

LOS INDICADORES COMO HERRAMIENTAS PARA LA EVALUACIÓN DE LA CALIDAD DE LOS SISTEMAS EDUCATIVOS

Por: Felipe Martínez Rizo

Currículo: licenciado en Ciencias Sociales por la Universidad de Lovaina. Desde 1974 profesor de la Universidad Autónoma de Aguascalientes, donde ha sido director general de Planeación y de Asuntos Académicos, decano del Centro de Ciencias Sociales y rector. Sus temas de investigación incluyen la educación superior y la básica, en temas de calidad, planeación, evaluación y desigualdad. Investigador nacional nivel II. Miembro de la Academia Mexicana de Ciencias. Director general fundador del Instituto Nacional para la Evaluación de la Educación, de 2002 a 2008.

Resumen

En este artículo se presentan, primeramente, algunos antecedentes del desarrollo de los indicadores sociales y educativos, como parte de una tendencia generalizada a la cuantificación, en relación con las políticas sociales; en seguida, se hacen precisiones sobre la noción, en especial cuando se aplica a la evaluación de la calidad de los sistemas educativos. Luego, se incluyen sugerencias para una sistematización de la metodología para la construcción de indicadores, y se hacen consideraciones sobre los alcances y las limitaciones de este tipo de herramientas, para prevenir el peligro de esperar demasiado de los indicadores, o el de ignorar la complejidad de las tareas que implica su desarrollo, al pensar que pueden llevarse a cabo en tiempos muy breves.

Abstract

This work presents the background regarding the development of social and educational indicators, as part of a trend to measure social policies. We provide fine points concerning these concepts, in particular when applied to the evaluation of the quality of the educational systems. Proposals are offered for a systematization of the methodology for the construction of indicators, as well as considerations on the possibilities and the restrictions of these tools, to prevent the risk of expecting too much from the indicators, or to ignore the complexity of the tasks that its development implies (i.e. believing they can be carried out in a brief time).

INTRODUCCIÓN

Las discusiones entre partidarios de los acercamientos llamados cuantitativos y los cualitativos, en el campo de las metodologías de investigación en las ciencias del hombre, parecen interminables. Se pueden encontrar versiones tempranas desde la segunda mitad del siglo XVII (cfr. Martínez Rizo, 1991 y 1993), y todavía en el XXI no parecen cerca de acabarse (cfr. Anderson y Herr, 1999; Shavelson y Towne, 2002).

A los partidarios de los enfoques cuantitativos suele gustarles la cita de Lord Kelvin, que en 1883 dijo: "Cuando no podemos medir, nuestro conocimiento es pobre e insatisfactorio". A sus oponentes les gusta una réplica expresada, en 1929, por Jacob Viner: "...y cuando podemos medir nuestro conocimiento sigue siendo pobre e insatisfactorio" (cfr. Bulmer, 1986, p. 182).

Conocer la complejidad del mundo social no es fácil y, en efecto, las insuficiencias del conocimiento al respecto no se resuelven sencillamente por asignarles números mediante procedimientos poco rigurosos, como muchas veces se hace, lo que daría la razón a Viner. Por otro lado, la expresión de Lord Kelvin encierra su parte de verdad, en especial cuando se trata de conocimientos que se pretende sirvan de sustento para tomar decisiones de gran calado, como es el caso de políticas públicas que involucran recursos tan considerables como los que los países dedican a sus sistemas educativos.

Ejemplos de tales decisiones son el llevar o no la escolaridad obligatoria hasta el fin de la enseñanza secundaria; incrementar los días de clase o la duración de la jornada escolar; reducir el tamaño promedio de los grupos escolares; dar mayor o menor hincapié a las matemáticas, la lengua materna o una segunda lengua, entre otras. En estos casos, no pueden bastar como sustento las opiniones de sentido común de las autoridades respectivas, o las de las partes involucradas, como los maestros o los padres de familia. Hacen falta bases más sólidas que, en una u otra forma, y dada la escala de que se trata, deberán incluir información precisa, válida y confiable, sobre el costo de una u otra opción, pero también sobre sus probables repercusiones.

Sin duda, por lo anterior en todos los estados modernos, y en particular en los ministerios de educación, surgieron desde hace tiempo oficinas a cargo de la producción de estadísticas cada vez más complejas. Como una extensión de esta tendencia, las últimas décadas han visto el desarrollo de los llamados indicadores sociales, y los últimos años el de los indicadores complejos o compuestos.

En relación con estos últimos, y además de discutir sus ventajas y desventajas, una reciente publicación especializada señala que "...se les reconoce cada vez más como una herramienta útil para el análisis de políticas", y añade que cada año aumenta su número. Se cita una reciente revisión del tema (Bandura, 2006), que encontró más de ciento sesenta indicadores compuestos en uso por diversas organizaciones de todo el mundo para comparar el desempeño de los países (OCDE-European Commission, 2008, p. 3).

LA CUANTIFICACIÓN EN RELACIÓN CON LAS POLÍTICAS SOCIALES

El desarrollo de los indicadores: perspectiva histórica

En el siglo XVII, el término *estadística* se refería a la información sobre algún aspecto de la situación del Estado, en forma de datos numéricos o cualitativos; luego, el sentido se reservó a la información cuantitativa. El uso de medidas burdas de ciertas cuestiones de la vida pública es antiguo, pero sólo en épocas recientes la cuantificación se manifiesta de manera más amplia. Inicialmente, eso ocurrió en relación con elementos demográficos o económicos, que se prestan para un tratamiento numérico. El desarrollo de la noción de *indicador social* es posterior, por la necesidad de complementar cifras de población y datos económicos con información sobre otras dimensiones de la vida social.

Biderman (1966) muestra que, desde fines del siglo XVIII, los presidentes de Estados Unidos de América utilizaron estimaciones numéricas en sus mensajes anuales *sobre el estado de la Unión*. El número de estas medidas (*indicadores*) pasó de un promedio de tres en cada informe de fines del siglo XVIII a once a mediados del siglo XX. La tendencia se afirmó a partir de 1933: desde entonces, las expresiones oficiales de los objetivos del desarrollo nacional se acompañan por lo regular de medidas numéricas.

El interés por los indicadores numéricos se manifestó también en la preocupación por la calidad de las escuelas. Según Shavelson y colaboradores, al terminar la Guerra de Secesión se reflejó en el Common School Movement; la creación del Departamento de Educación, en 1867, se justificó, entre otras razones, por la necesidad de elaborar indicadores (1989, p. 1).

