejo y difícil que aceptar el principio en teoría. ¿Cuál es la fuente de esta dificultad?

¿Resistencias políticas o características inherentes a la profesión?

Una primera respuesta podría apuntar a la defensa, por parte de las organizaciones sindicales docentes, de derechos adquiridos e intereses creados a favor de un ejercicio profesional carente de rendición de cuentas o presiones hacia la productividad. Poca duda puede haber de que la dinámica política de la resistencia a la evaluación docente toma esta forma a primera vista en muchos casos. Las experiencias revisadas en la sección precedente, sin embargo, sugieren hacerse por lo menos la pregunta de si el fracaso de los intentos por introducir evaluaciones del desempeño docente pueden atribuirse a las presiones sindicales y a la oposición de los docentes, o si estas últimas son más bien el producto de intentos fracasados y poco convincentes desde el punto de vista técnico y gerencial de implantar el principio de evaluación.

Así, bien puede construirse el argumento de que la

complejidad contractual de la profesión docente crea problemas muy serios a cualquier tentativa de introducir sistemas de evaluación del desempeño docente. Es bien sabido que una gran cantidad de factores contextuales –apoyo familiar, nutrición, exposición a fuentes extraescolares de información– influyen poderosamente en que el aprendizaje de un estudiante en particular sea diferente al de otro. Por otro lado, típicamente varios maestros trabajan sobre un estudiante: ¿cómo saber si el buen desempeño de un niño en un cierto grado se debe a la excelente maestra que tuvo ese año o a la excelente preparación provista por los maestros de los grados anteriores? Adicionalmente, se tiende a pensar que la colaboración y el trabajo en equipo de los maestros en lugar de relaciones competitivas entre los mismos constituyen una mejor contribución al aprendizaje. Después de todo esto, ¿cómo separar el esfuerzo individual?

La separación del resultado del esfuerzo individual, sin embargo, no es más que uno de los varios problemas que tiene el diseño de contratos eficientes entre un sistema escolar y un maestro. De gran importancia son también las dificultades para la observación del esfuerzo y para la medición del producto. Una vez cerradas las puertas del aula, ¿cómo saber si el maestro está haciendo todo lo que puede y debe? Más aún, no existe consenso respecto de que todos los maestros deberían hacer exactamente lo mismo, y si bien hay acuerdos generales en cuanto a las ventajas de ciertos tipos de pedagogía activas y constructivistas, es bastante general también la aceptación de que existen diversas fortalezas en distintos maestros y distintas necesidades en distintos tipos de estudiantes que pueden justificar formas muy diversas de adelantar una clase. Por otra parte, la controversia acerca de qué miden exactamente las pruebas de evaluación de apren-

dizajes, acerca de hasta qué punto miden lo más importante o miden correctamente, es permanente, a pesar de los considerables avances técnicos en la materia: como mínimo, hay que decir que es de aceptación general que varios de los productos de la enseñanza efectiva no quedan capturados en las pruebas estandarizadas de desempeño.

En la práctica, en fin, mucha información acerca de la calidad de la actividad docente es lo que se conoce como información de tiempo y lugar. En otras palabras, en cada escuela circula abundante información acerca de quién es quién: qué docente es exigente y cuál no, quién domina su materia y quién no, quién trata mejor o peor a sus alumnos, quién tiene una personalidad más atrayente que otros, quién es más dedicado, ameno, cumplido, etc. No hay más que recordar nuestra experiencia personal en la escuela para saber que esta información existe y circula. Sin embargo, es prácticamente imposible que esta información circule a escala masiva, que pueda reducirse a una serie de indicadores o variables registrables en un sistema de información a disposición del Ministerio u otra autoridad educativa por encima del director de la escuela.

Una consecuencia de todo esto es que escribir contratos eficientes, que maximicen los resultados del trabajo del maestro, es mucho más difícil en el caso de la docencia que lo que parecería común en muchas otras profesiones. Los sistemas de supervisión, inspección y vigilancia son una alternativa a los incentivos por evaluación del desempeño, pero en todas partes han experimentado un deterioro considerable y hoy en día, en la mayor parte de los países, son lo suficientemente imperfectos y poco confiables cuando abarcan a sistemas escolares masivos que resulta difícil hacer depender de ellos algún reconocimiento en la forma de ascensos,

aumentos de salario, premios, etc. Esos sistemas existen en la práctica como una última línea de defensa contra casos extremos, tales como los de maestros que cometen acciones de abuso físico o psicológico de sus estudiantes o simplemente dejan notoriamente de cumplir con obligaciones tan básicas como la de asistir regularmente a clase. No en vano, en la encuesta a la que fueron sometidos los maestros en el estudio comparativo al que se ha aludido en este trabajo, y ante la pregunta acerca de qué instancia del sistema educativo era relevante para su práctica profesional, la supervisión fue completamente ignorada por los interrogados.

El balance de esta rápida revisión de la complejidad contractual de la profesión docente no puede ser ajeno a la explicación de la resistencia de los maestros a tentativas de introducir tanto la evaluación del desempeño como los esfuerzos por combinarlas con la información procedente de las pruebas de medición de la calidad. Dicho más directamente, los docentes tienen poderosas y buenas razones para resistir, a no ser que pueda mostrarse que el sistema o propuesta de evaluación representa un serio intento de tener en cuenta la complejidad contractual de su profesión.

Conclusión: Hacia la construcción de sistemas de evaluación que reflejen la complejidad de la actividad docente

Si se toma en cuenta la discusión precedente, es posible esbozar las líneas generales de un sistema de evaluación del desempeño docente que tome en serio las dificultades propias de semejante empresa. Podría decirse que un sistema tal debería:

- Al evaluar el desempeño de los maestros, introducir el uso de los resultados de las evaluaciones del logro de los alumnos en las escuelas, pero no como puntajes absolutos, sino relativos a su nivel del período anterior, y en comparación no con un promedio nacional, sino con un conjunto de escuelas identificadas porque tienen características similares en varios aspectos fundamentales, en particular en la composición socioeconómica del alumnado.
- No distribuir incentivos a un maestro aislado identificado como particularmente esforzado o exitoso, sino a escuelas, premiando el desempeño de equipos docentes y directivos, antes que el de maestros individuales.
- Para evitar crear incentivos contraproducentes a las escuelas, valorar para la puntuación de cada establecimiento el que no tenga prácticas discriminatorias en la selección de estudiantes que entran al mismo.

