

Calidad y evaluación: matrimonio del cielo y el infierno

HÉCTOR VERA* | MIGUEL ALEJANDRO GONZÁLEZ-LEDESMA**

El presente artículo desarrolla dos tesis: 1) a más de 30 años de políticas sobre evaluación educativa, se ha conformado un *ecosistema de la evaluación* donde los elementos que lo componen se afectan mutuamente, pero sin una lógica coherente; 2) no se ha mostrado que exista relación causal entre la reingeniería del gasto público y la “calidad educativa”; pero los instrumentos de política pública actúan sobre la capacidad de agencia de los actores educativos y crean lógicas diferenciadas de *gestión de la evaluación*. El artículo está dividido en cuatro partes: 1) antecedentes sobre la relación entre evaluación y “calidad”; 2) imbricación de las ideas de evaluación, calidad y cuantificación; 3) instrumentos de política pública y lógicas de gestión pública relacionados con la evaluación; 4) críticas y propuestas para pensar de manera más adecuada el problema de la evaluación y los cambios que serían aconsejables en las prácticas y programas de evaluación.

Palabras clave

Evaluación
Calidad
Cuantificación
Políticas públicas
Educación

* Investigador del Instituto de Investigaciones sobre la Universidad y la Educación (IISUE) de la Universidad Nacional Autónoma de México (UNAM) (México). Doctor en Sociología y Estudios Históricos por la New School for Social Research (EUA). Miembro del Sistema Nacional de Investigadores, nivel II. Líneas de investigación: sociología de la educación; sociología del conocimiento; evaluación cuantitativa del mérito académico. Publicación reciente: (2018) “Breaking Global Standards: The anti-metric crusade of American engineers”, en D. Pretel y L. Camprubí (coords.), *Technology and Globalisation: Networks of Experts in World History*, Nueva York, Palgrave, pp. 189-215. DOI: https://doi.org/10.1007/978-3-319-75450-5_8. CE: hectorvera@unam.mx

** Becario posdoctoral en el Departamento de Investigaciones Educativas del Centro de Investigación y Estudios Avanzados (DIE-CINVESTAV) del Instituto Politécnico Nacional (IPN) (México). Doctor en Ciencias Políticas por la Scuola Normale Superiore (Italia). Miembro del Sistema Nacional de Investigadores, candidato. Líneas de investigación: gobernanza y políticas de educación superior; educación superior privada; movimientos estudiantiles; análisis cualitativo comparado. Publicación reciente: (2018, en coautoría con Germán Álvarez), “Marketing Context and Branding Content of Private Universities in Chile and Mexico”, en A. Papadimitriou (ed.), *Competition in Higher Education Branding and Marketing. National and global perspectives*, Cham (Suiza), Palgrave-Macmillan, pp. 37-62. CE: miguel.ledesma@cinvestav.mx

INTRODUCCIÓN

¿Qué pasaría si evaluáramos todos los programas y políticas del gobierno federal como se evalúa a la educación y a los maestros y directivos que forman parte de la Secretaría de Educación Pública (SEP)? Pongamos por ejemplo el deporte, ¿cuánto mejoró la salud de la población con los programas de la Comisión Nacional de Cultura Física y Deporte (CONADE)? ¿Cuántos jóvenes se mantuvieron alejados de las drogas gracias a las nuevas canchas de básquetbol construidas en el estado de Nuevo León? Es, por supuesto, algo difícil de saber con exactitud.

Podemos saber cuántas medallas se ganaron en una competencia internacional, y se puede comparar ese número con la cantidad de medallas obtenidas en pasadas ediciones de esa competencia, pero reducir una evaluación de políticas nacionales del deporte a ese indicador es esencialmente absurdo. Ganar o perder medallas es, muchas veces, un hecho contingente. Además, las medallas en deportes de alto rendimiento no dicen nada de muchas otras dimensiones del deporte y la cultura física del país. Quizá las políticas de la CONADE en un sexenio fueron exitosas para promover que más personas incrementaran su actividad física cotidiana, pero eso no tendrá ningún reflejo en las visitas al podio de los mexicanos en las Olimpiadas; y ése es el indicador imperfecto con el que, en última instancia, se evaluará el desempeño del director de la Comisión.

El problema de muchos indicadores es que miden con rigor y exactitud dimensiones irrelevantes de aquello que se quiere evaluar. Pero ante la incapacidad de poder medir *lo importante* (¿se puede verdaderamente medir lo importante?), los evaluadores —y los consumidores de evaluaciones (periodistas, lectores de la prensa noticiosa, organizaciones no gubernamentales, organismos internacionales, académicos, etcétera)— prefieren conformarse con una medición exacta, aunque intrascendente, que aceptar la incertidumbre de no poder medir lo sustancial.

La sed de métricas, sin embargo, no se expresa sólo del lado de quienes tienen la responsabilidad de formular políticas públicas. Todos nosotros, quienes más quienes menos, sumamos esfuerzos en la colonización numérica de la vida cotidiana. Desde el hecho de llevar un teléfono “inteligente” que monitorea constantemente nuestros hábitos en línea (los sitios que frecuentamos, el tipo de interacciones que tenemos con la publicidad, etcétera), hasta el acto deliberado de medir nuestro desempeño físico (el número de kilómetros que corremos al día, las calorías quemadas, el tiempo de sueño, entre muchos otros), social (cuántos “amigos” en Facebook, y cuántas interacciones tuvo un tuit, por ejemplo) y hasta laboral (cuántas personas visitan o interactúan con mi perfil de LinkedIn).

Además de participar voluntariamente en esta nueva lógica de interacción social, tratamos de incidir en la *calidad* de los datos que generamos, ya que el valor de esta información constituye una llave de acceso a un mundo cuyas reglas, recompensas y castigos tienen un impacto cada vez más visible en nuestras vidas. Y, al hacerlo, corremos el riesgo de que nuestras

apreciaciones acerca de la información que compartimos sean tan absurdas como la pretensión de evaluar las políticas nacionales de promoción del deporte a través del número de medallas obtenidas por los atletas olímpicos. Sin embargo, por encima de la capacidad de individuos e instituciones para cuantificar “correctamente” lo sustancial, está la necesidad de asignarle un valor a toda costa, pues, de lo contrario, eso que *no medimos* corre el riesgo de no llegar a existir socialmente (Comité Invisible, 2017). La diferencia, en todo caso, es que mientras los individuos interactúan con este nuevo contexto social en función de los recursos a su disposición (económicos, educativos, sociales y simbólicos, entre otros), los administradores de las redes sociales, las empresas de evaluación, las instituciones y los tomadores de decisiones (a través de políticas públicas), establecen las *reglas* que estructuran dichas interacciones.

La voluntad cuantificadora está conformando una nueva forma de socialización que involucra de manera activa —y asimétrica— tanto a las instituciones como a los individuos, y que se fundamenta en dos lógicas, una de asignación de *valor* y otra de índole *política*.

En primer lugar, el *valor* al que nos referimos es por fuerza contingente: se puede conocer sólo a partir de la comparación entre objetos (o sujetos) de la misma clase. Para obtener el valor, se debe partir de una hipótesis causal sobre las cualidades o el desempeño de ese algo (o alguien) que se está evaluando. En esencia se trata —y esto es importante subrayarlo— de hipótesis acerca de qué es un *individuo exitoso*; una política del deporte *eficiente*, o la *calidad* de la educación. Curiosamente, el carácter especulativo de la hipótesis desaparece apenas se cuenta con *datos*, mas no porque éstos la demuestren, sino porque el acto de obtenerlos se realiza como un imperativo, un estándar.

El Instituto Tecnológico de Massachusetts (MIT, por sus siglas en inglés) y la Universidad de Stanford, de acuerdo con QS World University Ranking-2018, o Corea del Sur y Finlandia en la prueba PISA, son ejemplos de entidades cuyo alto valor coincide íntegramente con lo que *son* en relación con el resto de los sujetos/objetos evaluados. No obstante, el estándar permanecerá siempre por encima de quienes más se le acerquen, al tiempo que los mejor calificados corren el riesgo permanente de ser defenestrados por sus competidores. Además, un estándar (o, mejor, sus creadores), se encuentra en competencia con otros cuyas hipótesis e indicadores son distintos, de manera que el MIT y Stanford, primeros lugares para QS, aparecen por debajo de Oxford y Cambridge en el *ranking* elaborado por Times Higher Education-2018.

En segundo lugar, el hecho de no poder establecer una correspondencia numérica confiable entre lo sustancial y su valor no sólo desnuda la subjetividad inherente al acto de evaluar, justamente ahí donde se pretende más impersonal y objetivo; además, coloca las consecuencias de esta práctica bajo una luz del todo distinta a la de su justificación pública. Con la evaluación se propone corregir los supuestos problemas derivados de la falta de correspondencia entre los intereses individuales y el “interés general”, y

ofrecer criterios para la distribución de incentivos y desincentivos que promuevan o eviten ciertas conductas. Al respecto, Buendía *et al.* (2017) describen cómo, a partir de 1990, las políticas de estímulos implementadas en las universidades autónomas desplazaron la relación directa entre el quehacer de los académicos y los ejes de la misión universitaria (docencia, investigación, difusión de la cultura), para imponer una racionalidad meritocrática en nombre de la calidad académica.

De acuerdo con estos autores, se creía que la calidad de las instituciones de educación superior (IES) estaba vinculada al número de académicos de tiempo completo con estudios de posgrado. En consecuencia, se promovió que los profesores obtuvieran posgrados. ¿Ha mejorado la calidad en las universidades en las últimas tres décadas gracias estas políticas? Depende dónde se mire. Si nos enfocamos en el incremento de académicos con posgrado (lo cual sería tautológico), se podría decir que sí; pero si observamos el número de estudiantes titulados en el mismo periodo de tiempo, diríamos que no tanto. Es claro que la discusión cambiará de acuerdo con el indicador (o hipótesis sobre la identidad de lo sustancial) en cada caso. Lo que sí sabemos es que estas políticas produjeron “un segundo régimen y tabulador salarial que gobierna la actividad académica en las instituciones. Y, aunque es voluntario, es un hecho que la mayoría del personal académico de tiempo completo acude a su convocatoria” (Buendía *et al.*, 2017: 201).

Donde quiera que se mire, la implementación de estímulos en el mundo académico se ha justificado públicamente con argumentos no-económicos a pesar de que sus consecuencias sí lo sean. No podemos saber a ciencia cierta si ha mejorado la calidad de la docencia y la investigación al amparo de programas como el Sistema Nacional de Investigadores (SNI), pero es evidente que su implementación ha terminado por estratificar —económica y simbólicamente— a los académicos del país.¹ La *lógica del valor*, en ese sentido, es una pieza fundamental en las nuevas formas de asignación de recursos públicos (y, cada vez más, privados) hacia la educación. Pero ¿cómo es que se terminó estableciendo un nexo causal entre la reingeniería del gasto y la calidad educativa?

La respuesta está estrechamente relacionada con la *lógica política*. La obsesión con la métrica está reconfigurando el tipo de relaciones entre los actores y —de acuerdo con lo que se dice— son los fines que debería perseguir la educación en nuestro país. Dichas relaciones, surgidas al amparo de las políticas públicas implementadas desde hace poco más de tres décadas, dependen de la creación de una identidad entre los intereses (individuales) de los actores y las *reglas* del juego establecidas por los decisores políticos. De esta manera, las políticas de evaluación han terminado por interponerse entre los sujetos y el hecho educativo, imposibilitando —quizá definitivamente— la creación de un horizonte compartido sobre los fines que debería perseguir la educación en México.

¹ La jerarquización de los científicos y académicos que son parte del SNI crea una idea de inclusión diferenciada, pero también de exclusión: aquéllos que no forman parte de este grupo padecen las consecuencias económicas y simbólicas de “no cumplir con el perfil”.

A lo largo del presente artículo desarrollaremos dos tesis vinculadas entre sí:

1. De la triada *calidad, evaluación y cuantificación* ha surgido un complejo aparato de control entre la sociedad y la educación que busca mitigar nuestra desconfianza en la capacidad de la primera por obtener lo que se propone alcanzar a través de la segunda. Sin embargo, a poco más de 30 años de políticas destinadas a la evaluación de individuos, programas e instituciones, creemos que más que un sistema integrado y con reglas claras, se ha conformado un *ecosistema de la evaluación*, donde si bien los distintos elementos que lo componen se afectan mutuamente, éstos comparten un espacio al que llegaron sin premeditación.
2. No hay evidencia de que exista una relación causal entre la reingeniería del gasto y la “calidad educativa”, pero los instrumentos de política pública desplegados bajo esa premisa actúan sobre la capacidad de agencia de los actores educativos y crean lógicas diferenciadas de *gestión de la evaluación*. En efecto, luego de comparar los instrumentos de evaluación y mejoramiento de la calidad más significativos que se aplican en los niveles básico y superior, mostramos que el grado de coerción es mucho mayor en el primero que en el segundo, y que divide a sus actores —de acuerdo con su capacidad de agencia— en “operadores” del sistema educativo y “maximizadores” de la carrera académica.