En este sentido, fue importante la iniciativa del presidente Hoover, quien formó, en 1929, una comisión de especialistas para desarrollar un estudio sobre las tendencias sociales en Estados Unidos. El informe de esa comisión, publicado en 1933 con el título *Tendencias sociales recientes en los Estados Unidos* (Comité Presidencial, 1933), es la obra que inaugura la extensa bibliografía sobre indicadores sociales.

Una nueva ola de interés se dio en la década de 1960, en la estela de la lucha por los derechos civiles, y en el marco de los programas de *la Gran Sociedad* y *Lucha contra la Pobreza*, del Presidente Lyndon Johnson. La importancia de esta tendencia fue tal que, para designarla, se acuñó la expresión *movimiento de los indicadores sociales*, que tuvo un componente educativo importante, a consecuencia del impacto del primer satélite artificial, cuya puesta en órbita por la URSS fue interpretada como síntoma del retraso de la enseñanza de matemáticas y ciencias en Estados Unidos.

El Special Study Panel on Education Indicators observa que el famoso informe de la National Comisión on Excellence in Education, publicado en 1983 con el título de *Una nación en peligro*, se basó en comparaciones internacionales del desempeño de los alumnos de diversos países, y que un año después el secretario de Educación comenzó la difusión de las *Wall Charts*, que publican cada año algunos indicadores de los sistemas educativos estatales (1991, p. 11).

A partir de la década de 1980, la preocupación por la calidad educativa alcanzó niveles sin precedentes en muchos países, ante la creciente competencia económica internacional. En ese marco adquieren sentido los esfuerzos por tener evaluaciones educativas más completas y confiables, lo que incluye mejores sistemas de indicadores.

En Estados Unidos, estas preocupaciones se reflejaron en diversos esfuerzos estatales durante la década de los ochenta (cfr. Odden, 1990); en las Metas 2000, adoptadas por el presidente tras la Cumbre Educativa de los gobernadores en octubre de 1989; en el establecimiento de estándares nacionales, el reforzamiento de los sistemas de pruebas y, finalmente, en las disposiciones de la ley No Child Left Behind, firmada por el presidente Bush a principios de 2002. En todos los casos el desarrollo de mejores sistemas de indicadores formó parte destacada de los esfuerzos de evaluación.

Los indicadores y las deficiencias de las estadísticas educativas

La relación entre indicadores y estadísticas es muy clara. Unos y otras representan esfuerzos por concretar nociones abstractas y cuantificar cualitativo. Hoy parece claro que un buen sistema de estadísticas es necesario para que las decisiones que se toman en los sistemas educativos tengan base sólida. Pese a ello, todavía a fines del siglo XX pocos países contaban con buenos sistemas de este tipo, en parte por las dificultades técnicas para desarrollarlos, junto con el crecimiento que experimentaron los sistemas escolares en todo el mundo. En 1993, Puryear decía al respecto:

...Las estadísticas sobre educación de aproximadamente la mitad de los estados miembros de la UNESCO, incluyendo al menos cinco de los nueve países

en vías de desarrollo más grandes, tienen lagunas y debilidades serias. Con mucha frecuencia las bases de datos sobre el tema carecen de confiabilidad. Aún en los países industriales avanzados los datos sobre costos y gastos educativos son muy deficientes. En casi todos los países los datos sobre analfabetismo no son confiables. Y los expertos opinan que las estadísticas educativas de 20 o 30 países son un verdadero desastre (1993, p. 4).

La mayoría de los países carece de medidas sistemáticas sobre resultados educativos y las evaluaciones comparativas internacionales son raras. Puryear señalaba que suele haber estadísticas mejores sobre cuestiones económicas, demográficas o de salud, y precisaba:

...Las estadísticas educativas existentes tienen un enfoque muy estrecho. Dejan fuera las medidas de calidad, de procesos y de productos. Los gobiernos han centrado la atención en un solo enfoque, bastante estrecho, para monitorear sus sistemas educativos –los conteos de tipo censo– y en un solo tipo de datos: los relativos a insumos. Se han concentrado casi enteramente en registrar el número de profesores, alumnos y edificios en el sistema formal, y en la importancia del gasto público destinado a la educación. Casi no han prestado atención a documentar cómo funcionan las escuelas o qué aprenden los alumnos [...] Un buen indicador de la existencia de estadísticas educativas de muy buena calidad, lo constituye la capacidad de reportar datos sobre la edad de los alumnos, que son indispensables para calcular tasas netas de matrícula. Únicamente unos 60 de los 175 estados miembros de la UNESCO reportan tales datos en forma consistente (1993, p. 6).

Poco antes, los primeros pasos de la OCDE para desarrollar su sistema de indicadores educativos comenzaban con un diagnóstico semejante:

Aunque muchos países publican impresionantes anuarios estadísticos, la cantidad y la calidad (validez, consistencia, etc.) de los datos recolectados es sumamente desigual, no sólo de un país a otro, sino incluso dentro de un mismo país [...] aunque la mayor parte de los 24 países participantes reportó que recababa datos para construir algunos indicadores educativos, su tipo y uso varía mucho y pocos países están en condiciones de ofrecer conjuntos completos [...] muchas presentaciones coincidieron en señalar problemas similares: datos incompletos o faltantes; falta de confiabilidad, debida en parte a técnicas de muestreo deficientes; insuficiente validez de las variables de contexto; dificultad para decidir qué datos recoger; problemas en cuanto al control de la información y el acceso a ella...

Aún los sistemas más completos de estadísticas nacionales presentan hoyos negros y lagunas de información en temas de los que casi no hay datos, como costos unitarios, gasto educativo privado o la contribución económica de empresas y familias al gasto educativo total. La mayoría de los sistemas son notablemente débiles en cuanto a información sobre el

conocimiento que tienen los maestros de las materias que deben enseñar, el aprendizaje de los alumnos, en especial en niveles cognitivos altos, o los procesos que tienen lugar dentro de la escuela (CERI-INES 1991, pp. 8-9).