El estudio sobre la carrera de maestros en Chile en el proyecto de referencia muestra como, en el caso del SNED, se ha hecho un intento convincente de tomar en cuenta estos principios. Es significativo que la información especialmente rica disponible acerca de la actitud de los maestros chilenos antes y después de la introducción del SNED sugiera que, cuando los maestros están expuestos a un sistema de evaluación competentemente adaptado a las características propias de la actividad docente, reaccionan favorablemente, según lo indica el crecimiento de la aceptación entre ellos de las prácticas de evaluación. Otras de las pocas experiencias exitosas en introducir elementos de evaluación del desempeño docente, tal el caso de una reciente iniciativa en El Salvador, han sido diseñadas de forma que se acercan al cumplimiento de las características esbozadas. La construcción de sistemas todavía

mas refinados que apuntan a la evaluación docente, como el reciente Programa de Acreditación para la Asignación de Excelencia Docente, también en Chile, es probablemente una avenida disponible para desarrollos futuros en esta materia.

Por lo pronto vale concluir, recapitulando, que si bien se registra resistencia de los docentes a introducir prácticas de evaluación de su desempeño y muy en particular al uso de resultados de las pruebas de evaluación de la calidad en este contexto, hay evidencia de que esta resistencia no obedece a una oposición de principio a la introducción de la evaluación, sino a la desconfianza provocada por la dificultad natural de evaluar la práctica docente con criterios relativamente más simples aplicables a otras profesiones, y alimentada por una historia poco feliz de intentos fallidos de implementar sistemas de evaluación en varios países de la región. También cabe destacar que existe evidencia preliminar que apunta a que la resistencia de los docentes puede modificarse tanto gracias a medidas que mejoren el ambiente escolar y la satisfacción con la profesión, como también mediante la introducción de sistemas de evaluación que encarnen una respuesta eficaz a los problemas contractuales que caracterizan a la docencia como actividad.

Los docentes y la evaluación

Emilio Tenti **Fanfani**³²

Introducción

El diccionario ofrece dos términos para designar dos fenómenos distintos. Uno es evaluar, el otro es examinar. A estas dos acciones se asocian dos sustantivos: el examen y la evaluación. Evaluar es una acción relacionada con el hecho de dar un valor o de apreciar. El examen, según el diccionario de Pedagogía de Paul Foulquié (México 1980), es “la acción de observar o de estudiar con atención”. Mientras que en el primer caso se trata de formular un juicio de valor (en términos de bueno o de malo, de adecuado o de inadecuado, de suficiente o insuficiente) en el segundo caso se trata de una simple acción de medición o diferenciación. Pero el examen tiene una acepción particular en el “lenguaje de la pedagogía”. En efecto, Foulquié lo define como aquel “conjunto de observaciones o de pruebas destinadas a apreciar el saber o la cultura de los candidatos con vistas a la concesión de un diploma, para entrar en una escuela, o la aptitud para una función determinada”. Como puede observarse, la definición de examen engloba la de evaluación (entendida como apreciación).

³² El autor agradece la colaboración de Leonardo Tenti Vuegen en la construcción de variables y en el manejo de las bases de datos.

La historia enseña que el examen y la evaluación fueron herramientas sistemáticamente utilizadas por la escuela moderna (es decir, la escuela que surge y crece con el Estado capitalista moderno). No hay institución que examine (en el doble sentido de observar, medir y también valorar) más intensivamente que la escuela. Y esto, al menos, por dos razones fundamentales:

a) La primera es estructural y se asocia con el carácter graduado que tiene la apropiación del conocimiento. La cuestión es relativamente simple: hay ciertas cosas que hay que aprender primero para aprender otras (generalmente más complejas) después. La racionalización de la pedagogía institucionalizó determinadas técnicas de examen para controlar la progresión del aprendizaje de los alumnos.

b) La segunda tiene que ver con un dato fundamental, característico del sistema escolar moderno, por lo general controlado por el Estado. Esta escuela no sólo se propone desarrollar conocimientos en las personas, sino que también distribuye credenciales, es decir, documentos públicamente reconocidos (los famosos títulos) que certifican que el poseedor de los mismos ha incorporado determinados conocimientos, valores y competencias. El examen es un requisito de los títulos o certificados públicos (garantizados por el Estado) que distribuyen las instituciones escolares (sean de carácter público o privado).

Durante los últimos años, por una serie de circunstancias complejas, los sistemas educativos (es decir, los docentes y las instituciones) ya no son sólo los sujetos activos de la evaluación, sino que ahora son también objeto de evaluación. ¿Qué es lo que explica que esta vieja institución que desde

siempre hace un uso intensivo de la evaluación y el examen, ahora se haya convertido en objeto de observación y valoración? ¿Qué es lo que justifica que luego de tantas evaluaciones a las que son sometidos los alumnos desde que entran hasta que salen de determinado nivel educativo sean nuevamente objeto de examen, esta vez por instancias extra escolares (los Ministerios de Educación, Institutos Internacionales, entes autónomos, etc.) generalmente ubicados en el ámbito del servicio público? A primera vista es tentador decir que el sistema educativo es examinado porque existe una sospecha. En otras palabras, la institucionalización generalizada (en todo el mundo desarrollado y en la mayoría de los países de América Latina) de sistemas nacionales de evaluación está indicando que “la sociedad” (ciertos y determinados actores sociales representativos y poderosos) desconfía de lo que hace y produce el sistema educativo (incluso sus propios exámenes). ¿Cómo justificar que luego de seis, siete o nueve años de escolaridad obligatoria los alumnos que han aprobado todos los exámenes y han cumplido todos los requisitos para obtener un diploma sean sometidos a una especie de “examen de Estado”? Todo parece indicar que sobre el sistema educativo planea una sospecha que puede expresarse del siguiente modo: no existe necesariamente una correspondencia entre el conocimiento y las competencias efectivamente desarrolladas en los alumnos y los certificados, diplomas o títulos efectivamente distribuidos por las instituciones escolares. Nunca existió una correspondencia estrecha entre estos dos modos de existencia del “capital cultural” (el incorporado en las personas y el institucionalizado en los títulos de Estado), pero en esta etapa del desarrollo de las sociedades y los sistemas educativos esta distancia pareciera haberse incrementado hasta el punto de despertar una sos-

pecha y una **preocupación**³³. De allí el desarrollo de las políticas e instituciones especializadas en “evaluar la calidad de la educación” efectivamente alcanzada por los alumnos que pasan por la escuela.

En este contexto no debería extrañar que los propios agentes escolares (en proporciones variables, según las circunstancias) se hayan sentido un tanto incómodos frente a este tipo de políticas que ponen un manto de duda sobre lo que ellos hacen y producen cotidianamente en sus aulas. De hecho, en casi todos los países, los programas de evaluación de la calidad no recibieron la aceptación inmediata de los docentes y menos aún de los docentes organizados en las organizaciones sindicales. Por el contrario, en casi toda América Latina las evaluaciones fueron objeto de fuerte resistencia por grupos significativos de docentes. Pero, desde un punto de vista sociológico, este es un hecho que no debe sorprender, dadas las circunstancias y el sentido que tienen la evaluación y el examen en la historia escolar.