El artículo está dividido en cuatro partes: en la primera trazamos algunos antecedentes sobre la relación entre evaluación y “calidad”, e introducimos el concepto de *ecosistema de la evaluación* para caracterizar los distintos tipos de evaluación que permean al sistema educativo mexicano; en la segunda parte mostramos cómo se fueron vinculando, en el discurso y la práctica, las ideas de evaluación, calidad y cuantificación; en la tercera describimos los instrumentos de política y lógicas de gestión pública relacionados con la evaluación; y en la última sección presentamos algunas críticas y propuestas en el sentido de cómo se podría pensar de manera más adecuada el problema de la evaluación, y qué tipo de cambios serían aconsejables en las prácticas y programas de evaluación.

ANTECEDENTES Y ECOSISTEMA DE LA EVALUACIÓN

Antecedentes

En los últimos 40 años la evaluación educativa ha pasado de ser un instrumento de carácter fundamentalmente diagnóstico, que se aplica en algunos países, a convertirse en una práctica de alcance mundial, considerada como un mecanismo necesario para la gobernanza de los sistemas educativos. La evaluación ha experimentado cambios importantes, desde sus orígenes como herramienta para el diagnóstico, hasta ampliar sus funciones

como *medio para incentivar o disuadir conductas*. Un antecedente común al cambio de rumbo en las políticas educativas a nivel mundial ha sido la incapacidad de los gobiernos de incrementar el financiamiento requerido por este sector a la par de las necesidades para su ulterior desarrollo y crecimiento. Al respecto, las acciones llevadas a cabo en cada país para enfrentar esta situación han dependido de las características y necesidades de sus sistemas educativos, pero, sobre todo, de su situación ante el contexto económicamente adverso experimentado a partir de la década de 1970 (González-Ledesma, 2014).

En los países occidentales el cambio se da en un momento donde la universalización de la educación básica y media se había conseguido al menos una década antes, al tiempo que los niveles de acceso a educación superior eran muy elevados (en torno al 40 por ciento, UNESCO, 2018). En estos países los primeros cambios se experimentaron precisamente en el nivel superior. Así, para la década de 1980, Estados Unidos ajustó el gasto en investigación de modo que sólo 20 instituciones tuvieran acceso a la mayor parte de los fondos federales destinados a ello. Paralelamente, varios países de Europa occidental (el Reino Unido *in primis*) optaron por fomentar una mayor estratificación dentro del mundo académico, y condicionaron el acceso a fondos públicos otrora de disposición directa e inmediata (Paradeise y Thoenig, 2017). En ambos contextos se trató de una reingeniería del gasto público como parte de proyectos más amplios de reforma, que contemplaban —entre otras cosas— tiempos, diseño de políticas, ajuste de instrumentos y, sobre todo, objetivos de mediano y largo plazo.

Diez años más tarde, se registraron importantes iniciativas de evaluación en el nivel de educación básica. En Estados Unidos, los estados de Texas, Maryland y Carolina del Norte fueron pioneros en la aplicación de exámenes de desempeño escolar como condición para otorgar fondos públicos a las escuelas, experiencia que a partir del 2001 se difundiría en todo el país, tras la aprobación de la legislación federal *No Child Left Behind* (Hennaway y Woodroffe, 2003). Paralelamente, en España se aprobó la Ley Orgánica General del Sistema Educativo (LOGSE), que dio la pauta para una reforma radical del sistema. Y es que, entre otras cosas, la LOGSE contemplaba la creación del Instituto Nacional de Calidad Educativa (INCE), como la entidad responsable de evaluar la educación no universitaria en el Estado español (OCDE, 2014).

En el caso de México, el cambio en materia de evaluación estuvo relacionado inicialmente a las condiciones económicas que impuso la crisis de la deuda de principios de los años ochenta. Los recortes al gasto público, que se implementaron como medida de emergencia para amortiguar la debacle económica, se llevaron a cabo sin una reprogramación mínima de los recursos disponibles. La crisis vino a agravar el rezago educativo de nuestro país, que apenas había logrado universalizar el acceso a la educación primaria, aunque en condiciones dramáticas de desigualdad (Martínez, 1992). Por su parte, la educación superior, que había experimentado un crecimiento importante de la matrícula hasta 1980, arrastraba sus propios

problemas en cuanto a burocratización, escasa diferenciación institucional, bajas tasas de egreso, desarrollo de la investigación, etc.²

No es, pues, sino hasta principios de los años noventa que nuestro país comenzó a esbozar una reforma que contemplaba la necesidad de implementar formas modernas de evaluación. Sin embargo, al igual que en otros países de América Latina, en un primer momento se trataba de prácticas dispersas, y en su mayoría destinadas a paliar las consecuencias económicas de la crisis que se venía arrastrando desde la década anterior. Tal es el caso de la administración de la demanda de educación media y superior a través de pruebas estandarizadas como el Examen de Habilidades y Conocimientos Básicos (EXHCOBA), creado en 1992; o los exámenes diseñados y aplicados por el Centro Nacional para la Evaluación de la Educación Superior (CENEVAL), a partir de 1994 (Backhoff *et al.*, 2000).³

Además, con base en la experiencia del Sistema Nacional de Investigadores (SNI), creado en 1984, se experimentaron formas novedosas de reingeniería del gasto salarial de docentes e investigadores universitarios, a través de programas de sobresueldos vinculados a la productividad. Es el caso del Programa de Estímulos al Desempeño del Personal Docente (ESDEPED), de 1992; el Programa Nacional de Superación del Personal Académico (SUPERA), en 1994, y el Programa de Mejoramiento del Profesorado (PROMEP), que entró en vigor en 1996 en sustitución de SUPERA (Buendía *et al.*, 2017).

En nuestro país, la difusión de prácticas evaluadoras es mayor en el nivel superior con respecto al básico. Esta diferencia se debe a que, desde el gobierno de Miguel de la Madrid (1982-1988), los recortes al gasto público fueron más agresivos en las universidades y otras IES que en la educación preescolar, primaria y secundaria. Sin embargo, los primeros intentos por crear un sistema que institucionalizara la práctica de la evaluación se dieron en el nivel básico, en el marco del Programa para la Modernización Educativa (PME), que guio la política sectorial durante el sexenio de Carlos Salinas de Gortari (1988-1994). El PME no sólo proyectó la idea de la evaluación con visión de sistema, sino que, además, la concebía por primera vez como un mecanismo para *mejorar la calidad* (Buendía *et al.*, 2017). Fue hasta el gobierno de Ernesto Zedillo (1994-2000), con el Programa de Desarrollo Educativo 1995-2000, que se establecieron las bases para la modernización del área de evaluación de la Secretaría de Educación Pública, al reconocer que los indicadores

2 Entre 1970 y 1979 la matrícula en la educación superior aumentó 213 por ciento, es decir que se pasó de 271 mil a 848 mil estudiantes. Entre 1980 y 1989 —en el contexto de la crisis económica— el aumento de la matrícula fue de tan sólo 34 por ciento (Rodríguez y Ordorika, 2011).

3 En 1994 se creó el Centro Nacional para la Evaluación de la Educación Superior (CENEVAL), cuya función es la de elaborar y aplicar exámenes de admisión para el bachillerato y la licenciatura. El EXHCOBA fue desarrollado en 1992, por parte de un grupo de investigadores de la Universidad Autónoma de Baja California (UABC) y de la Universidad Nacional Autónoma de México (UNAM), para seleccionar a los “mejores estudiantes que desean ingresar a la UABC y, hasta donde fue posible, a otras instituciones de educación superior”. Hoy en día la prueba EXHCOBA es aplicada por Métrica Educativa, una asociación civil que vende servicios de evaluación estandarizada a varias instituciones educativas del país. Ver: <http://metrica.edu.mx> o Backhoff *et al.*, 2000: 13.

...educativos existentes son resultado, antes que nada, del levantamiento de la estadística escolar —matrícula y número de maestros, grupos y escuelas— al principio y al final del ciclo lectivo. Estos indicadores reflejan el interés por *medir* lo que durante muchos años fue lo primordial, es decir, el crecimiento de la cobertura. No obstante, resultan insuficientes para hacer una evaluación completa del sistema educativo, principalmente en el aspecto de la *calidad* (Programa de Desarrollo Educativo 1995-2000, 1996. Énfasis añadido).

El eje rector de la política educativa durante el sexenio de Vicente Fox (2000-2006) fue el Programa Nacional de Educación, cuyo proyecto evaluativo se basó en dos ejes: el Compromiso Social por la Educación y la creación del Instituto Nacional para la Evaluación Educativa (INEE). Ambas iniciativas contaron con el apoyo del Sindicato Nacional de Trabajadores de la Educación (SNTE) y de su lideresa, Elba Esther Gordillo, y junto con el Programa Sectorial de Educación del gobierno de Felipe Calderón (2006-2012), mantuvieron una línea de continuidad con respecto a los ejes del PME de Salinas de Gortari.

La novedad de las políticas sectoriales de Fox fue la intención de crear un sistema de evaluación para la educación superior a través del Consejo Acreditador de la Educación Superior (COPAES), un sistema de evaluación de pares, y el Programa Integral de Fortalecimiento Institucional (PIFI), que buscaba anclar, definitivamente, el financiamiento a la evaluación (Buendía, 2013).⁴

Ecosistema de la evaluación

En México predomina lo que aquí definimos como *ecosistema de la evaluación*, término con el que aludimos a la difusión no coordinada de las instituciones, convenciones y prácticas de evaluación en todos los ámbitos del sistema educativo: evaluaciones institucionales, de programas, de individuos, etcétera. Este concepto es relevante porque el panorama de la evaluación se ha vuelto tan complejo que resulta difícil saber quién exactamente está evaluando, quién o qué es evaluado, con qué propósito se evalúa, a quién le importan los resultados de las evaluaciones y quiénes sufrirán las consecuencias del resultado de las evaluaciones.

Esto se debe, en parte, a que la evaluación de las instituciones, actores y acciones educativas en México —y, en mayor o menor grado, en cualquier otro país— no es resultado de un sistema cerrado y centralizado de evaluación diseñado para que todas sus partes constitutivas se complementen y se entrelacen de manera coherente. Lo que existe, más bien, es lo que

⁴ El antecedente del PIFI foxista fue la Comisión Nacional de Evaluación (CONAEVA), creado en 1989 durante el gobierno de Carlos Salinas de Gortari. De acuerdo con Angélica Buendía, el propósito de la Comisión “era formular y desarrollar la estrategia nacional para la creación y operación del Sistema Nacional de Evaluación a partir de tres líneas de acción: la autoevaluación de las instituciones, la evaluación del sistema y los subsistemas a cargo de especialistas e instancias, y la evaluación interinstitucional de programas académicos” (2013: 20).

podríamos denominar un “ecosistema”, donde si bien los distintos elementos que lo componen se afectan mutuamente (a través de cooperación, competencia o influencia indirecta), éstos comparten un espacio al que llegaron de manera accidental —no premeditada— mediante trayectorias independientes. Así, por ejemplo, para muchas universidades es importante saber cuál es su posición en los *rankings* internacionales, lo mismo que conocer el factor de impacto de las revistas donde publican sus investigadores; pero los *rankings* y el factor de impacto (aunque a veces se utilice el segundo como ítem de evaluación de los primeros) fueron creados y son administrados por entidades desvinculadas, que persiguen fines divergentes.

Entender plenamente las evaluaciones educativas requiere, pues, tener una visión panorámica del ecosistema de la evaluación. Y comprender que las partes que componen este ecosistema están en un equilibrio cambiante de relaciones de poder y de vínculos de interdependencia.

Hay varias formas posibles de bosquejar el ecosistema de la evaluación en México. Una manera que podría ser útil y clara es clasificar las distintas evaluaciones a las que están sujetos los múltiples actores educativos en las cuatro categorías que utilizan las instituciones que diseñan y/o aplican las evaluaciones: 1) instituciones educativas; 2) agencias de gubernamentales; 3) organismos internacionales; 4) empresas privadas y ONG.

Como resultado de esta agrupación, tendríamos un panorama (ecosistema) que luciría más o menos de esta forma:

1. Evaluaciones realizadas por las instituciones educativas. Aquí se encuentran todas las pruebas, exámenes, comités, tabuladores y reglamentos de evaluación que son diseñados e implementados de manera autónoma por cada institución educativa (cada escuela, colegio, academia, instituto y universidad en particular). La variedad en el tipo de evaluaciones es formidable: puede ir desde las escuelas primarias cuyo modo de evaluar a sus profesores y trabajadores se limita a los criterios idiosincráticos del director o del encargado de recursos humanos de la institución, hasta programas perfectamente reglamentados en las IES, donde se manejan fondos multimillonarios como parte de la distribución de sobresueldos vinculados al pago por mérito —un ejemplo de lo último sería el Programa de Primas al Desempeño del Personal Académico de Tiempo Completo (PRIDE), de la UNAM—. Entre estos extremos hay muchísimas variantes, donde se pueden contar: los cuestionarios que diseña cada institución para que los alumnos califiquen a los profesores (que en muchos casos es el factor preponderante —a veces único— para determinar la posible recontractación de los docentes); los exámenes de admisión para aceptar nuevos alumnos; los procedimientos para otorgar becas a los alumnos; los requisitos para obtener un grado o título profesional, etcétera. Dada su variedad —por el inmenso número de instituciones educativas en el país, y sus múltiples tipos y naturaleza—, ésta es la dimensión menos homogénea del ecosistema.