En la misma época, expertos encargados, por parte del gobierno de Estados Unidos, del diseño de un sistema de indicadores educativos a escala nacional identificaban problemas semejantes, que muestran que la pobreza y la inconsistencia de las estadísticas educativas no son exclusivas de países menos desarrollados:

Algunos problemas técnicos serios de los indicadores actualmente existentes, junto con grandes lagunas en las fuentes de datos disponibles, plantean problemas formidables a la tarea de construir un sistema de indicadores educativos en el nivel nacional. El Council of Chief State School Officers, a solicitud del National Center for Education Statistics, ha analizado las fallas de medidas de resultados educativos comparables, observando por ejemplo que hasta 1987 los estados americanos empleaban al menos diez maneras distintas de contar sus escuelas y unos doce métodos diferentes de reportar las cifras de matrícula. Los datos sobre deserción son notoriamente poco confiables y representan unos de los datos más ambiguos que se reportan sobre la educación norteamericana (Special Study Panel, 1991, p. 15).

PRECISIONES TERMINOLÓGICAS Y CONCEPTUALES

La noción de indicador

El término *indicador* se utiliza en sentidos no coincidentes. En la tradición metodológica de Lazarsfeld, es parte del proceso de operacionalización: *variables, dimensiones e indicadores*. El concepto clave es el de variable, que denota un aspecto de la realidad que se quiere explorar que adopta valores distintos en la población que se estudia.

Algunas veces, las variables se conceptualizan de tal manera que resulta muy sencillo identificar la realidad a la que corresponden. Es el caso de variables como las de sexo, edad o estado civil, tan utilizadas en muchas investigaciones.

En otros casos, las variables son conceptualizadas de modo menos claro, de suerte que, para manejarlas en forma práctica, es necesario buscar conceptos que se utilicen con mayor precisión y sean equivalentes o representativos del concepto más abstracto. Si interesa, por ejemplo, el aspecto (variable) *nivel socioeconómico*, se podrá recurrir en su lugar al concepto más fácil de manejar: *ingresos mensuales*. Este último es un *indicador* de la variable anterior. Sin duda, la traducción de ciertas variables a sus indicadores no necesariamente resulta afortunada.

En forma relacionada con la anterior, pero con algunas diferencias, desde los años sesenta se trató de identificar conceptos particulares precisos que permitieran concretar y valorar otros tan amplios e imprecisos como *desarrollo integral o bienestar social*. El concepto de *producto interno bruto*, tan caro a los economistas, fue durante mucho tiempo el único criterio para apreciar el grado de desarrollo de un país. Evidentemente, el concepto ignoraba importantísimas dimensiones de una realidad tan compleja. Por ello, se trató de concebir otros *indicadores* de conceptos tan comprensivos.

En la obra colectiva ya citada, editada por Raymond Bauer con el título *Social Indicators* (1966), se puede encontrar una consideración similar:

El problema clave de un sistema de indicadores sociales [...] es que nunca podemos medir directamente las variables que nos interesan, sino que tenemos que seleccionar substitutos en el lugar de esas variables. Podemos, por ejemplo, estar interesados en saber si una persona es o no ambiciosa. Pero no podemos observar la ambición en sí misma. Podemos hacer preguntas a la persona, y escuchar sus respuestas, o bien observar qué tan intensamente trabaja y en busca de qué tipo de retribución; a partir de ello podemos inferir si la persona es o no ambiciosa (Bauer, 1966, p. 45).

Bauer toma de Gross un esquema que ejemplifica el paso de una *gran abstracción* (abundancia) a conceptos más precisos, *indicadores cuantitativos* (en la forma de cantidades de bienes y servicios, producidos en ciertos periodos, y expresados en unidades físicas o monetarias), pasando previamente por *abstracciones intermedias* (riqueza: cantidad, distribución) (Bauer, 1966, p. 45).

En el sentido que ahora nos ocupa la palabra indicador, no se refiere de modo necesario a un aspecto particular de una variable más amplia, sino que, al contrario, puede tener un carácter sintético, e integrar varios más particulares, pero siempre con la pretensión de hacerlo de tal suerte que sea posible un tratamiento preciso, cuantitativo. Así puede entenderse la siguiente definición de indicador: “Estadística sintética (*summary*) sobre el estado actual de un sistema educativo” (Wyatt, 1994, p. 99).

Según Wyatt, la definición más aceptada hoy se deriva de Oakes, quien expresa que un indicador debe informar acerca, al menos, de uno de los siguientes aspectos:

- Logros de un sistema educativo para obtener ciertos resultados; el indicador se liga a los objetivos, y es un referente para medir los avances (*benchmark*).
- Características que la investigación ha mostrado se relacionan con resultados; el indicador tiene valor predictivo, porque su modificación traerá consigo otras.
- Rasgos centrales del sistema educativo para entender cómo funciona.
- Información relacionada con problemas o aspectos relevantes para la política educativa, que permitan apoyar la toma de decisiones (cfr. Oakes, 1986).

Wyatt coincide con Oakes en que un indicador debe:

- Medir aspectos que se encuentren en todos los ámbitos del sistema que se evalúa (*ubiquitous*), de suerte que puedan hacerse comparaciones.
- Medir aspectos duraderos del sistema, de tal modo que puedan analizarse tendencias y cambios en el tiempo.
- Ser fácilmente inteligible para una audiencia amplia.
- Tener factibilidad, teniendo en cuenta el tiempo, el costo y la capacidad técnica requeridos para su obtención.
- Ser en general aceptado por sus cualidades técnicas de validez y confiabilidad (Oakes, 1986, pp. 1-2; Wyatt, 1994, p. 105).

De acuerdo con Jaeger, quien muestra la poca consistencia de las definiciones de indicadores en la literatura especializada (1978), Shavelson y colaboradores con-

sideran que conviene adoptar una postura pragmática, y proponen la siguiente definición de trabajo: “Un indicador es una estadística simple o compuesta que se relaciona con un constructo educativo básico y es útil en un marco de políticas públicas” (1989, pp. 4-5).

La noción del Special Study Panel on Indicators es similar:

Un indicador es una estadística que mide nuestro bienestar colectivo. Un verdadero indicador mide la salud de un sistema, sea económico, de empleo, de servicios médicos o educativos [...] A diferencia de otras estadísticas, un indicador debe ser relevante para la toma de decisiones, en función de ciertos problemas; debe ofrecer información sobre un rasgo significativo del sistema al que se refiere; y generalmente incluye algún estándar contra el cual pueda juzgarse si hay progreso o retroceso (1991, p. 12).