En las notas que siguen aportamos algunos elementos de información y análisis relacionados con las actitudes de los docentes con respecto a la cuestión de la evaluación. Para ello, subsumiremos la cuestión en el tema más general de la “diferenciación, ordenamiento, clasificación y jerarquiza-

³³ En verdad hay que reconocer que ni siquiera el propio sistema educativo confía en sus propios productos. En efecto, en muchos casos el título o certificado de estudios que otorga un nivel del sistema no es suficiente para ingresar al nivel siguiente. Para ingresar a la educación superior, por ejemplo, no basta el título de bachiller sino que es preciso aprobar exámenes nacionales de acreditación de nivel y además superar exámenes específicos de admisión administrados por las instituciones del nivel correspondiente. Esta es otra evidencia de que cada nivel educativo tiene su propia racionalidad, la cual no siempre está articulada con el nivel subsiguiente.

ción” como dispositivos que operan en forma implícita y explícita en todos los campos de la vida social y que tienen sus implicaciones políticas y culturales bien precisas. Para realizar este análisis haremos uso de los datos producidos en el contexto de la aplicación de un cuestionario a muestras representativas de docentes de Argentina, Perú y Uruguay³⁴. Este contiene varios ítems relacionados con el tema que aquí nos interesa y los resultados nos permiten obtener un primer panorama acerca de las diversas posiciones que se manifiestan en el cuerpo docente de los países analizados. En primer lugar se analizan algunas orientaciones que los docentes manifiestan hacia la evaluación como un componente estructural de su trabajo en el aula. En un segundo momento se analizan sus opiniones en cuanto a las estrategias y procedimientos que se usan para evaluar su propio trabajo (el docente evaluado). Luego se presentan las actitudes que los docentes manifiestan frente a los sistemas nacionales de evaluación del rendimiento escolar y su eventual uso para evaluar la calidad de su propio trabajo. Por último se discuten algunas posiciones de los docentes frente a la temática general de la igualdad como valor social.

³⁴ El programa de estudios “Profesionalización de los docentes en América Latina” se lleva a cabo en el IIPE - UNESCO, en su sede de Buenos Aires. Al presente se han presentado tres informes nacionales (Argentina, Perú y Uruguay). Las bases de datos fueron construidas a partir de la aplicación de un cuestionario a muestras representativas nacionales de docentes de primaria y secundaria que trabajan en el ámbito urbano (Argentina y Uruguay) y urbano y rural (Perú).

1. El docente como evaluador

Diseñar, aplicar y evaluar exámenes constituye un componente básico del trabajo docente. Es en el examen cuando se manifiesta de un modo más evidente la asimetría estructural que caracteriza a la relación profesor/alumno. La práctica del examen tiene un componente técnico-profesional y su realización exitosa requiere de determinadas competencias por parte del docente.

Cuando se interroga a los docentes acerca del grado de dificultad de algunas tareas que realizan (**Cuadro N° 1**) el evaluar es un problema para poco más de un quinto de los docentes argentinos (21%). Pero es sintomático que “el tiempo disponible para corregir cuadernos, trabajos, etc.” constituya el problema más frecuentemente señalado como tal por los docentes (69% de los argentinos y 36.9% de los peruanos).

Los datos indican que hay otras dimensiones de su trabajo que son problemáticas para porcentajes mayores de docentes, tales como el manejo de la disciplina en clase o la relación con los padres. En el caso de la Argentina, el porcentaje de maestros que señalan como un problema la práctica de evaluación es superior a los que señalan como problemáticos “el dominio de los contenidos” (10%) o bien “organizar el trabajo en clase” (9%).

A su vez, y también en la Argentina, hay varios factores que están asociados al hecho de considerar la evaluación como un problema. Esto es más frecuente entre los profesores de secundaria (24%) que entre los maestros de primaria (19%), y entre los docentes del sector público (22%) que entre sus colegas del sector privado (17%). Es probable que la evaluación presente características más complejas (y eventualmente conflictivas) entre los profesores de secundaria. A su vez, los docentes del sector privado probablemente tengan

reglas más claras e institucionalmente definidas para realizar las tareas de evaluación. Pero estas son meras hipótesis para las cuales no existen evidencias empíricas que las sustenten.

Cuadro N° 1: Situaciones que constituyen un problema en el trabajo diario de los docentes (en %).

	ARGENTINA	PERÚ
Evaluar	21.1	15.3
El tiempo disponible para corregir evaluaciones, cuadernos, etc.	69.1	36.9
Manejar la disciplina en clase	32.9	19.5
La relación con los directivos	7.9	14.2
El trabajo con los colegas	7.7	13.8
El dominio de los nuevos contenidos	9.9	19.1
La falta de definiciones y objetivos claros sobre lo que hay que hacer	54.8	39.4
Las formas de planificar	19.9	30.1
La relación con los padres	28.6	31.7
Las características sociales de los alumnos	50.4	52.2
Organizar el trabajo en clase	8.8	8.1
El tiempo disponible para el desarrollo de las tareas	57.8	38.6
La disponibilidad de ámbitos de asesoramiento y supervisión	53.0	46.6

Otra prueba de que la evaluación es un desafío de cierta relevancia para los trabajadores de la educación lo constituye el hecho de que cuando se les ofrece a los docentes una lista de actividades que ellos preferirían realizar si se les otorgara más horas rentadas (si se les extendiera el tiempo pago para su actividad docente), las tareas de “corrección de exámenes, cuadernos, trabajos, etc.” es elegida por una exi-

gua minoría de docentes: 17% de los docentes argentinos y sólo 6.4% de los peruanos (Ver Cuadro N° 2). Este resultado es sintomático, más aún si se tiene en cuenta que ellos mismos señalan que el tiempo asignado a este tipo de tarea resulta el problema más frecuentemente señalado como tal por los docentes. Es evidente que tomar exámenes, corregir pruebas, revisar tareas, etc. no es una dimensión particularmente interesante para la mayoría de los docentes.

Cuadro N° 2: Actividades que los docentes elegirían hacer si les dieran 10 hs. más rentadas.

ACTIVIDADES	ARGENTINA	PERÚ
Corrección de exámenes, cuadernos, trabajos, etc.	17.0	6.4
Trabajo con alumnos sobre temas que les interesan	45.4	49.7
Actividades de autoperfeccionamiento (leer, indagar, etc)	43.8	41.8
Formación y perfeccionamiento en instancias formales	35.3	11.8
Trabajar con sus colegas	31.6	10.4
Trabajos de investigación	29.5	38.6
Producción de materiales pedagógicos	21.7	40.0
Proyectos de servicio a la comunidad	19.2	27.0
Preparación de clases	19.0	14.4
Enseñanza en el aula	15.6	26.5
Atención y actividades con padres	11.3	19.7

Son los profesores varones quienes menos se encuentran predispuestos a emplear un eventual tiempo extra en tareas de evaluación (solo un 7% en la Argentina). Lo mismo puede decirse de quienes son jefes de hogar (respecto de los que no lo son), aquellos cuyo salario docente representa el 100% del ingreso del hogar, los que se perciben como perteneciente a la clase media alta (en relación con los de media

baja y baja), los de más de 45 años y los que trabajan en el sector público. En Perú se observan otras asociaciones. En efecto, los datos indican que los más jóvenes, los que trabajan en establecimientos públicos, son quienes tienen menos interés en emplear tiempo suplementario para realizar este tipo de tareas asociadas con la evaluación.