2. *Evaluaciones realizadas por agencias gubernamentales.* En esta categoría se concentran las evaluaciones que llevan a cabo las múltiples dependencias de los gobiernos estatales y federal relacionadas —completa o parcialmente— con la educación. Fungen aquí como evaluadores lo mismo la SEP que los cada vez más abundantes órganos autónomos descentralizados. El tipo y número de estas evaluaciones —si bien abundante y compleja en su naturaleza— es más reducido que el de la categoría anterior; pero las *consecuencias* y el *número de afectados* por cada programa de evaluación es mucho mayor (pueden llegar a ser cientos de miles).

Considerar solamente algunos de los principales programas y entidades vinculados con la SEP del gobierno federal sería suficiente para apreciar la profundidad que pueden alcanzar estas evaluaciones. Por ejemplo, en el Plan Nacional para la Evaluación de los Aprendizajes (prueba Planea) —que antes fue la Evaluación Nacional de Logros Académicos en Centros Escolares (prueba ENLACE)— participan todas las escuelas públicas y privadas de nivel básico; se aplica una serie test para determinar en qué medida los estudiantes logran dominar un conjunto de aprendizajes básicos en español y matemáticas. También se ubica aquí el Concurso de Asignación a la Educación Media Superior, que organiza la Comisión Metropolitana de Instituciones Públicas de Educación Medio Superior (COMIPEMS), un órgano descentralizado de la SEP; se trata de una prueba que presentaron, en 2018, más de 300 mil aplicantes que aspiraban a entrar a una institución pública de educación media superior en la Ciudad de México y área metropolitana. Está, igualmente, el Programa Integral de Fortalecimiento Institucional (PIFI) con el que la SEP distribuye recursos para “lograr mejores niveles de *calidad* en sus programas educativos y servicios”⁵ entre las universidades públicas estatales (y otras instituciones afines) que logran obtener dictámenes favorables de parte de un comité de pares académicos externos a la institución.

También forman parte de estas evaluaciones varios programas del Consejo Nacional de Ciencia y Tecnología (CONACyT); uno de los más conocidos es el Sistema Nacional de Investigadores (SNI) y sus programas de becas para estudiantes de posgrado. El SNI evalúa periódicamente a aproximadamente 28 mil profesores-investigadores de tiempo completo para otorgarles el reconocimiento simbólico de “investigador nacional” y estímulos monetarios diferenciados por niveles según su pretendido desempeño (principalmente, la publicación de artículos en revistas académicas indizadas). Las becas de posgrado son asignadas anualmente a miles de estudiantes —unos en el extranjero y otros en México— cuando están enrolados en uno de los más de 2 mil programas de posgrado que pertenecen al Padrón Nacional de Posgrado de Calidad (PNPC). El CONACyT evalúa periódicamente tanto a los estudiantes (a través de informes anuales) como a cada programa de posgrado individual (es decir, una universidad con varios programas pertenecientes al PNPC es evaluada tantas veces como programas tenga).

⁵ Puede consultarse en: <https://goo.gl/z5gsjq> (consulta: 17 de octubre de 2018).

Otro tipo de evaluaciones que pertenece a esta categoría son los reportes y mediciones que realizan organismos públicos autónomos. Algunos de ellos no tienen la encomienda explícita de evaluar temas educativos, como el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), pero dado que entre sus objetivos está el de generar información sobre políticas sociales y medición de la pobreza, evalúan numerosos programas e instituciones vinculados con la educación. Otro de estos organismos, el Instituto Nacional para la Evaluación de la Educación (INEE), ha estado en el corazón mismo de la relación entre política pública, calidad y evaluación durante el último sexenio. El INEE se ha encargado de operar uno de los engranes centrales de la reforma educativa:

definir los métodos e instrumentos de evaluación destinados a medir las capacidades y aptitudes de quienes aspiran a ingresar al magisterio, así como de quienes se proponen obtener el ascenso a cargos de dirección y de supervisión en los ámbitos de educación básica y media superior.⁶

A la fecha, de acuerdo con el propio INEE, se han evaluado cerca de 1.1 millones de docentes en alguno de los cuatro procesos que regula la ley del Servicio Profesional Docente.⁷

La masividad y el alto impacto que sus resultados tienen sobre los evaluados, hacen de las evaluaciones llevadas a cabo por agencias gubernamentales la columna vertebral del ecosistema evaluativo en México. Las consecuencias potenciales de estas evaluaciones son —irónicamente— incommensurables. Están en juego las trayectorias educativas de niños y jóvenes; la continuidad laboral de más de cientos de miles docentes y directivos de educación obligatoria; los recursos económicos que recibirán escuelas e IES; las becas para los estudiantes de posgrado; los complementos salariales, así como los símbolos de estatus de científicos y profesores-investigadores de tiempo completo, etc.

3. *Evaluaciones realizadas por organismos internacionales.* Éstas son menores en número, pero sus consecuencias no son insignificantes. Aquí se incluyen evaluaciones planeadas por organismos de cooperación internacional que buscan intercambiar información y armonizar políticas entre los países miembros para intentar asegurar su crecimiento económico, como la Organización para la Cooperación y el Desarrollo Económicos (OCDE). Con la intención de tener información estandarizada y comparable sobre el rendimiento de los alumnos en matemáticas, ciencia y lectura, la ODCE realiza el Informe del Programa Internacional para la Evaluación de Estudiantes (Informe PISA). Dado que los resultados de la prueba PISA se presentan públicamente para comparar el desempeño de varios países, esto produce una presión permanente para que cada nación mejore (o al menos mantenga) su posición en el *rankeo* internacional. Cada vez que los

6 Puede consultarse en: <https://goo.gl/2waoHE> (consulta: 17 de octubre de 2018).

7 Puede consultarse en: <https://goo.gl/75oAAa> (consulta: 17 de octubre de 2018).

resultados son publicados se presentan en la prensa y en la discusión pública demandas por mejorar la posición del país en los resultados de esa prueba (por ejemplo, en 2016 México ocupó el lugar 57 en ciencias, el 55 en habilidad lectora y el 56 en matemáticas). Obtener mejores lugares en la prueba PISA se convierte, así, más que en una guía, en un fin por sí mismo a la hora de planear y justificar acciones gubernamentales para la educación.

Otro tipo de organismo internacional que evalúa la educación son las organizaciones financieras internacionales, como el Banco Mundial y el Banco Interamericano de Desarrollo, que realizan préstamos condicionados a los países y publican reportes y recomendaciones para influenciar las políticas educativas.

4. *Evaluaciones realizadas por empresas privadas y organizaciones no gubernamentales.* Estas empresas y organizaciones privadas (mexicanas o extranjeras) son de muy distinto tipo: empresas de la información, de entretenimiento, de servicios educativos, editoriales, asociaciones sin fines de lucro y *think tanks*. Siguiendo fines y métodos poco homogéneos entre sí, estas entidades realizan o facilitan la infraestructura para que se lleven a cabo muchas evaluaciones en el medio educativo. Están, por ejemplo, los célebres *rankings* internacionales de universidades hechos anualmente por empresas periodísticas, como *Times Higher Education* y *U.S. News & World Report*. El principal objetivo de quienes realizan estos *rankings* muchas veces se reduce a vender más ejemplares de su publicación, pero uno de sus efectos ulteriores incluye una aguerrida competencia internacional (Enders, 2015) entre universidades de todo el mundo por mejorar su posición en esos listados jerárquicos y estar en una mejor posición para allegarse estudiantes de élite, recursos monetarios y prestigio. Por otra parte, la posición que se obtiene en estos *rankings* se convierte posteriormente en objeto de queja o aplauso público y en justificación para distribuir o modificar el presupuesto público que reciben las IES públicas.

Empresas editoriales como Elsevier y Clarivate Analytics, por su parte, administran bases de datos bibliográficas de citas de artículos de revistas científicas (Scopus y Web of Science, respectivamente). Éstas y otras bases de datos sirven para estratificar a las revistas académicas y a los autores de artículos a través de índices y factores de impacto. La importancia de estas mediciones y cálculos de citas se ha incrementado con el paso del tiempo, pues cada vez se les utiliza más como una fuente de información para sustentar ulteriores evaluaciones a las IES y a los profesores-investigadores. Estos índices de citación —y otras mediciones derivadas de ellos— se emplean en programas internos de estímulos económicos de las IES, en programas públicos (como el SNI) y en los mismos *rankings* de universidades.

Otro tipo de compañías que ha entrado a la evaluación de aspectos de la educación son empresas de entretenimiento que promueven la calificación de los profesores universitarios como parte de una estrategia para producir contenidos y atraer flujos de visitantes a sus portales de Internet. Es el caso, por ejemplo, de Viacom, a la que pertenece el canal de videos musicales y de

programación juvenil MTV; esta empresa es dueña del portal ratemyprofessors.com, donde estudiantes de universidades de Estados Unidos y el Reino Unido pueden calificar a los profesores con los que tomaron clase. La versión hispanoparlante de ese portal —misprofesores.com— ha tenido una notable penetración en México. Este programa cuenta con listas de más de 3 mil instituciones (casi todas ellas IES, aunque se incluyen algunos bachilleratos) y publica las evaluaciones que han recibido decenas de miles de profesores; se muestran sus nombres, los cursos que imparten y la calificación que cada estudiante les otorga. Tan sólo para la UNAM, misprofesores.com exhibe las evaluaciones de más de 20 mil docentes. Estas evaluaciones, sobre las cuales las IES no tienen ninguna injerencia, y se realizan sin ningún control metodológico, llegan a influir en aspectos relevantes de las instituciones educativas, como la elección de cursos que realizan los alumnos, la imagen pública de los maestros o, incluso, su (re)contratación.

Finalmente, hay que mencionar otros organismos privados que se dedican a evaluar diversos aspectos de la educación. Se puede mencionar, por ejemplo, a México Evalúa, un *think tank* dedicado a “elevar la efectividad y *calidad* de la gestión gubernamental a través del análisis del diseño de las políticas públicas”.⁸ Y, también, organizaciones de acción política como Mexicanos Primero, que enlista entre sus principios algunos eslóganes como “La verdadera independencia está en la educación de *calidad*”.⁹ Igualmente, hay asociaciones civiles sin fines de lucro, como el Centro Nacional de Evaluación para la Educación Superior (CENEVAL), la cual realiza una amplia gama de exámenes —entre ellos, pruebas de cobertura nacional— que evalúan el nivel de conocimientos y habilidades académicas de los recién egresados de múltiples áreas de especialidad.

Si consideramos en conjunto las evaluaciones que hemos presentado, es fácil apreciar que todos los actores del sistema educativo están atrapados dentro de una espesa y desorganizada red de evaluaciones.¹⁰ Es importante recalcar que esta telaraña no fue tejida por una sola araña, extraordinaria y voraz; es más bien un apretado tejido que se ha ido formando mientras numerosas arañas, de manera no coordinada, contribuyeron con unos cuantos hilos que se han entrelazado con el tiempo, sofocando cada vez más a quienes se encuentran dentro. No estamos, pues, dentro de un panóptico (una construcción centralizada de control y vigilancia) ni de una jaula de hierro (una prisión mantenida por la racionalización y la burocratización). Se trata, más bien, de una pegajosa telaraña de seda. Las acciones no coordinadas de numerosas instituciones (locales, nacionales e internacionales; públicas y privadas; comerciales y sin fines de lucro) actúan siguiendo fines pertinentes y racionales desde su propio punto de vista y desde su propia

⁸ Puede consultarse en: <https://goo.gl/ypUdrP> (consulta: 17 de octubre de 2018).

⁹ Puede consultarse en: <https://goo.gl/zgteYs> (consulta: 17 de octubre de 2018).

¹⁰ Y no se debe olvidar que, en esta incompleta lista, ni siquiera hemos mencionado la práctica de evaluación más añeja y abarcadora del mundo educativo: las evaluaciones que reciben los alumnos de sus maestros (usualmente en la forma de calificaciones numéricas y la posterior construcción de los promedios de calificación). Es una práctica tan naturalizada y dada por supuesta, que a pesar de su enorme importancia casi nunca se le considera cuando se hacen reflexiones sobre el fenómeno de la evaluación.

escala de prioridades, pero terminan formando un (eco)sistema cuya lógica de operación y cuyas consecuencias prácticas no fueron buscadas o planeadas por nadie en particular.

Sin embargo, independientemente de la ausencia de una coordinación gubernamental de la práctica evaluativa o, mejor, una coherencia propia de un sistema de gobernanza de la educación, lo que tenemos son niveles de *agencia* (i.e., capacidad de transformar, en menor o mayor medida, el contexto y las condiciones que en las se vive y actúa) por parte de los sujetos de la evaluación, que aumentan de acuerdo con el nivel educativo del que se trate. En otras palabras, consideramos que los profesores en IES cuentan con mayor grado de agencia que los profesores de educación obligatoria.¹¹ No obstante, lo que nos interesa resaltar aquí es cómo las prácticas de evaluación han cambiado radicalmente la relación entre los individuos y la consecución de los fines que, se dice, deberían perseguir nuestros sistemas educativos.