Fuera de Estados Unidos, los esfuerzos por precisar la definición del término indicador han seguido caminos similares, y se han inclinado por un acercamiento pragmático. Un reciente documento al respecto de la Comisión Europea establece:

Un indicador puede definirse como la medición de un objetivo a perseguir, de un recurso a movilizar, de un efecto alcanzado, de una estimación de calidad, o una variable de contexto. Un indicador ofrece información cuantificada con el propósito de ayudar a los participantes en acciones públicas a comunicarse, negociar y tomar decisiones [...] Un indicador cuantifica un elemento que se considera relevante para el monitoreo o evaluación de un programa (1999, p. 17).

Tipos de indicadores

Al igual que Shavelson y otros, Oakes distingue dos tipos básicos de indicador: los que consisten, respectivamente, en estadísticas simples o compuestas (1986, p. 3). En una clasificación más elaborada, la Comisión Europea ofrece una tipología de indicadores que los cataloga a partir de seis dimensiones, según:

- El grado de procesamiento de la información: indicadores *elementales, derivados o compuestos*.
- La comparabilidad de la información: indicadores *específicos o genéricos*.
- El alcance de la información: indicadores de *programa* y de *contexto*.
- Las fases del programa: de recursos o insumos; de *salidas* o productos inmediatos; de *resultados* o productos mediatos; y de *impacto*, o de largo plazo.
- El criterio de evaluación: de *relevancia, eficacia, eficiencia o desempeño*.
- Su utilización: de *monitoreo*, para uso inmediato, o de *evaluación*, al final del proceso (Comisión Europea, 1999, pp. 19-38).

Sistemas de indicadores y enfoques para su desarrollo

La complejidad de los sistemas educativos hace que ningún indicador particular pueda ofrecer una imagen suficientemente amplia del conjunto. Por ello, es necesario construir un *sistema de indicadores* teniendo en cuenta que un *sistema* no es la simple acumulación o yuxtaposición de indicadores particulares, sino un conjunto articulado de ellos, con base en cierta estructura:

Otro concepto central en la discusión es el de sistema de indicadores. Sea que se trate de estadísticas simples o compuestas, un indicador singular difícilmente podrá proporcionar información útil sobre fenómenos tan complejos como los educativos. Los sistemas de indicadores se diseñan por lo general para generar información más amplia y precisa. Pero debe precisarse que un sistema de indicadores es más que una simple colección de estadísticas. Idealmente, un sistema de indicadores mide diversos componentes del sistema educativo, y ofrece también información sobre cómo interactúan los componentes singulares para producir el efecto de conjunto. En otras palabras, el conjunto de la información que ofrece un sistema de indicadores es más que la suma de sus partes (Shavelson *et al.*, 1989, pp. 5-6).

Cuando se trata de diseñar indicadores educativos, puede haber dos enfoques: algunas veces, el propósito será monitorear el grado en que se alcanzan algunas metas de política, en cuyo caso bastará un número relativamente pequeño de indicadores, que se definirán en función de tales metas. Otras ocasiones, se pretenderá cubrir de manera más completa los aspectos relevantes del sistema educativo para valorar su calidad en sentido amplio, sin limitarse a los aspectos más relacionados con las políticas vigentes; en este caso se requerirá un número mayor de indicadores, organizados de modo que constituyan también un verdadero sistema.

El segundo de estos dos enfoques tiende a la exhaustividad y se podría considerar más deductivo o teórico; el primero se limita conscientemente a algunos aspectos considerados de especial relevancia, puesto que han sido previstos en las metas, y se podría ver como más inductivo o práctico. Ambos enfoques pueden ser sólidos y, de hecho, los dos han dado lugar a sistemas de indicadores interesantes.

Los sistemas de indicadores y la noción de calidad en educación

No es sencillo ponerse de acuerdo en una definición de calidad educativa. Quizá la dificultad se deriva, en parte, del hecho de buscar el consenso en el terreno de los conceptos abstractos. Parece más factible conseguirlo en términos concretos. La dificultad también se debe a la variedad de los objetos a los que puede referirse la noción de calidad, que pueden ser un currículo, un texto escolar, un maestro, un alumno o una escuela.

El uso de sistemas de indicadores como herramientas de evaluación, sin embargo, no suele referirse a esos objetos, sino a uno diferente y muy complejo: un sistema educativo. Por ello, parece más sencillo ponerse de acuerdo en qué es un buen sistema educativo, para luego subir de nivel de abstracción y definir el concepto de calidad de un sistema educativo, como paso necesario para el diseño de un sistema de indicadores de la calidad, en ese sentido preciso. Propongo la siguiente definición:

Un buen sistema educativo es aquel que reúne las siguientes cualidades, que definen dimensiones abstractas de la noción de calidad:

- Establece un currículo adecuado a las necesidades individuales de los alumnos (*pertinencia*) y a las de la sociedad (*relevancia*).
- Logra que la más alta proporción posible de destinatarios acceda a la

escuela, permanezca en ella y egrese alcanzando los objetivos de aprendizaje (*eficacia interna y externa*).

- Consigue que los aprendizajes se asimilen duraderamente y den lugar a comportamientos sociales fructíferos para la sociedad y los individuos (*impacto*).
- Cuenta con recursos suficientes (*suficiencia*) y los aprovecha bien (*eficiencia*).
- Tiene en cuenta la desigualdad de situaciones y apoya a quienes lo requieran, para que los objetivos sean alcanzados por el mayor número posible (*equidad*).

Ninguna de las dimensiones anteriores es suficiente por sí sola para definir la calidad; es la integración armoniosa de todas ellas lo que la precisa. Por ello, puede proponerse una definición de la noción en términos abstractos, como sigue: *La calidad del sistema educativo es la cualidad que resulta de la integración de las dimensiones de pertinencia y relevancia; eficacia interna; eficacia externa a corto plazo; eficacia a largo plazo o impacto; suficiencia; eficiencia; y equidad.*

Sistemas de indicadores y evaluación de la calidad de un sistema educativo

Precisar las dimensiones de la calidad de un sistema educativo tiene una consecuencia importante para su evaluación: no debe reducirse a la aplicación de pruebas estandarizadas que midan el grado en que los alumnos de ciertos niveles alcanzan los aprendizajes que se esperan. Un buen sistema de información y evaluación de la calidad de un sistema educativo deberá comprender, además de pruebas de rendimiento, indicadores derivados de las estadísticas educativas tradicionales y otros estudios sobre aspectos particulares, como los recursos de las escuelas y los procesos que tienen lugar en su interior.