Es probable que las fuentes de dificultad que genera la actividad de evaluación sean diversas. Algunas tienen un componente técnico (evaluar requiere el dominio de teorías del aprendizaje, estrategias metodológicas e instrumentos técnicos precisos, etc.). Por otra parte, no cabe duda de que la evaluación tiene un claro componente “político” (supone una clasificación, un ordenamiento, una valoración, una jerarquización, etc.) dotado de alto potencial conflictivo, dadas las consecuencias que tiene en la carrera escolar y en la misma biografía de los alumnos y sus familias. Pero la investigación que comentamos no nos permite ir más allá de estas apreciaciones generales e hipotéticas.

2. El docente como objeto evaluado

Pero el docente no sólo es un “evaluador sistemático y profesional”, sino que también es, por decirlo así, un objeto evaluado, tanto por sus superiores jerárquicos –directores, supervisores, etc.– como por el sistema educativo como tal –Instancias nacionales de evaluación de la calidad docente–. La evaluación de los docentes tiende a instalarse en la agenda de política educativa de la mayoría de los países latinoamericanos. Dadas las implicaciones laborales de esta práctica, es un tema que interesa y preocupa a los líderes y militantes de la mayoría de los sindicatos docentes.

El propio sentido común indica que no es fácil “evaluar al evaluador”. En todos los países existen regulaciones y dispositivos que conforman un sistema de evaluación de los docentes. Esta evaluación determina el lugar que ocupan los docentes en la estructura ocupacional del sistema educativo y por lo tanto contribuye a definir el salario, carrera, etc.

Una clara mayoría de docentes de la Argentina y el Perú se manifiesta disconforme con los mecanismos vigentes de evaluación de su trabajo (Cuadro Nº 3). Esta posición crítica es también mayoritaria en el Uruguay, aunque en este país las posiciones parecieran estar más equilibradas. En la Argentina este desacuerdo es mayor en el resto del país que en la Capital, y entre los del sector público, los de clase media alta y alta, los profesores de media varones y los que son jefes de hogar.

Cuadro Nº 3: Grado de adecuación de los mecanismos vigentes de evaluación de los docentes.

	ARGENTINA	PERÚ	URUGUAY
Adecuación			
Sí	14.7	16.8	42.2
No	60.8	75.6	47.6
Ns/Nc	24.5	7.6	10.2
Total	100.0	100.0	100.0

El cuestionario aplicado preguntó a los docentes que calificaran el grado de pertinencia de determinados criterios para determinar las categorías salariales. Entre estos criterios se incluyó la “evaluación periódica del desempeño profesional”, además de otros factores clásicos tales como el título de grado, la antigüedad docente, y la zona geográfica de desem-

peño. Para la Argentina los resultados indican que “la evaluación” es colocada en tercer lugar de importancia, luego de la antigüedad y del título de grado (Cuadro N° 4) . Este ordenamiento de criterios indica el predominio de una visión tradicional respecto de los factores que determinan el salario docente. Sin embargo, existen algunos factores que están asociados con la probabilidad de valorar más el desempeño. Estos son: ser profesor de media varón, desempeñarse como docente en instituciones privadas y estar ubicado en los estratos sociales medio alto y alto.

Cuadro N° 4: Criterios que deben ser utilizados para determinar categorías salariales de los docentes (calificación de 1 a 10 según la importancia acordada).

CRITERIOS	ARGENTINA	PERÚ	URUGUAY	
Antigüedad docente		8.3	7.0	9.0
Título académico de grado		8.2	7.8	9.7
Evaluación periódica		8.0	7.9	9.5
Otros antecedentes académicos		7.7	6.3	8.7
Zona geográfica de trabajo		7.8	6.8	8.8 (*)

(*) En Uruguay se refiere a “contexto social del establecimiento”.

La evaluación periódica aparece valorada en primer lugar por los docentes del Perú, en segundo lugar por los del Uruguay y en tercer lugar por los argentinos. Sin embargo, puede decirse que en general todos incluyen de alguna manera la evaluación periódica como un criterio importante. El consenso desaparece cuando se trata de valorar los sistemas concretos de medición y valoración de criterios que se emplean en cada uno de los países considerados.

Se preguntó a los docentes cuáles son los agentes más adecuados para llevar a cabo la evaluación de la calidad de su trabajo; las respuestas muestran que no todos gozan del mismo grado de legitimidad. Algunos de estos agentes son más “tradicionales” que otros. Entre los primeros pueden citarse las autoridades de la institución escolar (supervisores, directores, etc.). Más recientes son las propuestas de incluir a los agentes que conforman la demanda del sistema educativo (alumnos, padres de familia, representantes de la comunidad, etc.) en los sistemas de evaluación de docentes. El Cuadro N° 5 presenta algunos datos acerca de la distribución de las opiniones de los docentes encuestados.

Cuadro N° 5: Agentes que deberían participar en la evaluación de los docentes.

AGENTES	ARGENTINA	PERÚ	URUGUAY
Autoridades de la institución (director, superv.)	73.7	45.5 (a)	76.0 (c)
Supervisor o Inspector	no corresp	no corresp.	76.8
Otros expertos y especialistas en educación	48.3	36.1	41.1
Colegas docentes más calificados	41.4	38.3	54.2
Técnicos del Ministerio de Educación de la Nación	30.6	52.3	no corresp.
Técnicos de los ministerios provinciales	30.4	27.4 (b)	no corresp.
Los alumnos de los establecimientos	23.5	17.8	25.6
Padres de familia y la comunidad escolar	20.0	34.4	12.6
Comunidad docente de cada escuela	no corresp.	31.5	no corresp.
Ns/Nc	2.5	0.8	13.4

(a) El equipo directivo del centro educativo.

(b) Los técnicos y expertos de los organismos intermedios (USES, Región).

(c) El director.

Es evidente que las instancias que gozan de la mayor aceptación son las autoridades de las instituciones (directores y supervisores). El resto de las instancias, salvo excepciones, sólo son reconocidas por menos de la mitad de los docentes. Estos datos están indicando que los docentes se ven a sí mismos como funcionarios de una organización que tienen una posición bien determinada en una jerarquía organizacional. De esta manera, quienes tienen la autoridad formal sobre los docentes, poseen la responsabilidad de evaluar la calidad de su trabajo. Otro tipo de autoridades, no formales o institucionales, por ejemplo los expertos o los colegas más destacados, son aceptados por una minoría de docentes como agentes de su propia evaluación. Cabe destacar que el 54% de docentes uruguayos reconoce a los “colegas más destacados” para el cumplimiento de esta función.