Sobre la “calidad” de la educación en México

Durante el sexenio de Enrique Peña Nieto la “educación de calidad” se convirtió en un eje propagandístico de primer orden. En los documentos y planes del gobierno federal más importantes, la idea de que el país necesitaba una “educación de calidad” se repetía constantemente. Así quedó reflejado en el Plan Nacional de Desarrollo, documento concebido como el programa rector del poder Ejecutivo federal en cada sexenio. Ahí se detallan los objetivos, estrategias y prioridades que el gobierno se planteó a sí mismo, y se justifican sus acciones futuras. El Plan se elaboró en el primer semestre del sexenio y se presentó públicamente como una serie de metas que se buscaba alcanzar en los siguientes años de gobierno. Todas las acciones posteriores de los cientos de dependencias del ejecutivo federal quedaron justificadas por sus supuestas contribuciones a ese plan.

La “educación de calidad” jugó un papel tan preponderante en la administración del presidente Peña Nieto, que quedó definida como una de las “cinco metas nacionales” en su Plan Nacional Desarrollo (PND): 1) México en paz; 2) México incluyente; 3) *México con educación de calidad*; 4) México próspero; y 5) México con responsabilidad global. La meta educativa fue explicada diciendo que:

El futuro de México depende en gran medida de lo que hagamos hoy por la educación de nuestros niños y jóvenes. Por tanto, es fundamental que la nación dirija sus esfuerzos para transitar hacia una sociedad del conocimiento. Un México con educación de calidad propone implementar políticas de Estado que garanticen el derecho a la educación de calidad para todos los mexicanos, fortalezcan la articulación entre niveles educativos, y los vinculen con el quehacer científico, el desarrollo tecnológico y el sector productivo, con el fin de generar un capital humano de calidad que detone la innovación nacional (PND, 2013: s/p).

¹¹ En este artículo nos ocupamos casi exclusivamente de ellos, pero debemos mencionar que la evaluación de programas y contenidos es de suma relevancia.

Esta enunciación, y su afición por el término “calidad”, no se redujo a una floritura verbal. Si bien la revisión de las acciones del gobierno deja claro que aumentar la calidad de la educación con un proyecto meditado y pertinente no estaba entre las prioridades (o no estaba al alcance de sus capacidades reales); la invocación a la “calidad” se convirtió en una coartada para justificar las acciones más evidentes del gobierno federal en el área educativa: domesticar al SNTE y apretar el yugo laboral sobre los maestros, mediante la redefinición de las reglas del juego en lo que respecta a su contratación y promoción. El eufemismo con el que el gobierno federal hacía referencia a estas acciones era la necesidad de “recuperar la rectoría de la educación” por parte del Estado.

El mismo mes (febrero de 2013) en que fue encarcelada la lideresa histórica del SNTE, Elba Esther Gordillo, se publicó en el *Diario Oficial de la Federación* un decreto que modificó la Constitución Política de los Estados Unidos Mexicanos. El artículo tercero constitucional se reformó en varios puntos y recibió numerosas adiciones; entre ellas, se agregó el siguiente párrafo:

El Estado garantizará la *calidad en la educación obligatoria* de manera que los materiales y métodos educativos, la organización escolar, la infraestructura educativa y la idoneidad de los docentes y los directivos garanticen el máximo logro de aprendizaje de los educandos (énfasis añadido) (Decreto por el que se reforma..., 2013).

Se subraya que el criterio que orientará a la educación “será de *calidad*, con base en el mejoramiento constante y el máximo logro académico de los educandos” (énfasis añadido).

Otra adición al tercero constitucional —materializada ese mismo mes— sostiene que

Para garantizar la prestación de *servicios educativos de calidad*, se crea el Sistema Nacional de Evaluación Educativa. La coordinación de dicho sistema estará a cargo del Instituto Nacional para la Evaluación de la Educación [INEE]... Corresponderá al Instituto *evaluar la calidad*, el desempeño y resultados del sistema educativo nacional en la educación preescolar, primaria, secundaria y media superior (énfasis añadido).

Entre los deberes del INEE, dice desde entonces la Constitución, está:

Generar y difundir información y, con base en ésta, emitir directrices que sean relevantes para contribuir a las decisiones tendientes a *mejorar la calidad de la educación* y su equidad, como factor esencial en la búsqueda de la igualdad social (énfasis añadido).

También se realizaron otros cambios para que la “calidad educativa” quedara presente en otras disposiciones legales que regulan la educación.

La Ley General de Educación, por ejemplo, se reformó en noviembre de 2013, en varios de sus artículos; entre ellos:

- El artículo segundo: “*Todo individuo tiene derecho a recibir educación de calidad* en condiciones de equidad, por lo tanto, todos los habitantes del país tienen las mismas oportunidades de acceso, tránsito y permanencia en el sistema educativo nacional”.
- El artículo tercero: “*El Estado está obligado a prestar servicios educativos de calidad* que garanticen el máximo logro de aprendizaje de los educandos”.
- El artículo octavo: “*El criterio que orientará a la educación* que el Estado y sus organismos descentralizados impartan... *será de calidad*, entendiéndose por ésta la congruencia entre los objetivos, resultados y procesos del sistema educativo, conforme a las dimensiones de eficacia, eficiencia, pertinencia y equidad” (énfasis añadido) (Ley General de Educación, 2013).

CALIDAD, EVALUACIÓN Y CUANTIFICACIÓN: FORJAR UN TRIÁNGULO

“Calidad”: el ariete retórico

Las alusiones a la “calidad” de la educación en los niveles más altos de la pirámide jurídica mexicana (Constitución y Ley General de Educación) necesitaban, por supuesto, una traducción más específica para convertirlas en algo que diera una dirección clara y puntual al quehacer de cada uno de los múltiples actores del sistema educativo para conseguirla. Esta tarea, sin embargo, resultó elusiva. La “calidad” era frecuentemente invocada, pero rara vez definida. Desde el INEE se hicieron intentos para llenar de contenido pedagógico a lo que ya se había convertido en una categoría de combate político. Al final del sexenio, la consejera presidenta de la junta de gobierno del INEE publicó en la *Gaceta* del Instituto un texto que resumía la postura de éste para aclarar qué se debería de entender por ese término. En él sostiene que el “ piso básico ” para evaluar la calidad educativa incluye cuatro principios:

1. *Principio de universalidad.* Todas las personas deben tener acceso a la escuela y permanecer en ella hasta concluir su educación obligatoria.
2. *Principio de equidad.* No debe existir entre las personas diferencias de acceso, permanencia o logro educativo en función de su género, grupo étnico, adscripción cultural, nivel socioeconómico, nacionalidad o cualquier otro motivo.
3. *Principio de logro.* Las personas deben desarrollar las mismas competencias (o equivalentes), en los mismos niveles y en cada punto del sistema educativo, cualesquiera que sean los contenidos, conocimientos y valores que un sistema nacional se proponga enseñar.
4. *Principio de suficiencia y calidad de la oferta.* Para que los principios anteriores se cumplan,

el Estado tiene la obligación de generar recursos humanos capacitados y con las condiciones materiales adecuadas a tal fin (Bracho, 2018: 25).

Pese a los intentos por darle un sentido técnico y sustantivo al concepto, nunca se borró la impresión de que “calidad educativa” era, principalmente, un arma de choque retórico con la que el gobierno buscaba derribar las barreras que se alzaban para detener sus planes de reforma. Estudiosos del fenómeno educativo, como Arnaut (2017; 2018), llegaron a sostener, con sorna, que la “calidad” se había convertido en una especie de deidad que está en todas partes y en ninguna; invocada como un pensamiento mágico, su único criterio de existencia eran los exámenes de opción múltiple aplicados masivamente a los docentes: “Se inventaron un nuevo ídolo, un nuevo dios, el de la calidad, que ni el INEE alcanza a definir, y al que sólo se puede acceder y tener contento mediante el sacrificio del magisterio en la piedra de los sacrificios de la evaluación” (Arnaud, 2017: s/p).

Un matrimonio a la fuerza: calidad y evaluación

En medio de la andanada retórica de la “calidad” se realizó un maridaje de mal agüero entre la “calidad” y una de las prácticas ubicuas del mundo educativo contemporáneo: la evaluación. En el discurso de funcionarios de la SEP, de políticos de los partidos que aprobaron las leyes de la “reforma educativa”, y de miembros de algunas organizaciones no gubernamentales, se asumió —sin mayor reflexión o justificación— que el camino más corto, o acaso el único, para “alcanzar” la calidad, era a través de la evaluación. Este credo quedó nítidamente resumido en uno de los principios de Mexicanos Primero: “Lo que no se evalúa, no se puede mejorar”. Y si bien desde antes del sexenio de Peña Nieto ya había una marejada evaluativa bien entrada en cauce, las reformas legislativas —y la campaña mediática que las acompañó— le dieron un impulso aún más decidido. En la práctica esto significó, en lo que respecta a los niveles de la educación obligatoria, evaluar a los docentes y atar su futuro laboral a los resultados de esas examinaciones.

Así, la compulsión evaluadora que se vive actualmente en la educación —y en muchos otros ámbitos: empresarial, deportivo, médico, etcétera— está anclada en un par de dogmas (Vera, 2016): el primero es la *pantometría*, el supuesto de que “todo es medible” y “todo debe ser medido”; el segundo, hermanado con el anterior, es la *mesurofilia*, el amor por la medición, que se traduce en la creencia de que medir es la forma más exacta y válida de conocimiento. La ubicuidad evaluativa se ha traducido en acciones masivas y de consecuencias profundas para el mundo de la educación: desde la adjudicación de becas escolares según el *promedio de calificación* de los alumnos, hasta la repartición de sobresueldos para científicos por publicar en revistas con un alto *factor de impacto*, pasando por la admisión de alumnos a instituciones de educación media superior de acuerdo con el *número de aciertos* en un examen estandarizado. Programas nacionales enteros, a todos los niveles del sistema educativo, se han articulado alrededor de estos principios. Algunos de los más visibles son el Examen Nacional de

Conocimientos, Habilidades y Competencias Docentes (que se encuentra entre los más recientes de estos esquemas) y el Sistema Nacional de Investigadores (que es uno de los más añejos).

“Calidad” y evaluación se convirtieron en las puntas de una pinza ideológica y política que atenaza a numerosos actores del sistema educativo. Su poder se basa, entre otras cosas, en que nadie puede pronunciarse en contra de la calidad. ¿Quién podría decir, razonablemente, que no desea que la educación mejore? Pero ahí reside la efectividad de esta trampa. La calidad es un horizonte inalcanzable, está siempre “un poco más allá”. En palabras de la exconsejera presidenta de la junta de gobierno del INEE: “la calidad nunca se alcanza totalmente: siempre es posible proponerse metas más elevadas” (Schemelkes, 2018: 19). Nada es perfecto; todo puede ganar en calidad. La calidad es una fuga hacia adelante. Y si vemos a la evaluación como un camino para “alcanzar” la calidad, pero la calidad es intrínsecamente inalcanzable, entonces las evaluaciones serán eternas y, previsiblemente, cada vez más abarcadoras e intrusivas. Esto está sucediendo en México y en muchos otros países donde los actores educativos también están siendo azotados con el látigo evaluativo. Cualquier evaluación mostraría que la Sorbona (una de las universidades más influyentes del planeta) *podría* ser una mejor universidad, o que la enseñanza de las matemáticas en Corea (que en las pruebas PISA se encuentra notoriamente por arriba del promedio de los países miembros de la OCDE) *podría* ser más efectiva. Esto justifica y alimenta el llamado a seguir evaluando porque todavía no se ha “llegado” a la calidad.

Este efecto de perpetua incompletitud de la calidad, convierte a la calidad educativa en una instancia discriminatoria. Como ha señalado Plá al hablar de “inclusión diferencial”:

...todo dispositivo de calidad debe discriminar qué posee calidad y qué no. Si todo es excelente, nada lo es... La discriminación se ejerce a través de la racionalidad instrumental que define técnicamente lo que es calidad y lo que no lo es, estableciendo a su vez los niveles de desviación y déficit de desviación de los diferentes sujetos en relación con los estándares establecidos por un grupo de especialistas. En este sentido, la inclusión diferencial requiere expertos tanto para la producción de la desigualdad como para la autorreproducción del sistema, pues son ellos quienes establecen su conocimiento como superior, aplican los criterios y determinan qué es calidad de aprendizaje y conocimiento (2018: 28-29).

Pese a las afinidades que pueden compartir, la relación entre calidad (educativa) y evaluación (educativa) no es autoevidente ni necesaria. En su sentido amplio, *calidad* refiere a un atributo o propiedad de un objeto que es considerado superior o excelente. Se puede poseer ese atributo o propiedad sin tener que pasar necesariamente por una evaluación (y, sobre todo, una evaluación estandarizada). Las universidades de Salamanca o Estambul fueron excelentes centros de enseñanza décadas o siglos antes de

que se diseñaran los indicadores y *rankings* con los que se evalúa y compara hoy a las universidades del mundo. Un objeto, una práctica, un procedimiento pueden cumplir cabal y efectivamente las tareas para las que fueron concebidas sin que tengan que ser sometidas a una determinación exhaustiva de su mérito o valía según una serie de criterios simplificados.