El interés por las pruebas de rendimiento no debe llevar a abandonar los indicadores tradicionales, sino a perfeccionarlos y manejarlos junto con los más recientes, además de seguir explorando el desarrollo de los que todavía están mal atendidos, en especial los de *relevancia y pertinencia, impacto y equidad*.

Cuando lo que se pretende no es supervisar el cumplimiento de algunas metas estratégicas de política, sino evaluar la calidad de un sistema educativo de manera amplia, debe matizarse la frecuente recomendación de que conviene que el número de indicadores que se utilice sea reducido, ya que un número grande dificultaría la toma de decisiones basada en ellos. En una dirección opuesta, el Special Study Panel on Education Indicators del National Center for Education Statistics (NCES) señala:

...a medida que nuestro trabajo avanzaba nos convencimos de que la búsqueda de un número limitado de indicadores educativos clave era equivocada. Como ningún conjunto limitado de indicadores haría justicia de la complejidad de la empresa educativa, un conjunto limitado no sólo reflejaría cierta agenda educativa, sino que definiría una agenda así. Si, por ejemplo, se establece que la enseñanza de las matemáticas y la geografía son suficientemente importantes para merecer que se les dedique un indicador, pero no ocurre lo mismo con la enseñanza de la música o de una lengua extranjera, los educadores tomarán eso como una señal a seguir (1991, p. 7).

METODOLOGÍA PARA EL DISEÑO DE INDICADORES

A diferencia de lo que ocurre con las pruebas de aprendizaje, la metodología para diseñar indicadores educativos no está sistematizada o codificada con suficiencia. En lo relativo a pruebas, desde hace medio siglo se dispone de manuales que explican con detalle los pasos para elaborar instrumentos adecuados. No es sencillo, desde luego, seguir al pie de la letra todas las indicaciones técnicas, pero si se consigue hacerlo en un grado aceptable, el resultado será un instrumento razonablemente válido y confiable, aun si no se incorporan los desarrollos de las últimas décadas.

No se encuentran, en cambio, textos similares para guiar el proceso de construcción de sistemas de indicadores. Los escasos apuntes que hay al respecto deben buscarse en documentos técnicos y artículos poco accesibles, pero no puede hablarse todavía de un cuerpo estándar de orientaciones en este campo. Lo anterior es uno de los factores que explican que muchos de los sistemas de indicadores existentes en nuestro medio, sobre tema educativo u otro, sean el resultado de procesos poco rigurosos de diseño, y adolezcan de deficiencias graves.

En seguida se presenta la metodología para el diseño de sistemas de indicadores sistematizada en el Instituto Nacional para la Evaluación de la Educación, a partir de la experiencia del autor, y recogiendo elementos de Shavelson et al. (1989 y 1991) y de la Comisión Europea (1999). La metodología comprende cuatro etapas.

Elaboración de la lista de indicadores

El punto de partida es la idea de que un sistema de indicadores no se reduce a una lista de ellos, simplemente yuxtapuestos. Para constituir un auténtico sistema, por el contrario, los indicadores deben integrarse de manera lógica, según criterios precisos.

Si el propósito que guía la construcción de un sistema de indicadores es evaluar la calidad del sistema educativo con amplitud, precisar esa noción de calidad es fundamental. Como se ha visto, una noción rica de calidad debe incluir dimensiones como relevancia y pertinencia, eficacia interna y externa, impacto, suficiencia y eficiencia y equidad.

Un marco de referencia para construir indicadores, que se propone como alternativa al enfoque sistémico, es el del derecho a la educación. No parece que esos dos marcos deban oponerse, sino que más bien se complementan. Un sistema educativo, en efecto, no podrá ser considerado bueno si no consigue que se cumpla con las exigencias del derecho a una buena educación. Es necesario, pues, precisar lo que se entiende por una buena educación, lo que lleva de nuevo a la necesidad de precisar dimensiones como relevancia, eficacia y eficiencia, entre otras, como se hace en el enfoque sistémico expuesto en este texto.

Si se decide desarrollar un sistema de indicadores a partir de ciertas metas de política previamente definidas, este primer paso será muy sencillo, pero no habrá que perder de vista algunas dimensiones de la calidad educativa que el conjunto de metas de que se trate no incluya, por dar prioridad a otros aspectos. Esa prioridad puede ser justificada, y será razonable centralizar los esfuerzos en cierta dirección, pero no por ello deberán olvidarse las demás dimensiones de la calidad.

Los enfoques sistémico y estratégico tampoco se contraponen. El primero, con su pretensión de (relativa) exhaustividad, siempre será necesario, pero es razonable conceder especial atención al desarrollo de indicadores que atiendan las metas definidas como prioritarias en las políticas educativas en un momento determinado.

Deberá evitarse el frecuente error de diseñar un sistema de indicadores utili-

zando sólo la información ya existente. Un buen sistema de indicadores convendrá diseñarse a partir de lo que se defina como importante, sistémica o estratégicamente, aunque para muchos aspectos no se cuente con datos adecuados y deban realizarse después todos los esfuerzos para obtener la información necesaria.

Una vez establecida una primera lista de indicadores a desarrollar, sea con un enfoque o con otro, deberá pasarse a la siguiente etapa.

Desarrollo de cada indicador

Deberá precisarse el nombre y la definición de cada indicador, y ubicársele en el marco de referencia que se esté utilizando. Habrán de señalarse, además, las fuentes de obtención de la información de base; la fórmula de cálculo y, en su caso, los criterios para ello; las posibles formas de desagregar el indicador (v. gr., por género, entidad, modalidad de escuela, etcétera); y aportar elementos para la interpretación de los resultados.

Pese a todos los esfuerzos que se hagan, es seguro que el resultado de esta etapa no será un indicador perfecto, por lo que será necesario llevar a cabo, en la siguiente etapa, un proceso de depuración y refinamiento.