Por último, sólo una minoría de docentes otorga una legitimidad a los destinatarios de la educación, sus familias y la comunidad, como para participar en la evaluación de su trabajo, por lo que toda propuesta en este sentido es probable que suscite reacciones críticas de la mayoría del cuerpo docente.

3. Actitudes frente a la evaluación de los alumnos y el sistema educativo y sus consecuencias

Durante la década del '90 la mayoría de los países latinoamericanos institucionalizó las evaluaciones nacionales de rendimiento académico. Los docentes tienen una opinión formada respecto de la legitimidad de estos estudios. El Cuadro N° 6 presenta una distribución del grado de acuerdo que han tenido estas políticas en los tres países analizados.

Cuadro N° 6: Grado de acuerdo de los docentes con las Evaluaciones Nacionales de Rendimiento.

	ARGENTINA	PERÚ	URUGUAY
Muy de acuerdo	6.6	16.6	9.9
De acuerdo	34.3	53.2	48.0
Sub total	40.9	69.8	57.9
En desacuerdo	30.4	15.3	22.8
Muy en desacuerdo	17.3	3.2	6.5
Sub total	47.7	18.5	29.3
Ns/Nc	11.4	11.6	12.8
Total	100.0	100.0	100.0

Los datos indican que en los tres contextos analizados la mayoría los docentes tienen actitudes diferentes respecto de esta nueva dimensión de la política educativa de los Estados nacionales latinoamericanos. En la Argentina, la mayoría de los docentes se manifiesta en desacuerdo con los operativos nacionales de evaluación, mientras que esta actitud crítica es minoritaria en el caso del Perú y el Uruguay. En la Argentina, otras evidencias indican que éste es uno de los capítulos de las reformas educativas recientes que generaron más resistencias entre los docentes. Incluso otras dimensiones sustantivas de las reformas, tales como el cambio en los contenidos, las políticas compensatorias, etc., gozan de más aceptación que los operativos nacionales de evaluación de la calidad de la educación. Estas diferencias en las actitudes pueden deberse a una combinación de factores que tienen que ver con la cultura de los docentes y con el tipo y forma en que se llevaron a cabo las experiencias de evaluación por parte de los respectivos ministerios de Educación. También

puede ayudar a rendir cuenta de estas diferencias observadas la capacidad que tienen los sindicatos de orientar la actitud y la opinión de los docentes en los países considerados. En la Argentina los sindicatos mayoritarios tuvieron una actitud abiertamente crítica respecto de la reforma educativa en general y especial de los denominados “Operativos Nacionales de Evaluación de la Calidad de la Educación implementados desde 1993 en adelante.

La opinión de los docentes respecto de esta cuestión también varía según algunas características básicas de los docentes. En la Argentina, por ejemplo, la actitud crítica es más frecuente entre los docentes de más edad (45 años o más), las profesoras de enseñanza media, los que trabajan en el sector público, los que se perciben como formando parte de las clases media baja y baja y los que manifiestan estar en una situación social peor que sus padres (cuando ellos eran niños).

En el caso del Perú, país donde la evaluación de calidad goza de más consenso, las posiciones favorables a la evaluación son aún más frecuentes entre los docentes de menos de 30 años y los que trabajan en establecimientos religiosos católicos.

En el Uruguay también se registran algunas asociaciones interesantes. Por ejemplo, el acuerdo con las evaluaciones nacionales de aprendizaje es más frecuente entre los docentes que dicen tener una situación económica mejor o igual a la de sus padres (es decir, que mantienen o mejoran su posición en la estructura social). En cambio, quienes experimentan una situación de decadencia social son significativamente más críticos en relación con esta cuestión. Por último, los docentes que se identifican con el rol de “transmisor de cultura y conocimiento” son más favorables a las evaluacio-

nes de calidad (casi un 70% de acuerdo) que sus colegas que se definen como “facilitadores del aprendizaje”, quienes expresan un porcentaje de acuerdo significativamente menor (57%). Resulta plausible que aquellos que definen su oficio en relación con la idea de transmisión de conocimiento se muestren más favorables a la idea misma de evaluar la calidad de los aprendizajes efectivamente desarrollados en los alumnos.

Es sabido que los resultados de las evaluaciones de calidad de la educación tienen varios usos posibles, entre ellos, el que tiene que ver con la evaluación de la tarea de los docentes. Uno puede preguntarse cuál es la opinión de los docentes al respecto. El Cuadro N° 7 provee información acerca de la distribución de estas opiniones. Los datos indican que la opinión de los docentes tiene un comportamiento totalmente diferente en los países analizados. Mientras que en la Argentina, que es el país donde la opinión crítica respecto de las evaluaciones de rendimiento son más frecuentes, la opinión acerca de su uso en la evaluación de los docentes se muestra muy dividida. A su vez, en el Perú y en Uruguay las posiciones son completamente opuestas: mientras que en el primero una mayoría absoluta está de acuerdo con esta propuesta, en el Uruguay sucede exactamente lo contrario. En este último país, donde la mayoría acuerda con las evaluaciones de rendimiento, también se manifiesta una oposición contundente al uso de las mismas como factor para evaluar la calidad del trabajo docente.

Cuadro N° 7: Opinión acerca del uso de los resultados de las evaluaciones de calidad para evaluar a los docentes.

	ARGENTINA	PERÚ	URUGUAY
Acuerdo	42.4	68.4	29.2
Desacuerdo	47.1	29.1	63.9
Ns/Nc	10.5	2.5	6.9
Total	100.0	100.0	100.0

Resulta de interés conocer qué factores están asociados con las opiniones predominantes en cada país. En la Argentina, el desacuerdo con el uso de los resultados de las evaluaciones como insumos para la política laboral docente es más frecuente entre los docentes del sector público (50%) en relación con los del sector privado (38%), los de nivel socioeconómico bajo (51.3%) en relación con los de nivel alto (43.1%), y las profesoras de secundario (49%) en comparación con sus colegas varones (36.5%).

En el Uruguay, país donde se registra el más alto porcentaje de rechazo a esta propuesta, el desacuerdo es más fuerte entre los docentes del sector público (66%) que entre los privados (57.8%), y entre los que no son titulares (68%) frente a los titulares (60.9%).

En el Perú, donde el acuerdo con la proposición es ampliamente mayoritario, esta actitud es aún más frecuente entre los docentes del sector privado (laico y religioso) y entre aquellos que se perciben como pertenecientes a las clases medias (en relación a los que se perciben como de clase baja).

Todo parece indicar que de la posición favorable a la implementación de un sistema nacional de evaluación de la calidad de la educación no se infiere automáticamente un

acuerdo con su uso como herramienta para evaluar la calidad del trabajo de los docentes.