Por supuesto, vincular calidad con evaluación no es irremediablemente artificioso. Las evaluaciones *pueden* servir para diagnosticar cuáles son las áreas donde el espacio para mejorar es mayor. Una buena evaluación puede ser *una* de las herramientas para encontrar el modo de optimizar algo. Pero la evaluación no es indispensable para la calidad ni, mucho menos, un elemento suficiente para el perfeccionamiento.

Los funcionarios al frente de la política educativa no pensaban así. Los secretarios de Educación Pública durante el último sexenio recalcaron frecuentemente que, en la educación, la calidad necesitaba, obligatoriamente, de la evaluación. Emilio Chuayffet, el primer encargado de la SEP del último sexenio (entre 2012 y 2015), decía:

La reforma [educativa] es la confirmación del compromiso del gobierno de México con la educación pública, laica, gratuita, incluyente y de calidad... La evaluación, si queremos mejorar al sistema educativo de México, necesita ofrecernos con objetividad cuáles son sus puntos débiles para que, con base en valoraciones puntuales, puedan tomarse decisiones inteligentes y precisas para corregir (Chuayffet, 2015: s/p).

Aurelio Nuño, quien fuera secretario entre 2015 y 2017, y que se convirtió en el promotor más notorio de la política peñista en materia educativa, centró su defensa de las reformas siguiendo las mismas líneas generales:

Gracias a la Reforma Educativa... hoy, podemos contar con un sistema de educación que, además de ser laico, gratuito y obligatorio, es hoy también un derecho de todos los niños, las niñas y los jóvenes, que sea de calidad... La evaluación, y esto es muy importante decirlo, no es un fin, sino medio para garantizar que todos los niños y jóvenes reciban una educación de calidad. La evaluación ha hecho posible que estemos transitando de un sistema opaco, algo contrario e injusto a uno que privilegia la dedicación personal, el esfuerzo personal y el mérito de los maestros... ¿Para qué se evalúa? La evaluación nos permite valorar si el desempeño de docentes y directivos alcanza el nivel suficiente para asegurar una educación de calidad. De manera más específica, la evaluación nos ayuda a identificar las debilidades y fortalezas de los maestros con el fin de proporcionales programas de formación continua y desarrollo profesional dirigidos a mejorar sus prácticas de enseñanza, liderazgo y gestión escolar (Nuño, 2016: s/p).

El discurso público de los secretarios de Educación presentaba a la “evaluación” y a la recuperación estatal de la “rectoría de la educación” como los medios para obtener la “calidad”. Pero las acciones gubernamentales

contaban otra historia. Pronto se hizo evidente que la evaluación sí era un medio, pero no para aumentar la calidad de la educación, sino para que el gobierno federal pudiera controlar más metódicamente al magisterio; la “calidad”, en este plan, era un fin meramente manifiesto para justificar las acciones gubernamentales.

Un dato revelador respecto de la discrepancia entre los pretendidos fines y las acciones reales es el contraste entre el ejercicio del gasto de la SEP en comunicación social y en formación de profesores: en 2017, la Secretaría gastó 1,960 millones de pesos para publicitar la reforma educativa (lo que significó un sobreejercicio de 2,680 por ciento, pues el Congreso de la Unión únicamente había aprobado una partida de 70.6 millones de pesos); pero sólo destinó la mitad del presupuesto previsto para capacitar a maestros (a la SEP le fueron aprobados 1,654 millones de pesos para capacitación docente, de los cuales sólo gastó 949 millones de pesos, lo que implicó un subejercicio del 42 por ciento) (Salazar, 2018; Roldán, 2018). Por otra parte, la reforma en la evaluación de los docentes (aunque en efecto fuera para el alegado fin de “identificar las debilidades y fortalezas de los maestros”) reducía la calidad al espacio del aula y, de manera aún más simplista, a la figura del profesor.

De matrimonio a ménage à trois: calidad, evaluación y cuantificación

En congruencia con la visión generalizada internacionalmente, en México se ha aceptado una triple convicción sobre la naturaleza del acto de evaluar: 1) “todo es medible”; 2) “lo que no se puede medir, no se puede mejorar”; 3) “las evaluaciones más adecuadas son cuantitativas y estandarizadas”.

El tercero de estos puntos hace eco de un viejo precepto que William Thompson (mejor conocido como Lord Kelvin) hizo famoso en el siglo XIX: cuando no puedes medir algo, o no puedes expresarlo con números, “tu conocimiento es pobre e insatisfactorio” (1889: 73). Por supuesto, la mayoría de las veces que se cita esta línea como justificación para intentar cuantificar el análisis de los fenómenos humanos, se descontextualiza el viejo precepto de Kelvin, quien hablaba exclusivamente de la “ciencia física”; y subraya que encontrar principios numéricos y aplicar métodos de medición es sólo un “primer paso para el conocimiento” (Thompson, 1889: 73). Esa advertencia de Kelvin no ha sido tomada en cuenta. Hoy se piensa, de manera acrítica, que el conocimiento expresado con números es una meta en sí misma (en vez de un simple punto de partida) y que todos los fenómenos (no sólo los físicos) deben ser interrogados a través de métodos cuantitativos. Las consecuencias de esta tergiversación han sido numerosas y profundas.

En un trabajo reciente sobre lo que llama la “tiranía de las métricas”, Jerry Muller ha sintetizado los hallazgos sobre los efectos de la cuantificación en la vida social, incluida la educación,¹² que se han realizado en

12 El estudio señero sobre este tema es: Porter, 1995. Sobre las consecuencias generales de la cuantificación: Espeland y Stevens: 1998; O’Neil, 2016. Sobre los efectos de la cuantificación en la educación: Espeland y Sauder: 2007; 2016. Sobre el caso de la educación en México: Vera, 2017a; 2017b.

ciencias sociales desde los años noventa. Para dar cuenta de este problema, Muller acuñó el término de “obsesión métrica” (*metric fixation*) para nombrar la expectativa de que es deseable —y posible— sustituir el juicio personal (adquirido por medio de la experiencia y el talento individuales) con indicadores numéricos que comparan el desempeño de distintos actores o instituciones a través de información estandarizada (Muller, 2018: 18).

La obsesión métrica tiene dos características iniciales: primero, la creencia de que hacer públicas las mediciones de una evaluación es una garantía de que las instituciones rinden cuentas y cumplen sus propósitos; segundo, la suposición de que la mejor manera de motivar a quienes laboran en una institución es vincular recompensas (o castigos) con los resultados de sus evaluaciones de desempeño, recompensas que pueden ser pecuniarías (“pago por mérito”) o de reputación (nombramientos, premios, etc.).

Esto último implica lo que se conoce como “pago por mérito”: ofrecer a los trabajadores incentivos financieros que cumplan ciertos criterios cuantitativos. Esta estrategia laboral, que no siempre ha probado ser efectiva en organizaciones cuyo fin explícito es la ganancia económica, ha tenido peores resultados en instituciones que persiguen fines más idealistas (como las escuelas, universidades u hospitales). Donde sea que la evaluación está unida a premios monetarios se invita a la simulación, o a que se haga trampa en la medición.

Las principales deficiencias de la obsesión métrica son bien conocidas: en primer lugar, no todo lo que es importante es medible (o medible de manera adecuada) y frecuentemente lo que sí se puede medir rigurosamente, no es relevante. Las instituciones y los distintos tipos de trabajo tienen múltiples facetas, pero las evaluaciones (principalmente cuando son cuantitativas) se concentran en sólo algunos de sus aspectos, lo que incita a los actores a descuidar aquellas dimensiones de su labor que no están consideradas en la evaluación. Esto explica, por mencionar un ejemplo de sobra conocido, por qué muchos profesores en IES desdeñan las labores docentes (que pesan poco en las evaluaciones que determinan sus sobresueldos) y prefieren concentrarse en multiplicar sus publicaciones (que es el ítem mejor recompensado en la mayoría de las evaluaciones).

Otros problemas de la obsesión métrica son condensados por Muller, quien cita dos “leyes”: por un lado, la ley de Goodhart, que sostiene que “cualquier medida usada para controlar es poco fiable” (Muller, 2018: 20); y, por otro lado, la ley de Campbell: “mientras más se utiliza un indicador social cuantitativo para tomar decisiones, más estará sujeto a corromperse, y más propenso será a distorsionar y pervertir el proceso social que intenta monitorear” (Campbell, 1976: 49). El autor de este principio, Donald Campbell, lo ejemplificaba con el caso de las mediciones educativas diciendo que:

...las pruebas de aptitud pueden ser indicadores valiosos de logros escolares generales *bajo condiciones de enseñanza normales enfocadas en la capacidad general*. Pero cuando los exámenes se convierten en el fin del proceso de

enseñanza, pierden su valor como indicadores educativos y distorsionan el proceso educativo de maneras indeseables. Sesgos similares se presentan en el uso de pruebas objetivas utilizadas en cursos, o como exámenes de admisión (1976: 51-52).

Muller también subraya que cuando hay una falta de progreso real en lo que se busca mejorar, el progreso en la medición es disfrazado como un éxito. Es típico que se presenten los *fin*es de una política, y que después se justifiquen las acciones de los gobernantes citando *mediciones sobre los medios* (no sobre los fines mismos). Así hacía, por ejemplo, Otto Granados Roldán (el último secretario de educación del sexenio de Peña Nieto), quien dijo ante la UNESCO que

Si el gran logro del siglo XX fue por una cobertura universal en la educación básica, ahora la gran batalla del siglo XXI es por la calidad con equidad e inclusión... Como secretario de Educación de mi país celebro que, a cinco años de iniciada la reforma educativa, México avanza de manera firme hacia esos objetivos (Granados, 2018: s/p).

Los resultados que Granados presentó a continuación no fueron sobre cómo ha mejorado la calidad, la equidad o la inclusión, sino cuánto dinero se invirtió diariamente en la infraestructura de las escuelas y cuántos maestros fueron contratados o ascendidos mediante evaluaciones y concursos:

Estamos invirtiendo cerca de 4 millones de dólares diarios para mejorar la infraestructura de 33 mil escuelas en las que estudian 6 millones de alumnos, principalmente en las regiones de mayor rezago y comunidades indígenas. Casi 190 mil maestros han sido contratados o ascendidos mediante evaluaciones y concursos basados en el mérito (Granados, 2018: s/p).

Lo que se invierte en los *medios* para un fin, se presenta como los resultados de la política; y la finalidad última que se perseguía (la calidad educativa) queda olvidada a la hora de rendir cuentas concretas.

Una característica final de la obsesión métrica es que —paradójicamente— quienes predicán la fe en las evaluaciones desestiman los estudios que demuestran la falta de efectividad de las evaluaciones y piensan que las carencias de una medición se resuelven aplicando aún más evaluaciones (Muller, 2018: 20). La parte obsesiva de esta fijación evaluativa se observa en que sus proponentes se adhieren a ella a pesar de las consecuencias negativas de poner en práctica sus ideas. Es común que, cuando las instituciones se percatan de que hay inconvenientes con las evaluaciones, su respuesta no sea eliminar, reducir o ajustar las evaluaciones, sino aumentar el número de mediciones y cuantificar cada vez más cosas. Lo que esta escalada métrica produce, usualmente, no es un sistema más eficaz, sino mayor inversión de tiempo, trabajo y dinero para poner en marcha cada vez más evaluaciones. Con eso se crean burocracias evaluadoras-calificadoras-certificadoras

cada vez más costosas, sin que aumente la “calidad” de las instituciones, servicios o prácticas evaluadas. En resumen, el dinero y tiempo invertido en evaluar aumenta de manera sostenida, sin que haya evidencias de que mejore la calidad (Muller, 2018: 19).

Pese a que tales problemas son conocidos, el peso de la obsesión métrica es tal que podemos decir que hoy día *se prefiere medir mal, que no medir*. Esto pasa, entre otras cosas, porque cuando una autoridad cimienta la legitimidad de sus acciones y decisiones en ciertos tipos de prueba y evaluación, le resulta muy difícil librarse de éstas —aunque se haya hecho evidente que son defectuosas—.

Hay otro motivo importante que ayuda a comprender por qué los métodos cuantitativos de evaluación se han vuelto tan persistentes. En el discurso público de muchas sociedades contemporáneas los números son percibidos —merecidamente o no— como garantes de “objetividad”, “exactitud”, “confianza”, “racionalidad” y “universalidad”. Muchos políticos y administradores intentan revestir de autoridad sus tomas de decisión amparándose en la confianza pública que producen los números (Porter, 1995). Esto produce una demanda creciente para que se cuantifiquen cada vez más fenómenos sociales, incluida la educación (Vera, 2017a). Unido a esto, cuando las decisiones políticas son presentadas como respuesta a una medición “objetiva” (entendiendo lo “cuantitativo” de manera reduccionista, como hacen muchos, es decir, como sinónimo de “objetivo”), se produce la impresión de que se está ante juicios imparciales y se intenta despersonalizar las consecuencias de políticas que acarrearán efectos negativos.