Jueceo inicial

Antes de proceder a organizar costosos procesos de recolección de información para alimentar un sistema de indicadores, cada uno de los elementos que lo forman deberá someterse a un cuidadoso escrutinio, mediante un procedimiento de jueceo entre expertos y usuarios, para valorar a priori el grado en que cada indicador parece satisfacer los criterios de calidad aplicables que son, al menos, los siguientes:

- *Validez*, cualidad técnica básica, la propiedad del indicador que consiste en que mida realmente lo que se pretende que mida.
- *Confiabilidad*, la otra cualidad técnica esencial, consistente en que el indicador se defina conceptual y operacionalmente de forma tal que la información que permita obtener sea consistente a lo largo de sucesivas aplicaciones.
- *Comparabilidad*, si el indicador es aplicable en contextos (educativos) diferentes de modo tal que permita comparaciones significativas.
- *Estabilidad temporal*, si permite comparaciones a lo largo del tiempo.
- *Actualidad* de la información que ofrece el indicador (*freshness*), ya que uno que no pueda ofrecer información razonablemente reciente, no tiene mucho valor.
- *Sensibilidad*, que se dará si valores distintos de un indicador se asocian, en forma consistente, a diferencias significativas de los sistemas educativos.
- *Factibilidad de implementación* del indicador, si la información necesaria para construirlo puede obtenerse. Cuando esa información no esté disponible, deberá valorarse la importancia del indicador en relación con el costo que implicará.
- *Importancia*, en el sentido de que el indicador se refiera a un elemento del sistema educativo que tenga peso significativo en relación con alguna dimensión de la calidad.
- *Utilidad*, como el grado en que el indicador se refiera a un aspecto en el que las decisiones que tomen los responsables puedan impactar de manera significativa.

- *Claridad*, en el sentido de que la manera en que presente la información deberá facilitar una adecuada interpretación por parte de los usuarios.

Además de los criterios anteriores, aplicables a cada indicador, deberá valorarse la calidad del conjunto, y juzgar si cubre de manera suficiente y equilibrada el objeto de estudio con un número razonable de indicadores ni excesivo ni escaso, esto es, los criterios de *cobertura*, *balance* y *parsimonia* del sistema de indicadores como tal.

El análisis deberá involucrar un número razonable de jueces, de tres tipos: *expertos*, investigadores o especialistas de los temas cubiertos; *usuarios especializados*, responsables del sistema educativo con experiencia de nivel alto y medio; y *usuarios de base*, como directores de escuela, maestros y padres de familia.

El proceso de jueceo deberá comprender al menos tres vueltas, en forma similar a la que se emplea en la técnica *delphi*. En cada vuelta se solicitará a los jueces que califiquen cada indicador según los criterios de calidad en una escala ordinal de cinco valores, de muy adecuado a muy inadecuado, con las siguientes variantes:

- En la primera vuelta se solicitará simplemente la opinión de los participantes sobre cada uno de los aspectos mencionados de cada indicador.
- En la segunda se pedirá a los participantes que vuelvan a opinar sobre cada aspecto de cada indicador, y se tendrá a la vista información sobre las opiniones de los demás participantes en la primera vuelta, en la forma de medias y desviaciones de las puntuaciones obtenidas por cada indicador en cada criterio. A quienes tengan opiniones que difieran mucho de la tendencia promedio, se les pedirá que expresen argumentos para defender su posición.
- En la tercera se solicitará de nuevo la opinión de los participantes, pero se les dará, además, información sobre los argumentos de los participantes con opiniones extremas o discordantes en la vuelta anterior.

Con base en las opiniones de los jueces se podrá decidir cuáles indicadores satisfacen, de manera más o menos amplia, los criterios de calidad antes mencionados. Entre una vuelta y otra, podrán modificarse o descartarse los indicadores que, según el consenso recabado, no satisfagan los criterios de calidad. La opinión de los *expertos* deberá tener mayor peso en las decisiones sobre los criterios de validez, confiabilidad, comparabilidad y estabilidad temporal, así como los de actualidad y sensibilidad; la opinión de los *usuarios especializados* es de especial relevancia para decidir sobre la factibilidad, importancia y utilidad; y la de los *usuarios de base*, para juzgar sobre la claridad de los indicadores.

Teniendo en cuenta que algunos de los participantes en el proceso podrán abandonarlo antes de terminar, y el tiempo que implica recoger las opiniones de cada vuelta y procesarlas antes de la siguiente, habrá que considerar un tiempo razonable para cada una (de dos a cuatro semanas) e incluir, desde el inicio, a un número de jueces mayor que el que se espera termine el proceso.

Prueba piloto o primera aplicación

Tras las etapas anteriores podrá procederse a una primera aplicación del sistema, y alimentarlo con la información necesaria, sea ya existente o generada *ex profeso*, así como realizar los procesos de depuración y cálculo necesarios para obtener los

primeros valores de cada indicador.

Con los resultados de la primera aplicación se podrá hacer un nuevo análisis de la calidad de cada indicador y del sistema en conjunto, esta vez *a posteriori*, revisando la consistencia de la información y contrastándola con otra comparable, por ejemplo derivada de investigaciones independientes. Podrá también valorarse en qué medida la información que arroja el sistema es interpretada correctamente y utilizada para sustentar acciones de mejora. Tras lo anterior, el sistema podrá operar con regularidad, con dos procesos adicionales: uno de refinamiento permanente de los indicadores que lo requieran, con base en la experiencia que se irá acumulando, y otro de establecimiento y ajuste de parámetros o estándares de referencia, con los cuales se contrastarán los valores de los indicadores, para llegar a un juicio de valor.

No debe olvidarse que la evaluación implica la medición, pero no se reduce a ella. Un indicador captura o mide cierto aspecto o dimensión de la realidad a evaluar, en nuestro caso del sistema educativo, lo cual es necesario, pero no suficiente para evaluar ese aspecto: para llegar a un juicio sobre lo adecuado o inadecuado de la realidad medida hace falta un referente que defina lo que se considera adecuado o inadecuado.

Ese referente con que debe contrastarse el resultado de la medición es el *parámetro* o *estándar*. Es normativo; no empírico. Para definirlo, la pregunta que hay que formular es ¿con qué o con quién debemos compararnos? Los referentes pueden ser de tres tipos: *óptimos*, *promedio* y *mínimos*.

Tratándose de juicios de valor sobre sistemas educativos nacionales, los referentes óptimos pueden ser los resultados educativos de los países más avanzados. Al interior de un sistema educativo, cuando se trate de juicios de valor sobre escuelas, los referentes óptimos serán los resultados de las mejores. Usar como punto de comparación tal tipo de referentes óptimos equivale al enfoque conocido como *benchmarking*, o la identificación de las llamadas *mejores prácticas*.