En verdad, los tres países presentan situaciones dominantes diferentes. Mientras que en la Argentina se presenta una especie de polarización de opiniones respecto de estas dos cuestiones, en el Uruguay se combina una mayoría que manifiesta su acuerdo con los sistemas nacionales de evaluación de rendimiento, pero también un rechazo mayoritario para usarlos como instrumentos de la evaluación de los docentes. En el Perú, en cambio, la mayoría de los docentes legítima tanto una cosa como la otra.

CuadroN° 8: Relaciones entre acuerdo con la evaluación de la calidad de la educación y su uso para evaluar la calidad del trabajo de los docentes.

ARGENTINA

Uso de los resultados de la evaluación para evaluar la calidad del trabajo de los docentes.

Operativos de evaluación de la calidad de la educación	Sí	No	No sabe	Total
De acuerdo	53.8	38.3	7.9	100.0
En desacuerdo	35.7	55.9	8.4	100.0
No sabe	35.2	39.1	25.7	100.0
Total	43.1	46.8	10.2	100.0

PERÚ

	Sí	No	Ns / Nc	Total
De acuerdo	72.6	25.3	2.2	100.0
En desacuerdo	59.5	38.0	2.4	100.0
No sabe	57.1	38.1	4.8	100.0
Total	68.4	29.1	2.5	100.0

Como puede observarse en el Cuadro N° 8, en el caso de la Argentina y del Perú existe una cierta asociación positiva entre el acuerdo con los sistemas nacionales de evaluación y el uso de sus resultados para evaluar la calidad del trabajo docente. Sin embargo, también son frecuentes los casos de discordancia entre estas dos tomas de posición. A modo de hipótesis, este fenómeno puede explicarse con el siguiente razonamiento: muchos docentes están en desacuerdo con el modo o la forma en que se realizan las evaluaciones de calidad (la metodología y técnicas de evaluación empleadas, las dimensiones de los aprendizajes que se evalúan, el uso dado a los resultados, la legitimidad de las instituciones evaluadoras, etc.) pero no con el principio mismo de la evaluación, su necesidad y pertinencia, y su eventual uso para definir la calidad del trabajo docente. Si este argumento tuviera alguna pertinencia, es obvio que obligaría a una revisión crítica de las políticas concretas de evaluación de la calidad de la educación que se han venido desarrollando hasta el presente, al menos en los países aquí analizados.

4. Actitudes frente a la diferenciación y clasificación

Toda evaluación se traduce en una clasificación y en una diferenciación. Evaluar consiste en establecer un orden, una jerarquía. En este sentido, consiste en algo más complejo que una simple medición. Decimos que una escuela que obtiene un promedio elevado de rendimiento en un área del saber es “mejor” que otra que obtiene promedios más bajos. Como tal implica formular juicios de valor. Pero en todos los casos las evaluaciones (de alumnos o de docentes) tienen como resul-

tado una formalización y objetivación de diferencias (en la calidad y cantidad de aprendizajes desarrollados, en la calidad del trabajo docente, etc.). En este sentido resulta interesante interrogarse acerca de la legitimidad que tiene esta idea general en el cuerpo docente. Para aproximarnos a ella podemos recurrir al análisis de las respuestas dadas a dos cuestiones conceptualmente asociadas.

La primera tiene que ver con la idea misma de establecer diferenciaciones salariales con el fin de premiar a “los mejores docentes”. La pregunta, a primera vista, puede parecer ingenua, o incluso obvia si fuera planteada a una muestra de profesionales tradicionales, tales como médicos o abogados. Es probable que la mayoría de estos encuestados hubiera respondido positivamente a la pregunta “¿Usted cree que habría que encontrar los mecanismos adecuados para que los mejores ingenieros ganen más que los demás?”. Sin embargo, esta pregunta no tuvo una respuesta consensuada entre los docentes de los países aquí examinados. Por el contrario, en los tres casos se manifiesta una cierta polarización de las opiniones acerca de esta cuestión, que a primera vista parecía apelar a una respuesta uniforme y de sentido común.

Cuadro N° 9: Acuerdo con la proposición
“Habría que encontrar mecanismos adecuados para que los mejores maestros ganen más que los demás”.

Acuerdo	ARGENTINA	PERÚ	URUGUAY
Sí	42.9	64.1	39.1
No	40.1	31.8	45.9
No sabe	16.9	4.1	15.0
Total	100.0	100.0	100.0

Una vez más, el caso del Perú difiere del de la Argentina y el Uruguay. Mientras en ese país existe una clara mayoría en favor de la proposición, en el Uruguay sucede lo contrario y en la Argentina se presenta un cierto equilibrio entre las respuestas posibles. En el Perú también es bajo el porcentaje de docentes que no está en condiciones de tomar posición, mientras que esa actitud es más numerosa entre los docentes de la Argentina y el Uruguay.

Cuadro N° 10: Factores asociados con la oposición a la propuesta de establecer mecanismos adecuados para que los mejores docentes ganen más que los demás.

	ARGENTINA	PERÚ	URUGUAY
de nivel primario	X	X	X
del sector público	X	X	X
mujeres		X	X
de clase media baja y baja	X		
que viven en hogares pobres		X	
con movilidad social descendente	X	X	X
con movilidad social ascendente			
"facilitadores del aprendizaje"			X
"transmisores de conocimiento y cultura"		X	
inseguridad en el trabajo		X	

El Cuadro N° 10 muestra algunos factores asociados con la probabilidad de manifestar una posición crítica respecto de una propuesta que pretende establecer diferenciaciones y recompensas en función de la calidad del desempeño en el trabajo de los docentes. El desacuerdo es más probable entre los maestros de primaria, los que trabajan en el sector públi-

co, las mujeres y los que tienen posiciones y trayectorias sociales más desfavorecidas.

Otra variable cuyo comportamiento puede ayudar a comprender las resistencias al establecimiento de diferencias salariales entre los docentes es la que tiene que ver con el valor que los mismos asignan a la igualdad (frente a la libertad). En efecto, el cuestionario propone a los docentes que opten entre las proposiciones siguientes:

- “A. Para mí la libertad y la igualdad son igualmente importantes. Pero si tuviera que decidirme por una o por otra, consideraría la libertad personal como lo más importante, es decir, el que cada cual pueda vivir en libertad y desarrollarse sin obstáculos”.
- “B. La libertad y la igualdad son las dos importantes, pero si tuviera que decidirme por una de las dos, consideraría la igualdad como lo más importante, es decir, que nadie se vea desfavorecido y el que las diferencias de clase social no sean tan fuertes”.

Cuadro N° 11: Orientaciones hacia la libertad y la igualdad.