Este panorama no es exclusivo de México; la experiencia de este país es paralela a lo que ha sucedido en otros lugares. Refiriéndose a la educación superior en el ámbito internacional, Paradeise y Thoenig (2017) muestran que se ha realizado una transición entre dos “régimenes de calidad”: de uno centrado en la reputación, a otro de “juicios basados en la excelencia”. En el segundo, los juicios sobre la calidad de la educación se enfocan en la estandarización de las medidas de desempeño y en el uso de grandes bases de datos con las que se organizan y jerarquizan dichas medidas. Esta transición ha implicado una contraposición entre las prácticas endógenas y las tradiciones de evaluación del trabajo académico en las IES, y un nuevo régimen exógeno de evaluación (o, como ellos dicen, “cuando los números definen la calidad académica”). Se trata de un “régimen exógeno” porque los criterios de evaluación, la selección de indicadores y el diseño de los instrumentos se realiza en entidades externas a las universidades, y muchas veces incluso ajenas al mundo académico (como los *rankings* realizados por revistas comerciales). La expresión más manifiesta de este cambio es la relevancia que han adquirido los *rankings* internacionales de universidades. Cada vez más los presupuestos de las universidades varían de acuerdo con sus resultados en esos índices; las IES dedican cada vez más recursos a mejorar en los aspectos particulares que se privilegian en esos indicadores y tienden a desatender aquellas áreas de su quehacer que no son consideradas por ellos. La difusión y consumo de la información de los *rankings*

ha tenido un enorme éxito por su simplicidad y su aparente transparencia; pero uno de los resultados prácticos de esto —y la consecuente comparación internacional de universidades— es usar como rasero un modelo único, homogéneo y estandarizado de lo que debe ser la educación superior (Paradeise y Thoenig, 2017).

La carambola entre calidad, evaluación y cuantificación es producto de una serie de simplificaciones. La calidad (que es un concepto multidimensional) fue reducida, en el discurso y la práctica, a obtener buenos resultados en un proceso de evaluación. La evaluación (otro concepto multidimensional), por su parte, fue reducida a superar una cantidad mínima de puntos en exámenes cuantitativos.¹³

EVALUACIÓN E INSTRUMENTOS DE POLÍTICA PÚBLICA

El acto de gobernar, nos recuerdan Frans van Vught y Harry de Boers, significa utilizar *instrumentos* de política pública, pues sin éstos, “las políticas públicas no serían más que ideas abstractas o fantasías” (2015: 41). Al inicio señalamos cómo la evaluación se construye sobre hipótesis acerca de la eficiencia, la eficacia y la calidad de los objetos/sujetos examinados. Asimismo, advertimos que los resultados de las evaluaciones suelen esgrimirse como argumento para justificar acciones o para orillar a los individuos a que actúen de una determinada manera. Los instrumentos de políticas, en este sentido, son el *medio* a través del cual los gobiernos, las autoridades educativas u otros actores facultados por la ley, garantizan que las evaluaciones tengan *consecuencias*. En esencia, los instrumentos constituyen el mecanismo a través del cual se induce a las personas a “hacer cosas que de otra manera no harían; o elegir la posibilidad de hacer cosas que de otro modo les resultaría imposible hacer” (Schneider e Ingram, 1990: 513).

En la literatura se pueden encontrar varias clasificaciones de instrumentos de políticas públicas que se han desarrollado desde los años cincuenta del siglo XX.¹⁴ Sus diferencias radican en el alcance del análisis que se proponen (escala nacional vs ámbito específico de *policy*) y, según los estudios más recientes, en la inclusión y análisis de instrumentos que han surgido a raíz del viraje mundial hacia el libre mercado. Pero, independientemente de sus diferencias, los estudiosos coinciden en agrupar los instrumentos en familias que se distinguen, a su vez, por el *grado de coerción* que éstos

¹³ Cuando los funcionarios de la SEP presentan sus políticas públicas hablan de la calidad de manera multidimensional; pero la práctica ha quedado siempre reducida a resultados de evaluaciones unidimensionales.

¹⁴ El tema de los instrumentos de políticas se ha venido desarrollando al menos desde 1954 a partir del trabajo de Robert Dahl y Charles Linblom, *Politics, Economics and Welfare*. Sin embargo, una de las contribuciones más significativas ha sido la de Theodore Lowi (1964) en *American Business, Public Policy, Case Studies, and Political Theory*, quien propuso cuatro familias de instrumentos, vinculadas al funcionamiento de las políticas estadounidenses del estado de bienestar durante esa época, a saber: regulativos, de gestión pública directa, de incentivos, y de desincentivos. Más recientemente, Hood y Margetts (2007) retomaron esta clasificación y agregaron otros instrumentos, basados en el manejo de la “información” (*nodality*) y creación de “capacidad organizativa” (*organization*) (para una síntesis ver Van Vught y de Boers, 2015; Bobbio *et al.*, 2017).

pueden ejercer sobre la ciudadanía. Luigi Bobbio *et al.* (2017) identifican dos grandes familias de instrumentos: *sustanciales*, que se dirigen a atender de manera directa un problema público determinado; y *procedimentales*, que postergan la solución de dicho problema, pero adelantan algunas condiciones para su atención en el futuro. A éstos se suma una tercera opción dentro de la “caja de herramientas” del gobierno, que es la de no-hacer o, mejor, dejar hacer (*laissez faire*).

Cada familia se caracteriza por la capacidad de sus instrumentos constitutivos de movilizar algunos de los recursos con los que cuenta el gobierno (autoridad, fondos públicos, información, organización) (Hood y Margetts, 2007), a través de cuatro modalidades de intervención, a saber:¹⁵

- a) *Regulación / desregulación*. Se basa en la movilización de la autoridad del gobierno a través de obligaciones y prohibiciones cuya inobservancia conlleva sanciones. La conducta de los individuos e instituciones viene definida casi capilarmente a partir de órdenes, directivas, normativas y procedimientos que son controlados a través de inspecciones periódicas, así como por la vigilancia de funcionarios, burócratas, jueces, etcétera. Con la llegada de la doctrina neoliberal, muchos gobiernos han optado por desregular enteros sectores de su economía y servicios públicos, en el entendido de que la libre competencia permitirá resolver problemas que el Estado no ha logrado solventar. Ello ha implicado, por ejemplo, abrir la provisión de servicios educativos a la participación del sector privado.
- b) *Gestión directa / indirecta*. Supone el despliegue de la autoridad, capacidad de organización y administración de recursos públicos. El gobierno se involucra directamente en la resolución del problema y produce bienes o servicios (escuelas, hospitales, infraestructura, etcétera). El viraje hacia el libre mercado ha traído consigo mecanismos a través de los cuales una autoridad puede delegar a terceros estas actividades como, por ejemplo, las escuelas *charter* en los Estados Unidos, o la selección de aspirantes a la educación media superior a través del examen aplicado por el CENEVAL, que es una organización privada facultada para tal efecto (Hannaway y Woodroffe, 2003).¹⁶

¹⁵ Para efectos de la presente argumentación hemos adaptado la clasificación de instrumentos de política elaborada por Bobbio *et al.* (2017: capítulo 5, ed. Kindle). Mientras los autores atribuyen las modalidades de acción únicamente al conjunto sustancial de instrumentos, nosotros vemos que funcionan también para aquéllos de índole procedimental, ello en virtud de que su aplicación es coherente (y empíricamente demostrable) con las características atribuidas a esta segunda familia de instrumentos. Por otro lado, hemos buscado encuadrar el impacto del viraje hacia el neoliberalismo incluyendo a la “desregulación” y a la “gestión indirecta”. De hecho, notamos que la categorización de Bobbio *et al.* se desarrolla a partir de los postulados de Lowi (1964) cuando empíricamente es demostrable la existencia de experiencias significativas contrarias que se contraponen a la regulación y la gestión directa.

¹⁶ En los países occidentales, la implementación de políticas reguladoras determinó tanto la percepción pública sobre la vía para solucionar los problemas colectivos, como la retórica de los gobiernos en el periodo que va de la posguerra y hasta la década de 1970, cuando la crisis económica mundial volvió más receptivos a los decisores políticos hacia las propuestas inspiradas en el libre mercado. En el caso latinoamericano, la

- c) *Incentivos y desincentivos / directrices*. Consiste en la movilización de recursos públicos para lograr que los individuos se comporten o no de cierta manera. Ello con la intención de resolver un problema o alcanzar un objetivo a través de la agregación de conductas esperadas por parte del grupo destinatario de la política pública. Las directrices, por su parte, especifican lo más detalladamente posible las responsabilidades, atribuciones y objetivos de los individuos en los organigramas de sus respectivas instituciones.
- d) *“Codazo suave” (nudge)*.¹⁷ Se basa en el despliegue de la versión menos coercitiva de la autoridad a través de lo que Sunstein (2014) define como “paternalismo libertario”. A partir de hallazgos recientes de la psicología cognitiva, Thaler y Sunstein (2008) sostienen que, en algunos casos, para modificar el comportamiento en la dirección deseada por los decisores políticos, se debe actuar sobre la *arquitectura de la elección* de los individuos, de manera que las opciones disponibles correspondan con el objetivo que se está persiguiendo.
- e) *Nodalidad*. El término fue utilizado por primera vez por C. Hood (1983) para señalar la centralidad del gobierno (como nodo) en la acumulación, procesamiento y difusión de información. La nodalidad se ejerce a través de requerimientos gubernamentales de información sobre un problema o ámbito de *policy* a instituciones o individuos de interés; o para difundir información a través de campañas sobre algún problema público que, por diferentes razones, se considera atendible a través de la concienciación de la población.

Finalmente, ante ciertos retos, un gobierno puede simplemente optar por no hacer nada y dejar que las cosas sigan su curso (*laissez faire*). Esta elección, que no es neutral, obedece a distintas razones: por lo general se trata de una acción deliberada con la que se busca evitar temas delicados en la agenda política, ya sea porque los tomadores de decisiones consideren que no es el mejor momento para abordarlos, o simplemente porque se espera que sean “otros” (funcionarios, instituciones, gobiernos, el mercado) quienes se hagan cargo. Este tipo de acciones supone, en muchos casos, el despliegue de instrumentos que le permitan al gobierno administrar un determinado ámbito de política con acciones de tipo inercial (por ejemplo, mantener estable el presupuesto asignado a un programa de becas que no se quiere eliminar, pero tampoco hacer crecer).

regulación se da en el marco de la política de sustitución de importaciones, que prometía incrementar los estándares de vida de la población a través de medidas de redistribución de la riqueza.

¹⁷ En inglés, *nudge* quiere decir, literalmente, “empujar con el codo”, y se refiere al acto de llamar sutilmente la atención de alguien hacia alguna dirección. Hemos optado por traducirlo como “codazo suave”. Un ejemplo reciente de este tipo de políticas en nuestro país es la decisión del Senado de la República de convertir a todos los mexicanos en donadores de órganos, a menos que se manifieste lo contrario. Al establecer como opción preestablecida (*default*) la donación, se apuesta por la supuesta propensión de los individuos a preferir dejar las cosas “como están” pues, de lo contrario, la acción de oponerse requeriría hacer “algo más” ante la entidad correspondiente, lo cual representaría un costo evitable.

En la década de los noventa, por ejemplo, las autoridades educativas de nuestro país acordaron, junto con la ANUIES, una contracción deliberada del crecimiento de la matrícula en las IES públicas. Esta decisión se tomó con base en la escasez de fondos públicos, producto de la crisis económica que se venía arrastrando desde la década anterior; pero, sobre todo, ante el asombroso crecimiento de la oferta privada en este periodo. En 1990 las IES privadas pasaron de absorber 17 por ciento de la matrícula, al 29 por ciento en el año 2000.¹⁸ Se esperaba que, al *dejar hacer* a los actores privados, se podrían cumplir las expectativas de una demanda creciente, lo que permitiría un ahorro significativo de fondos públicos, pues el gobierno se habría encargado solamente de financiar el crecimiento inercial de la oferta pública.¹⁹

Tanto las modalidades de acción como el cariz particular que adquieren los recursos gubernamentales utilizados dependerán de si el problema se atiende de manera directa (*sustancial*) o indirecta (*procedimental*). La forma en que se ha tratado la cuestión del Nuevo Aeropuerto Internacional de la Ciudad de México (NAICM) entre la administración saliente y la que presidirá Andrés Manuel López Obrador es un buen ejemplo de esta dinámica. Se sabe que la capacidad del actual aeropuerto ha sido superada y que es necesario encontrar una alternativa (*problema*). El gobierno de Peña Nieto optó por construir el NACIM en Texcoco, Estado de México (*enfoque sustancial*), concesionando la obra a constructoras y contratistas privados (*gestión indirecta*), y participando con una buena parte de la inversión requerida (movilizando *fondos públicos*). Por su parte, durante su campaña, López Obrador se opuso abiertamente a este proyecto debido al elevado costo que representa lo que falta para su construcción, y a las consecuencias ecológicas para el Valle de México; y propuso como alternativa la base aérea de Santa Lucía, también en el Estado de México. Sin embargo, tras haber triunfado en las elecciones, en vez de anunciar la cancelación de la obra en Texcoco, ha convocado a expertos en la materia (*gestión indirecta*) a un debate público en el que se expongan las razones a favor y en contra de las dos opciones sobre la mesa (*nodalidad*), como preámbulo a una consulta popular que resolvería definitivamente la cuestión del lugar y, con ello, el problema (*enfoque procedimental*).²⁰

18 Entre 1980 y 1990, la matrícula en las IES públicas creció 40.8 por ciento; el año 2000 esta cifra descendió a 23.5 por ciento; por su parte, las instituciones privadas pasaron de crecer 90 por ciento entre 1980-1990 a 144 por ciento entre 1990 y el año 2000 (González-Ledesma, 2015).