El uso de referentes promedio implica comparar los resultados de un sistema, una escuela o un alumno con la media de los sistemas, escuelas o alumnos que se evalúa. Es el tipo de comparación que se refleja en expresiones como: *tal escuela está por encima de la media nacional*; o bien: *la media de México es inferior a la de los países de la OCDE*. Cuando se compara a un país con otros de *desarrollo similar*, de alguna manera se utiliza este tipo de referente. El uso de referentes mínimos implica que se establezca el menor valor aceptable, cuyo cumplimiento permite, por ejemplo, que un alumno sea aprobado, o que una institución sea acreditada. El uso de los países de menor desarrollo como puntos de comparación podría verse como una aproximación al uso de referentes mínimos.

Cada referente (óptimo, promedio y mínimo) tiene cierto sentido; cada uno arroja luz sobre la realidad evaluada desde cierta perspectiva, y ninguno es suficiente. Parece conveniente usar los tres tipos, para lograr una apreciación más completa.

Tiene sentido comparar la situación educativa de México con la de países más avanzados. La situación de estos últimos es diferente de la del nuestro, pero no deja de ser un punto de referencia, una meta a perseguir, aunque sea a mediano o largo plazo; la globalización del mundo contemporáneo lo hace inevitable. Sin embargo, utilizar sólo tales referentes envuelve una dosis de injusticia, al no tener en cuenta numerosos factores demográficos, económicos, sociales y culturales. La equidad implica usar parámetros que consideren los contextos de las realidades

evaluadas, comparando lo realmente comparable. Por ello, el emplear también como referentes a los países de desarrollo similar, usando referentes *promedio*, es un complemento deseable en las comparaciones internacionales.

La evaluación de los sistemas educativos de las entidades federativas puede hacerse, de manera similar, en relación con referentes óptimos, promedio o mínimos. Como referente óptimo puede usarse el nivel de las entidades más desarrolladas, que pueden considerarse referentes de *benchmarking* para las menos avanzadas. La media nacional es el referente promedio natural. En cuanto a los referentes mínimos, las autoridades educativas podrán establecerlos. En el ámbito internacional, organismos como la UNESCO y la OCDE pueden fijar este tipo de parámetros mínimos.

Una forma diferente, y sugestiva, de definir referentes es la de emplear como tales la situación del sistema a evaluar en el pasado, el presente o el futuro. La situación en cierto momento del pasado puede verse como referente mínimo.

El futuro puede usarse como punto de referencia, en el sentido de que un sistema, o una escuela, puede valorar si alcanza o no, si se aproxima o se aleja, a mayor o menor velocidad, a las metas que se haya establecido en determinado horizonte temporal. Tal vez, estos parámetros sean los más pertinentes para valorar la calidad educativa: una escuela o sistema escolar de calidad es, finalmente, aquella o aquel que siempre mejora respecto a sí mismo, sin idealizar el pasado y con metas ambiciosas, pero realistas para el futuro.

REFLEXIONES FINALES

Quienes se dedican a desarrollar indicadores, así como los responsables del diseño de políticas, que los requieren para sustentar mejor sus decisiones, deben tener claro que, contra lo que a menudo se piensa, el proceso de desarrollo de estas herramientas es necesariamente largo, si se quiere que sea sólido.

No es raro que se piense que la tarea en cuestión puede hacerse en pocas semanas, si ya existen bases de datos con las estadísticas que suelen recogerse con periodicidad en todos los sistemas educativos. Se pierde de vista que, para asegurar la solidez del trabajo, los pasos indicados en la metodología antes presentada se deben seguir con cuidado; que se debe involucrar a grupos de actores calificados; y que los productos del trabajo deben revisarse y corregirse una y otra vez, antes de que se les considere establecidos. En el caso de indicadores educativos, hay que tener en cuenta el ritmo anual de los procesos de recolección de estadísticas que se obtienen de las escuelas al inicio o fin de cada ciclo escolar; así podrá entenderse que el tiempo que implica no sólo el primer planteamiento de un indicador, sino su refinamiento hasta alcanzar un grado aceptable de madurez, implica varios años.

Al sacar las lecciones del *movimiento de los indicadores sociales* de los años sesenta, una especialista advertía sobre el tiempo que se “necesita para el desarrollo de un concepto satisfactorio, luego para la producción de una medida práctica y, finalmente, para asegurar que el indicador resultante sea entendido y utilizado”, y recordaba que “los indicadores existentes han requerido veinte, treinta o más años, desde el momento en que aparecieron las primeras demandas públicas y los primeros análisis teóricos...”.

Precisando las ideas anteriores, esa experta señalaba:

Hay que contar con que la creación de un indicador nuevo toma largo tiempo. Este tiempo difícilmente puede abreviarse, dada la naturaleza iterativa de los procesos de formación de conceptos y de los esfuerzos de medición, así como la dificultad de identificar el carácter preciso de las preocupaciones públicas antes de producir una medida. Si se debe desarrollar un método nuevo es necesario además tener en cuenta el tiempo necesario para los ajustes por ensayo y error. Aun si ya existen algunos antecedentes de investigación y cierto consenso alrededor de un problema, es probable que pasen al menos 10 años entre el deseo de contar con una medida y la producción efectiva de esta (De Neufville, 1975, pp. 241-244).

Pensando en términos de *sistema de indicadores*, y no uno u otro aislado, no debemos esperar a tener perfectamente probado el primero para emprender el desarrollo del siguiente. Los tiempos de que habla De Neufville no deberán multiplicarse por el número de indicadores a desarrollar, pero tampoco podrán acortarse. Por ello, el desarrollo de un sistema de indicadores maduro es tarea de, al menos, una década.

Por otra parte, debe estar claro que los indicadores son sólo una herramienta que puede apoyar la toma de decisiones, pero que nada puede reemplazar el papel de autoridades responsables competentes.

Sheldon afirma que “con cada movimiento de indicadores, el entusiasmo creció y disminuyó, a medida que el optimismo sobre lo que podrían conseguir los indicadores dejó el lugar a la realidad de lo que efectivamente consiguieron, tanto en la educación como, más generalmente, en la sociedad” (según Shavelson *et al.*, 1989, p. 2).