	ARGENTINA	PERÚ	URUGUAY
la libertad	34.4	40.8	33.5
la igualdad	39.3	47.6	42.2
ninguna de las dos/depende	19.6	8.4	20.7
Ns/Nc	6.7	3.2	3.6
Total	100.0	100.0	100.0

Los resultados que aparecen en el Cuadro N° 11 indican que los docentes tienden a privilegiar el valor de la igualdad por sobre el de la libertad. Esta predisposición a valorar la igualdad tiene un significado particular en un contexto histórico social signado por el incremento de las desigualdades en la distribución de bienes y recursos tan estratégicos como la riqueza, los ingresos y el poder. Desde este punto de vista, los docentes constituyen una categoría social que se opone y resiste a las políticas públicas que producen desigualdades.

Pero esta actitud también puede estar asociada con las resistencias que se presentan en proporciones significativas de docentes al establecimiento de diferenciaciones salariales en función de criterios de calidad de su trabajo profesional. Cabe tener en cuenta que el estatuto mismo de este oficio, que se desempeña en contextos institucionales muy estructurados, con una matriz histórica de tipo legal burocrático, no favorece el establecimiento de criterios de desempeño que favorezcan las diferenciaciones en las recompensas materiales en función de su calidad. Al respecto, cabe recordar que en estos contextos todavía tiene cierta vigencia el clásico principio “a igual trabajo igual remuneración” que regía en las ocupaciones asalariadas.

También en este caso la preferencia por estos valores –que de alguna manera estructuran el campo ideológico cultural de la mayoría de las sociedades capitalistas contemporáneas– se asocia con determinadas propiedades de los sujetos estudiados. El cuadro N° 12 ofrece alguna información acerca de los factores positivamente asociados con la preferencia por la igualdad.

Cuadro Nº 12: Factores asociados con la preferencia por la igualdad.

FACTORES	ARGENTINA	PERÚ	URUGUAY
jóvenes (menores de 30 años)	X	X	
movilidad descendente	X		X
clase percibida baja	X	X	X
sector público	X	X	
sector privado			X
profesores varones	X		
docentes jefes de hogar	X		
inseguridad en el trabajo	X		
regiones pobres	X		

A primera vista la posición y la trayectoria de los docentes en la estructura social estarían asociadas a la probabilidad de adherir a determinados valores políticos y sociales. Los más desfavorecidos (tanto en términos objetivos como subjetivos) estarían más predispuestos a valorar la igualdad que la libertad. En el caso de la Argentina, los docentes que viven en las regiones más pobres del territorio nacional y se sienten inseguros en su puesto de trabajo manifiestan la misma preferencia.

Por otra parte, en el caso de la Argentina y el Perú, los docentes del sector público adhieren preferentemente al valor de la igualdad, mientras que en el Uruguay se da la situación inversa, es decir, esta preferencia es más frecuente entre los docentes del sector privado.

Otro indicador del valor que los docentes asignan a la igualdad se deduce de la respuesta dada a una pregunta acerca del salario que deberían percibir distintas categorías ocu-

pacionales (y entre ellas los docentes de primaria y secundaria). Las respuestas producidas permiten tener una idea de las diferencias salariales que los docentes están dispuestos a reconocer entre distintos grupos sociales. En general se observa que los docentes de Argentina y Uruguay comparten una imagen ideal de la estructura social que es muy igualitaria. La brecha entre los salarios más elevados y los más bajos va de 2.8 en Uruguay a 3.3 en la Argentina y 4.5 en el Perú. En todos los casos estas brechas ideales son “igualitarias” frente a las distribuciones reales de salarios e ingresos en las sociedades consideradas.

En síntesis, todo parece indicar que la cultura dominante entre los docentes privilegia fuertemente el valor de la igualdad. Este apego al igualitarismo debe ser tenido en cuenta cuando se trata de instituir mecanismos que de alguna manera pueden ser percibidos como una amenaza a la realización de este ideal objetivamente valorado por los docentes.

Resumen y conclusiones

A modo de breve síntesis de lo expuesto pueden hacerse las siguientes observaciones:

- 1. La evaluación es un aspecto particularmente problemático de la dimensión pedagógica del trabajo docente. Ello puede deberse al hecho de que esta operación requiere el dominio de un conjunto de competencias técnicas relativamente complejas que requieren un aprendizaje permanente. Por otra parte, la evaluación lleva tiempo y tiene un aspecto rutinario que requiere un esfuerzo particular. Muchos docentes manifiestan que es escaso el tiempo que

tienen para realizar tareas de corrección de exámenes, cuadernos, etc. Sin embargo, cuando se le ofrece tiempo adicional de trabajo, pocos lo emplearían en realizar actividades de evaluación. Por último, no hay que olvidar que la evaluación por su naturaleza (estructuralmente arbitraria) y función (acreditación, asignación de premios y castigos, clasificación y jerarquización de los alumnos, etc.) tiene implicaciones fuertemente conflictivas, tanto en relación con los alumnos como con sus familias.

- 2. Las opiniones de los maestros acerca de los mecanismos y procedimientos para su propia evaluación son predominantemente críticas. Cuando se les pregunta por los criterios a tener en cuenta, en general todos incluyen (con diversas valoraciones) la evaluación periódica de su trabajo, junto a otras dimensiones más tradicionales. El problema surge cuando se trata de construir un consenso acerca de los agentes que participan en su evaluación y en los mecanismos concretos empleados. Como el educativo es un sistema todavía fuertemente burocratizado y jerárquico, las autoridades formales (directores de establecimientos y supervisores) son las únicas instancias que una mayoría absoluta de docentes reconoce como agentes legítimos de su propia evaluación. Es probable que para introducir nuevos sistemas y, sobre todo, nuevos agentes –la comunidad escolar, representantes de organizaciones locales, etc.– en los procesos de evaluación de la calidad del trabajo docente haya que introducir primero reformas en los modelos de trabajo y de organización de las instituciones escolares.
- 3. La opinión de los docentes acerca de los sistemas na-

cionales de evaluación de la calidad de la educación y acerca de su uso eventual como insumo para determinar políticas laborales no es homogénea. En general, puede decirse que ante estas cuestiones las opiniones y actitudes de los docentes son bien diversificadas. En el caso de la Argentina llama la atención el bajo grado de legitimidad que tienen tanto las evaluaciones realizadas, como su uso eventual para medir la calidad del trabajo docente. Este no es el caso del Perú y el Uruguay. Para explicar los factores asociados al rechazo o la aceptación habría que hacer otro tipo de indagación. Sin embargo, es probable que la forma y el contexto en que se llevaron a cabo las evaluaciones, así como los usos que se hicieron de la información producida, pueden explicar en parte estas diferencias de opinión. El grado de legitimidad de una determinada política pública es siempre una construcción social. No existen secuencias únicas del tipo “primero el consenso, luego la política”. En muchos casos primero es preciso intervenir, aun contra la oposición de muchos, e incluso de una mayoría. El consenso puede llegar después, por ejemplo, cuando pueden observarse los efectos (positivos) de las políticas adoptadas.