19 Sin embargo, a partir del 2000, como demuestra Germán Álvarez (2011), se hizo evidente que el crecimiento de la oferta privada estaba condicionado al poder adquisitivo de sus clientes potenciales. Para paliar el déficit en la oferta, el gobierno de Vicente Fox optó por impulsar la expansión de las universidades tecnológicas.

20 Son varias las razones por las que López Obrador podría haber decidido postergar la cuestión NAICM, desde el mero cálculo político sobre el equilibrio que pretende construir con las fuerzas políticas y económicas del país sin perder el consenso popular que lo llevó al poder, hasta la intención sincera de instaurar mecanismos para la participación democrática e informada de la ciudadanía en temas de gran relevancia. Sin embargo, nuestro interés no es especular sobre las intenciones de presidente electo, sino ilustrar las características que adquieren los instrumentos de política en función de cómo se aborda el problema y qué tipo de recursos se movilizan.

Lógicas de gestión de los instrumentos de política para la evaluación educativa

En este apartado mostraremos cómo el ecosistema de la evaluación en México produce lo que aquí denominaremos *lógicas de gestión*, que combinen el tipo de enfoque con el que se aborda un problema educativo determinado, con la forma en que los instrumentos de política pública actúan sobre los sujetos evaluados. Cabe advertir que, si bien el ecosistema está lejos de constituirse como un verdadero sistema nacional de evaluación, ello no quiere decir que las entidades que lo conforman (instituciones, organismos internacionales, empresas privadas y ONG) actúen en un vacío de sentido práctico o institucional.

Un primer elemento común de las formas modernas de evaluación es que su diseño se basa en la idea de que no existe algo parecido a una identidad entre los intereses de los individuos y los fines que, supuestamente, debería perseguir la educación. La evaluación “corrige” este problema, ya sea al definir —*por encima* de los individuos— el valor de lo educativo (la mejor universidad, el plan de estudios más pertinente, el estudiante más capaz, etcétera), o al orientar —*a pesar* de los individuos— la toma de decisiones sobre los mecanismos para la asignación de recursos (por ejemplo, el ingreso al Padrón del Programa Nacional de Posgrados de Calidad [PNCP], los exámenes de selección universitaria y el PIFI). Un segundo elemento en común es, precisamente, el uso de la evaluación como mecanismo para la distribución de recursos económicos y simbólicos. Y es que, luego de establecer el valor de lo educativo y las acciones necesarias para obtenerlo, es necesario —recordando la definición de Schneider e Ingram (1990)— implementar instrumentos para persuadir u obligar a los individuos ya sea a *elegir* la posibilidad de hacer lo que se espera de ellos, o *actuar* de acuerdo con el modo en que deberían hacerlo. Las entidades evaluadoras despliegan acciones que van, desde incentivar a los actores a integrarse a los mercados del prestigio individual (como ResearchGate y [Academia.edu](https://www.academia.edu)), hasta orillarlos a interactuar con tabuladores salariales estratificados con base en el desempeño.

Ambas características se encuentran estrechamente relacionadas entre sí, en la medida que el grado de coerción es inversamente proporcional a la capacidad de agencia. Sin embargo, mientras los niveles de coerción son proyectados en el diseño de la evaluación a partir de las hipótesis sobre las que se apoya (“¿cómo se logra un buen desempeño docente?”, o “¿cómo debe de ser un artículo académico relevante?”), la capacidad de agencia está vinculada al tipo de instrumento que se utiliza para solucionar un problema o alcanzar un objetivo relacionado a dicha hipótesis (despido de los “malos” profesores, o asignación de puntajes a la calidad del artículo). Los *grados de coerción*, por lo tanto, aluden al nivel de especificación de los medios que deben utilizar los actores con relación a los fines establecidos por los decisores políticos. La *capacidad de agencia*, por su parte, no se refiere a la “libertad” sin más de los individuos, sino a las posibilidades de acción (o grados de libertad) determinadas *a priori* por los hacedores de políticas.

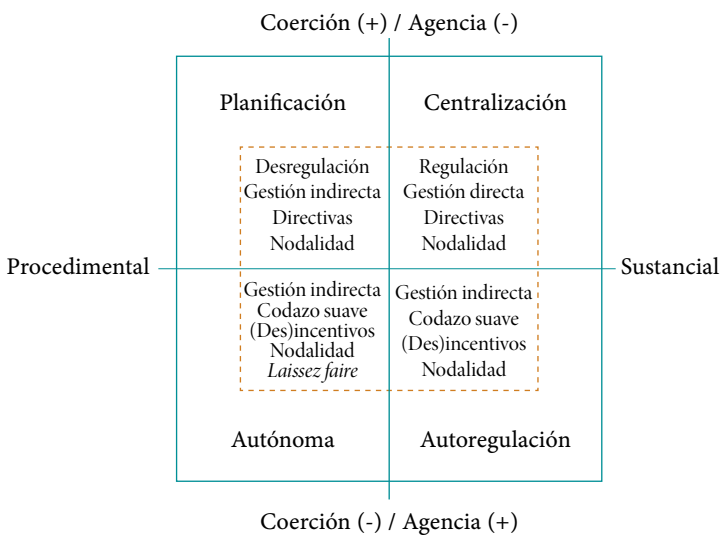
Las *lógicas de gestión* surgen de la combinación entre la dinámica generada por las dos características del ecosistema de la evaluación (+ o – coerción / + o – agencia), y el tipo de enfoque sobre el problema a resolver en cada caso (*sustancial* o *procedimental*). A su vez, cada lógica se distingue por las modalidades de intervención que despliega (regulación / desregulación; gestión directa / indirecta; incentivos y desincentivos / directrices; “codazo suave”, y nodalidad), y el tipo de recursos que moviliza (autoridad, fondos públicos, información, organización) a través de los instrumentos de política utilizados. A partir de estos elementos hemos identificado cuatro lógicas típico-ideales: a) de planificación; b) de centralización; c) de autorregulación; y d) de gestión autónoma (Figura 1).

- a) *Lógica de planificación*. El grado de coerción de esta lógica es muy alto, por lo que la capacidad de agencia de los individuos es reducida. El enfoque es procedimental: los hacedores de políticas establecen las características de los actores y/o entidades que habrán de participar en la creación de soluciones para resolver un problema determinado en el futuro. Las modalidades de intervención de esta lógica son la desregulación (políticas de descentralización educativa, eliminación de restricciones a la participación de actores privados en la educación); la gestión indirecta (permitir la conformación y operación de asociaciones como el CENEVAL, Métrica Educativa o, más recientemente, dotar de autonomía a agencias como el INEE); las directrices (establecer los tiempos y las características con las que habrán de operar políticas habilitadoras como la autonomía escolar); y la nodalidad (creación de campañas que buscan que la opinión pública acepte y, en su caso, defienda la necesidad del cambio). El carácter planificador de este arreglo alude a las acciones que han permitido la transición, desde una administración de lo educativo basada (formalmente) en el cuidado de los procesos y la asignación directa de fondos públicos, a una que se enfoca en los resultados y la distribución selectiva y condicionada de recursos o, en términos de Neave (2012), el pasaje desde el Estado proveedor al Estado evaluador.
- b) *Lógica de centralización*. El grado de coerción es mayor, por lo que la capacidad de agencia de los individuos es prácticamente nula. En este caso el enfoque es sustancial: los hacedores de políticas se involucran directamente en la resolución de los problemas. Esta lógica sigue una perspectiva racionalista de la toma de decisiones basada en la capacidad gubernamental de ejercer su autoridad para obligar a los actores a producir, entregar y procesar información (nodalidad),²¹ así como de requerir que actúen de acuerdo con las atribuciones, responsabilidades y objetivos preestablecidos (directrices). Tradicionalmente los procesos de concentración de facultades decisionales (y la centralización en sí) se asocian a las viejas

21 A través de la instauración del Sistema de Información y Gestión Educativa previsto por la Ley General de Educación del gobierno de Enrique Peña Nieto.

políticas características del Estado interventor; mientras se asume que lo opuesto (desconcentración y distribución de atribuciones) es una característica intrínseca de la administración, en clave neo-liberal, de lo público. Sin embargo, la evidencia acumulada en los últimos años demuestra que políticas de tipo gerencial o de mercado pueden llegar a ser sumamente verticales.²² En nuestro país, esta lógica se anunció a través de una retórica que sostenía la necesidad de “recuperar la rectoría del Estado en la educación” (gestión directa), a través de políticas como la Ley General del Servicio Docente (LGSD).²³ En el mismo sentido, la reingeniería del gasto público requiere de una fuerte intervención estatal, lo mismo que la definición de los beneficiarios de las asignaciones presupuestales, ya sea que se construyan criterios de mérito, necesidad o focalización.

Figura 1. Lógicas de gestión de la evaluación en México



Fuente: elaboración propia con base en Hood y Margetts, 2007; Bobbio et al., 2017.

c) *Lógica de autorregulación.* El grado de coerción de esta lógica es menor, por lo que los individuos cuentan con una amplia capacidad de agencia. El enfoque es sustancial: el gobierno establece metas genéricas vinculadas a problemas que deben ser resueltos a través de la investigación académica y científica.²⁴ Para ello delega a agencias gu-

22 Respecto del caso de Chile, por ejemplo, Alejandra Falabella (2016) utiliza el término, *Estado hipervigilante*, por la administración férrea y de alcance capilar de una evaluación que inspecciona y castiga en un sistema educativo que funciona como un mercado desde la década de los ochenta.

23 La “recuperación de la rectoría de la educación” se procuró mediante la definición de los requisitos para el ingreso, promoción y permanencia de los profesores, en lo que muchos entendieron como una reforma laboral o antisindical (restar agencia a actores tradicionalmente organizados) más que educativa. Así quedó establecido en la LGSD.

24 Como los siete temas y retos del Programa Especial de Ciencia, Tecnología e Innovación, al que los aspirantes a las Cátedras CONACyT deben referirse para ser contratados. El

bernamentales (gestión indirecta) la responsabilidad de administrar programas y recursos destinados a resolverlos (SEP, CONACyT, SNI). La amplia capacidad de agencia permite a los individuos interactuar con las entidades evaluadoras de manera más abierta, específicamente al incidir en los medios que utilizan para satisfacer los requerimientos estipulados por dichas entidades a cambio de obtener (o no perder) recursos económicos y simbólicos (incentivos y desincentivos). Al mismo tiempo, existen numerosos incentivos estructurales (“codazos suaves”) para que los individuos elijan las actividades académicas y las modalidades de investigación más redituables, y eviten aquéllas que no lo son (la publicación en una revista indizada por sobre la docencia; la hiperespecialización por sobre los enfoques multidisciplinarios). La gestión de tipo autorregulada establece pautas de comportamiento transversales a las comunidades académicas del país (PROMEP), lo mismo que en contextos institucionales específicos (PRIDE-UNAM). Esta lógica opera con los mismos principios en aquellas políticas encaminadas a gobernar indirectamente el comportamiento de las instituciones, sobre todo a través de incentivos económicos (PIFI, PNPC, COPAES). Finalmente, un rasgo fundamental de esta forma de gestión es que los actores educativos *aspiran* a encaminar sus proyectos académicos o científicos a través de las trayectorias establecidas por este tipo de políticas.

- d) *Lógica de gestión autónoma*. En este caso, el grado de coerción es prácticamente inexistente, al tiempo que los individuos cuentan con una elevada capacidad de agencia. El tipo de enfoque es procedimental, pero, dada su naturaleza, el carácter mediato de las modalidades de intervención, y de los instrumentos de política, hace muy borrosa la frontera entre las entidades evaluadoras que los utilizan; al tiempo que su intención, más que resolver problemas, es crear las condiciones para alcanzar ciertos objetivos (conductuales, políticos, económicos, sociales, etcétera). En el caso del gobierno, su presencia puede interpretarse con un no-hacer (o *laissez faire*), mientras que en el caso de las empresas privadas por lo general trae consigo un cálculo de tipo económico. Los organismos internacionales, por su parte, procesan grandes cantidades de información sobre los sistemas educativos para luego crear horizontes de sentido que inspiren las políticas de los gobiernos nacionales. Las modalidades de intervención por excelencia son la nodalidad y el “codazo suave”. Al igual que en la gestión de tipo autorregulada, los actores pueden sentirse atraídos a participar en este tipo de evaluaciones (ResearchGate; Academia.edu; MisProfesores.com; rankings QS y Times Higher Education), pero solamente como un medio para incrementar su visibilidad y prestigio. El indicador más importante del nivel de agencia característico de la gestión autónoma es el hecho de que son

PECiTI consta de siete ejes: ambiente, conocimiento del universo, desarrollo sustentable, desarrollo tecnológico, energía, salud y sociedad.

los actores los que aplican un cálculo de tipo procedimental ante las posibilidades de incrementar sus posibilidades de obtener más prestigio y recursos al participar (o no) en este tipo de evaluaciones. Por su parte, las entidades aspiran a que las consecuencias de sus mediciones les permitan llegar a ocupar un lugar entre las entidades que operan bajo una lógica de gestión autorregulada.