Como muestran esos especialistas, tras varias experiencias frustrantes las expectativas sobre lo que pueden aportar o no los sistemas de indicadores son más realistas. La literatura muestra que se ha llegado a un consenso en cuanto a que los indicadores no pueden, por sí mismos, fijar objetivos o prioridades, evaluar programas o establecer balances. Todo ello puede apoyarse en la información de un buen sistema de indicadores, pero implica otras acciones. Las expectativas sobre los indicadores que hay ahora en los sistemas educativos con mayor experiencia al respecto, son bastante modestas: pueden servir “para describir y plantear problemas con mayor precisión; para detectarlos más tempranamente; para tener pistas sobre programas educativos prometedores y cosas similares” (Shavelson *et al.*, 1989, pp. 7-8).

Por otra parte, al presentar la reedición de su obra de 1975, Innes de Neufville (2002) señala que casi todos los aspectos técnicos de su libro siguen vigentes, pero que ahora se enmarcan en una visión de la relación entre conocimiento y toma de decisiones más rica que la que prevalecía en la década de 1970, todavía marcada por una concepción ingenuamente positivista, en la que los indicadores eran el fruto de un trabajo técnico puro, sin interferencia de intereses de grupo o extraacadémicos.

En 2002, la autora contextualizaba sus ideas técnicas sobre los indicadores en una concepción más compleja de la relación entre conocimiento y acción, con apoyo en el pragmatismo de principios del siglo XX, las teorías interpretativas, las concepciones de Berger y Luckmann, y las de Habermas. En esta perspectiva, el conocimiento útil no surge sólo del trabajo técnico, sino que se mezcla, de modo inexorable, con puntos de vista que privilegian unos aspectos de la realidad y se confrontan o complementan de manera compleja. La utilidad de un sistema de

indicadores para apoyar decisiones no depende en forma exclusiva de su solidez técnica, sino también de la legitimidad que le dé la participación de actores varios en su construcción y de la riqueza del proceso de construcción mismo, con aportes técnicos, discusiones que aclaren puntos difíciles y consensos más o menos amplios laboriosamente alcanzados.

Las reflexiones anteriores sobre la necesidad de invertir recursos y tiempos considerables para desarrollar un buen sistema de indicadores, no deben desalentar a quienes se dediquen a esas tareas. Todo camino, por largo que sea, comienza con los primeros pasos:

...se presta considerable atención a la imprecisión de muchos datos estadísticos, a las distorsiones incorporadas sistemáticamente en algunos casos, a los frecuentes errores de interpretación y a la creciente posibilidad de manipular los datos [...] se necesitan mejoras tanto en la cantidad como en la calidad de las estadísticas [...] La conclusión parece ser que en vez de no hacer nada, es preferible comenzar con datos malos, advertir a todos sobre sus defectos y limitaciones, y buscar una mejora gradual gracias a su uso (Gross, 1966, p. xvi).

REFERENCIAS BIBLIOGRÁFICAS

- Anderson, G. L. y K. Herr (1999). *The New Paradigm Wars*, *Educational Researcher*, vol. 28 (5), pp. 12-21.
- Bandura, R. (2006). *A Survey of Composite Indices Measuring Country Performance: 2006 Update*. United Nations Development Programme-Office of Development Studies.
- Bauer, R. A. (ed.) (1966). *Social Indicators*. Cambridge, MIT.
- Biderman, A. D. (1966). Social Indicators and Goals, en Bauer, 1966, pp. 68-153.
- Bulmer, M. (1986). *The Chicago School of Sociology*. Chicago: The University of Chicago Press.
- CERI-INES (1991). *General Assembly on International Education Indicators Handbook*. París: OCDE.
- Comité Presidencial de Investigación sobre Tendencias Sociales (1933). *Recent Social Trends in the United States*. New Cork: McGraw Hill.
- Comisión Europea (1999). *Selection and use of indicators for monitoring & evaluation. Evaluating socio-economic programmes*, vol. 2. Luxembourg: Publications of the European Communities.
- Gross, B. M. (1966). Preface. A Historical Note on Social Indicators, en Bauer, 1966, pp. ix-xviii.
- Innes de Neufville, J. (2002). *Knowledge and Public Policy: The Search for Meaningful Indicators*. New Brunswick. Transaction Publ. 2a edición de *Social Indicators and Public Policy. Interactive Processes of Design and Application*. Amsterdam: Elsevier Scientific Publ. Co. 1975.
- Jaeger, R. M. (1978). About educational indicators: Statistics on the conditions and trends in education, *Review of Research in Education*, vol. 6, pp. 276-315.
- Martínez Rizo, F. (1993). La polémica sobre la cuantificación en el campo de las ciencias del hombre, *Papers* (Univ. Autónoma de Barcelona), núm. 42, pp. 13-34.
- _____. (1991). The Controversy about Quantification in Social Research, *Educational Researcher*, vol. 20 (december), núm. 9, pp. 9-12.
- Martínez Rizo, F. et al. (2007). *Propuestas y experiencias para desarrollar un sistema*

- de indicadores educativos*. México: INEE.
- Oakes, J. (1986). *Educational Indicators: A Guide for Policymakers*. New Brunswick: Center for Policy Research in Education, Rutgers University; Rand Corporation; Univ. of Wisconsin.
- Odden, A. (1990). Educational Indicators in the United States: The Need for Analysis, *Educational Researcher*, June-July, pp. 24-29.
- OECD-European Commission (2008). *Handbook on Constructing Composite Indicators. Methodology and User Guide*. París: OECD.
- Puryear, J. M. (1993). Status and Problems of International Educational Statistics and Research. *CIES Newsletter*, núm. 103, p. 4.
- Shavelson, R. J. y L. Towne (eds.) (2002). *Scientific research in education*. Washington: National Academy Press.
- Shavelson, R. J., L. McDonnell y J. Oakes (1991). Steps in designing an indicator system, *Practical Assessment, Research & Evaluation*. Electronic Journal. ERIC.
- _____ (eds.) (1989). *Indicators for Monitoring Mathematics and Science Education. A Sourcebook*. Santa Monica: Rand.
- Special Study Panel on Education Indicators (1991). *Education Counts. An Indicator System to Monitor the Nation's Educational Health*. Washington: National Center for Educational Statistics-USA Dpt. of Education.
- Tuijnman, A. y N. Bottani (eds.) (1994). *Making Education Count. Developing and Using International Indicators*. París: OCDE.
- Wyatt, T. (1994). Education indicators: a review of the literature, en Tuijnman y Bottani, 1994, pp. 99-121.