Por otra parte, y también a modo de hipótesis, el reconocimiento diferencial que tienen determinados dispositivos y políticas no sólo se explica en función de sus bondades intrínsecas o percibidas por los actores, sino también por el grado de legitimidad que tienen quienes los promueven y ejecutan. Argentina, donde se produce el más alto rechazo a los operativos de evaluación, es también el país donde las instituciones responsables de formular y ejecutar políticas educativas (Ministerio nacional y ministerios provinciales de Educación) son objeto

del nivel de desconfianza más alto por parte de los docentes.

- 4. Por último, el campo del debate sobre la evaluación pareciera estar organizado alrededor de dos polos típicos. Por una parte están los defensores de una especie de *igualitarismo formalista*. Este es el caso de aquellos que se niegan a medir y evaluar porque este operativo tiene como efecto una objetivación de las diferencias. La matriz burocrática del sistema educativo introduce en los miembros de las organizaciones una especie de igualdad estamental (por ejemplo entre alumnos y entre docentes). Desde una perspectiva racional/legal, los poseedores de los mismos títulos y certificados escolares son iguales, es decir, son intercambiables. El título, como capital cultural institucionalizado y garantizado por el Estado, es igualador, y en cierta medida oculta las diferencias reales que caracterizan a sus poseedores. Las evaluaciones de rendimiento ponen al descubierto diferencias (en competencias, conocimientos, actitudes, etc.) que los títulos y certificados ocultan y de esta manera introducen clasificaciones y jerarquías más detalladas y personalizadas. A partir de entonces un grupo puede compartir el mismo título (de bachiller, por ejemplo), pero al título se le agrega una calificación (por ejemplo, un promedio de rendimiento en una prueba final de evaluación) que distingue.

Por lo anterior, algunos sostienen que las evaluaciones de rendimiento tienen que servir para formalizar y hacer públicas esas diferencias entre individuos (o instituciones) que antes tenían sólo una existencia “de hecho” y no de derecho. Esta concepción se basa en que el simple hecho de medir y

hacer públicas las desigualdades tiene un efecto positivo en la medida que provee información tanto a los proveedores del servicio como a los usuarios (“consumidores”) del mismo. Sin embargo, cuando los resultados de las evaluaciones de individuos o instituciones se hacen públicos y se usan para establecer rankings “oficiales”, se produce un operativo de construcción social. El reconocimiento público y formal de las diferencias agrega su propia fuerza específica a esas diferencias reales. En este sentido contribuye a reproducirlas en el tiempo. Las evidencias indican que el simple conocimiento de la desigualdad y su “oficialización”, lejos de contribuir a superarlas, puede incluso “naturalizarlas” y, por lo tanto, consolidarlas en el tiempo.

Es posible construir una estrategia alternativa entre el igualitarismo formal y burocrático que cree superar las desigualdades por su negación burocrática y la consolidación de la desigualdad a través de su objetivación y formalización (mediante el establecimiento de jerarquías de calificaciones, tipologías de establecimientos, etc.). Ésta consiste en desplegar mecanismos que permitan conocer tanto las desigualdades como la lógica de su producción, pero no para establecer rankings o acreditaciones jerarquizadas (bachiller de 7, de 8 o de 10), sino para usar ese conocimiento para contrarrestar los factores que producen y reproducen las desigualdades, en especial en el salón de clases. Desde esta perspectiva, las evaluaciones ofrecen, más que una utilidad pública, un servicio técnico-pedagógico. Desde esta perspectiva es una herramienta que produce información para uso práctico de los educadores, los alumnos y sus familias. Esta estrategia se propone conocer las desigualdades, no para formalizarlas e institucionalizarlas, sino para combatirlas.

Si la evaluación ha de servir sobre todo como una he-

rramienta pedagógica, los docentes se constituyen en los usuarios privilegiados de sus productos. Por lo tanto, deben constituirse en protagonistas tanto de su diseño como de su implementación. Este no ha sido siempre el caso en los procesos de reforma educativa que se implementaron en la región durante las dos últimas décadas. Es de desear que de ahora en más las políticas que evalúen la calidad de la educación (la definición de qué es lo que se mide, cómo se mide y qué uso se le da a la evaluación) no sea un tema exclusivo de expertos y funcionarios de los ministerios de Educación.

Instituto Internacional de Planeamiento de la Educación

El Instituto Internacional de Planeamiento de la Educación fue creado por la UNESCO en París, en 1963 con el propósito de fortalecer las capacidades nacionales de los Estados Miembros de la UNESCO en el campo de la planificación y la gestión educativas.

El IPE contribuye al desarrollo de la educación en todo el mundo, difundiendo los conocimientos y formando a los especialistas en este campo. Su misión es la de promover el desarrollo de competencias en materia de definición e implementación de estrategias de cambio educativo.

La creación de su primera sede regional en Buenos Aires, en abril de 1997, se fundamenta en las especificidades de la situación latinoamericana y, particularmente, en el interés despertado por el proceso de transformación educativa que están llevando a cabo la mayoría de los países de la región.

Visite el sitio web del IPE - UNESCO Sede Regional Buenos Aires para acceder a información actualizada sobre nuestras actividades de formación, investigación, asistencias técnica, así como también a publicaciones y documentos referidos a distintos aspectos de la planificación y gestión educativas.

<http://www.iipe-buenosaires.org.ar>

Miembros del Consejo de Administración del IPE

Presidente:

Dato'Asiah bt. Abu Samah (Malasia)

Directora, Lang Education, Kuala Lumpur, Malasia.

Miembros designados:

Carlos Fortín

Secretario General Adjunto, Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD), Ginebra, Suiza.

Edgar Ortegón

Director, Programación de Proyectos e Inversiones, Instituto Latinoamericano y del Caribe de Planificación Económica y Social (ILPES), Santiago, Chile.

Jean Louis-Sarbib

Vicepresidente Senior, Banco Mundial, Washington D.C, Estados Unidos de América.

Ester Zurberti

Jefe, Servicio de Extensión, Educación y Comunicación, División de Extensión en Investigación y Capacitación, Departamento de Desarrollo Sustentable, FAO.

Miembros elegidos:

José Joaquín Brunner (Chile)

Director, Programa de Educación, Fundación Chile, Santiago, Chile.

Klaus Hüfner (Alemania)

Profesor, Free University, Berlín, Alemania.

Zeineb Faïza Kefi (Túnez)

Embajador Extraordinario y Plenipotenciario de Túnez en Francia y Delegado Permanente de Túnez ante la UNESCO.

Philippe Mehaut (Francia)

Director Adjunto, Centre d'études et de recherches sur les qualifications,

Una mirada política acerca de las evaluaciones

(Céreq), Marsella, Francia.

Teboho Moja (Sudáfrica)

Profesor de Educación Superior, Steinhardt School of Education, New York University, New York, Estados Unidos de América.

Teiichi Sato (Japón)

Embajador Extraordinario y Plenipotenciario y Delegado Permanente del Japón ante la UNESCO.

Tuomas Takala (Finlandia)

Profesor, University of Tampere, Tampere, Finlandia.