*Entre operadores y maximizadores:
¿hacia un sistema diferenciado de evaluación?*

Tal y como hemos señalado, el ecosistema de las evaluaciones es un espacio común al que se ha llegado de manera más o menos accidental, y siguiendo trayectorias —en principio— independientes. Sin embargo, el paso del tiempo y la influencia de las políticas orientadas a reforzar y ampliar los alcances de esta práctica han hecho que confluyan en aspectos esenciales. En primer lugar, independientemente de su objeto y alcance, la evaluación se rige por los principios de que: a) “todo es medible”; b) “lo que no se puede medir, no se puede mejorar”; y c) “las evaluaciones más adecuadas son cuantitativas y estandarizadas”. En segundo lugar, para que una evaluación sea efectiva debe tener *consecuencias* distributivas en términos económicos o simbólicos, a pesar de que su implementación se justifique con argumentos que no son ni económicos ni discriminatorios. En tercer lugar, la evaluación opera sobre las arquitecturas de la elección de los individuos y determinan sus grados de libertad con relación al tipo de decisiones y comportamientos posibles en sus respectivos contextos (instituciones, aulas, oficinas y laboratorios, entre otros). Finalmente, las políticas de evaluación están cargadas de sentidos acerca de *cómo* debería ser la realidad sobre la que operan, por lo que, independientemente del grado de coerción que ejerzan, su implementación implica la expropiación de la capacidad de los individuos de relacionarse directamente con el hecho educativo.

Las lógicas de gestión analizadas en este apartado reflejan el tipo de gobernanza diferenciada del sistema educativo mexicano, que se caracteriza por la separación de las políticas destinadas a la educación básica (modalidad *jerárquica*) de aquellas dirigidas a la administración del nivel superior (gobernanza de *control a la distancia*).²⁵ De acuerdo con la tipología que hemos desarrollado, las modalidades de *planificación y centralización* —que son las más coercitivas—, corresponden a la forma en que se administran las políticas de evaluación y aseguramiento de la calidad en el nivel básico. Por su parte, tanto la educación superior como las políticas que ponderan la producción científica y académica se rigen de acuerdo con los criterios de las lógicas de *autorregulación y gestión autónoma*, cuyo grado de coerción es escaso o nulo.

Ahora bien, si tomamos en cuenta la evolución de las políticas educativas en nuestro país, es interesante notar que, aunque las IES han sido objeto de diferentes tipos de evaluaciones desde principios de los años ochenta, los

²⁵ Para saber más sobre las modalidades de gobernanza sistémica en la educación, sugerimos revisar el artículo de Capano (2011). Para el caso latinoamericano, ver González-Ledesma (2014).

hacedores de políticas han demostrado un mayor interés en crear las condiciones para el funcionamiento de un sistema de evaluación en el nivel básico. Esto se explica por, al menos, tres razones: 1) el consenso surgido a principios de los noventa entre los decisores políticos, grupos de académicos y organismos internacionales de que los esfuerzos y recursos institucionales debían concentrarse en el nivel básico; 2) la implementación de políticas de autofinanciamiento en las IES públicas y el impresionante crecimiento de la oferta privada de educación superior hasta el año 2000; y (en la actualidad) 3) la necesidad política de neutralizar el poder de las organizaciones magisteriales (el SNTE y la Coordinadora Nacional de Trabajadores de la Educación, CNTE).²⁶

Las diferencias en la administración de las políticas de evaluación entre el nivel básico y el superior podrían considerarse hasta cierto punto “naturales” desde el punto de vista de los procesos formativos involucrados en cada caso; sin embargo, al observar el impacto de las lógicas de gestión entre los actores de ambos niveles, es claro que la diferencia no radica tanto en cuestiones de índole pedagógica, cuanto organizacional. El Sistema Nacional de Investigadores y la reforma educativa de Enrique Peña Nieto son una excelente prueba de ello.

Hace 34 años se creaba el SNI como un mecanismo de compensación ante el deterioro salarial ocurrido en la década de los ochenta, que buscaba retener al personal de mayor calificación en las universidades públicas (Buendía *et al.*, 2017). A través de este programa, las autoridades educativas no sólo trataron de afrontar una contingencia económica articulando una respuesta de tipo organizacional; también esperaban que esta respuesta tuviera consecuencias positivas en el plano académico. De hecho, apenas superada la contingencia económica el SNI no sólo no desapareció, sino que se convirtió en un programa vertebral de las políticas científicas en nuestro país. En términos organizacionales, y esto es interesante subrayarlo, la lógica de este programa sentó un precedente como mecanismo para regular la conducta de los académicos, históricamente reacios a la autoridad y celosos de su autonomía. El SNI no sólo era (y es) un incentivo económico basado en la producción y el mérito; se trataba, sobre todo, de un indicador del reconocimiento a la trayectoria académica y un símbolo de estatus en el marco de una jerarquía científica que se volvía tangible por primera vez. Es por ello que este instrumento se volvió recurrente en otras clases de estímulos, tanto transversales como de carácter institucional.

Los programas de estímulos constituyen un aliciente que orienta la conducta de los individuos hacia la satisfacción de los requisitos que se

26 En el marco del ambiente reformista de la década de los noventa, se habían generado amplias expectativas de que las instituciones de educación superior iniciarían un proceso de transformación que las acercaría, en términos de financiamiento y organización, a las IES privadas. Paralelamente, el gobierno planificó y comenzó a implementar cada vez más programas de financiamiento condicionado, destinados a representar porcentajes cada vez mayores del presupuesto de las IES. Por motivos cuyo análisis escapa a los objetivos del presente artículo, las expectativas en torno a las tendencias de la educación superior no se cumplieron y, ya para el año 2000, el gobierno implementaba políticas de tipo inercial en cuanto al financiamiento y el crecimiento de la matrícula. De hecho, la participación público-privada en la cobertura total de la educación superior se ha mantenido prácticamente idéntica desde entonces (+/- 70-30 por ciento) (González-Ledesma, 2015).

necesitan cumplir para acceder a ellos, lo que los convierte en verdaderos *maximizadores* de la carrera académica que calculan, crean estrategias y orientan su producción en su beneficio. Sin embargo, esto no quiere decir que la institucionalización y expansión de este tipo de programas viniera acompañada de una recuperación real de los salarios en el ámbito de las lógicas de gestión autorregulada y autónoma; lo que ha sucedido es que, en la mayoría de los casos, el sueldo se ha convertido en el complemento del incentivo. La capacidad de agencia en dichos ámbitos, por lo tanto, no está determinada solamente por la forma en que se distribuyen los (des) incentivos, sino, sobre todo, por las expectativas individuales de obtener las mejores condiciones posibles (económicas y simbólicas), de acuerdo con el perfil profesional con el que se cuenta en cada caso.

A lo largo del sexenio que está por concluir, fuimos testigos de la forma en que se esgrimió la reforma constitucional en materia educativa como una condición esencial para alcanzar metas de índole pedagógica e, incluso, de justicia social. Sin embargo, el énfasis puesto en la Ley General del Servicio Docente y, particularmente, en la evaluación de los profesores, demuestra hasta qué punto el meollo de la cuestión es más bien organizacional, es decir, político. Y es que, parafraseando a Harold Laswell (1936), la reforma no ha sido otra cosa que la disputa por definir *quién* conquista el control de la educación en México, *cuándo* lo logra y, sobre todo, *cómo* lo hace. La consigna de “recuperar la rectoría del Estado en materia educativa” reconocía implícitamente que el control de la educación estaba en *otras manos*: el magisterio disidente, agrupado en la CNTE; pero también —y ésta fue una sorpresa para muchos—, de la otrora aliada política, Elba Esther Gordillo, lideresa del SNTE.

Hacerse del control de la educación significaba, entonces, fragmentar al magisterio como gremio, al sustituir la negociación colectiva por mecanismos individualizados de administración laboral. A tal propósito, la LGSD define el perfil y los requisitos mínimos que deben tener los docentes, al tiempo que implementa el concurso de oposición para el ingreso de nuevos maestros y la evaluación, cada cuatro años, de quienes ya son parte del servicio. Una vez neutralizada la capacidad de interlocución del SNTE y eliminado el obstáculo representado por la CNTE, los profesores se convertirían en *operadores del sistema educativo*. Esto le permitiría al Estado, por fin, controlar lo que pasa al interior de las aulas, en términos de: 1) crear los contenidos educativos, administrar la forma en la que éstos se imparten e integrar curricularmente la educación obligatoria; 2) empoderar la figura del director a través de la autonomía escolar; y 3) crear sistemas de información, tanto para monitorear el cumplimiento de los objetivos de la reforma (a través del INEE), como para facilitar la administración del subsistema, con la creación del Sistema de Información y Gestión Educativa (SIGED).

Tanto los programas característicos de las lógicas de gestión de *autorregulación* y *autónoma*, como las de *planeación* y *centralización*, ilustran la forma en que lo organizacional se presenta como medio para lograr objetivos educativos, tanto en el marco de *reformas de largo aliento* (como en

el caso de la educación superior y las políticas de producción científica), como en aquellas de índole *estructural* (la reforma constitucional en materia educativa). En ambos casos llama la atención cómo, a nivel sistémico, el hecho de anteponer lo organizacional a lo educativo nos coloca de nueva cuenta ante la pregunta que interroga por el mejoramiento efectivo de la educación. Sólo para encontrarnos, de nueva cuenta, con que no sabemos a ciencia cierta si mejoró la calidad y el nivel de la investigación de nuestro país, o si está mejorando la educación básica. Obviamente no estamos dispuestos a conformarnos con respuestas tautológicas, como la expresada por el exsecretario de educación, Otto Granados Roldán, ante la UNESCO, ya que se basan en el fortalecimiento *del medio* (la cuantificación) como prueba del éxito de una política (cuántos investigadores se incorporaron al SNI este año / cuántos profesores aprobaron su examen de evaluación).

CRÍTICA DE LAS CRÍTICAS A LA EVALUACIÓN DE LOS PROFESORES UNIVERSITARIOS

Las evaluaciones de los académicos en IES, como muchos han señalado, muchas veces son opacas o llanamente ficticias, premian el productivismo, conspiran contra la calidad, son reduccionistas y utilizan formas obusas o abiertamente absurdas para “medir” el trabajo académico. ¿Cómo es eso posible? ¿Por qué miles de las personas con más estudios en el país permiten que tal situación se reproduzca en su propio medio profesional? ¿Por qué los métodos de evaluación adquirieron la forma que tienen actualmente? ¿Qué utilidad tienen las evaluaciones (con todo y sus defectos y las molestias que producen entre los evaluados)? ¿Por qué los evaluados participan en ellas?

Pensemos en la última de estas preguntas. Tanto en la literatura especializada como en las opiniones que externan los universitarios, es común oír que las evaluaciones les son impuestas desde afuera y que son forzados a participar en ellas. Al interrogar a los académicos, es poco común encontrar algún rasgo reflexivo donde se vean a sí mismos como copartícipes del sistema que los somete. Es raro encontrar opiniones como ésta de una investigadora de la Universidad Autónoma de Nuevo León (UANL):

El asunto es que hay una aceptación. Finalmente se legitima vía el silencio, vía la aceptación de que quien otorga [los estímulos] es el director, que es discrecional. Entonces todo contribuye a que lo mejor sea “portarse bien”. Para muchísimos maestros el repartir muchos [estímulos] nivel 1 [del tabulador de la UANL] en lo que repercute es en agradecimiento, en decir: “bueno, pues peor es nada; aunque sea que no me lo quiten eso”. Entonces, va uno convalidando las prácticas. Somos absolutamente corresponsables. Tenemos que dejar de sentir que somos las víctimas. Somos completamente responsables de lo que nos hace el CONACyT, de lo que nos hacen los sistemas de estímulos. Somos absolutamente corresponsables. Creo que, si pudiéramos reflexionar un poco más desde ese lado, tal vez podríamos hacer

otras cosas para que esto no se siga convirtiendo en esta situación tan perniciosa en gran parte de nuestras universidades públicas.²⁷

Se trata de una observación aguda que merece ser considerada seriamente, o incluso, ser llevada más lejos. Esta profesora enfatiza la corresponsabilidad de los evaluados por su *aceptación* apática de la situación. Pero hay algo más profundo en juego: la *colaboración activa* de los académicos para garantizar la reproducción del sistema de evaluación y distribución de bienes escasos que impera actualmente.

Que esta colaboración sea activa no significa que sea consciente o entusiasta, ni que los actores estén libres de ser coaccionados por algún tipo de violencia simbólica. Ésta, como se sabe, es un tipo de violencia que se ejerce sobre un agente social con su propia anuencia, y es particularmente efectiva cuando no es vista como violencia (Bourdieu, 1995). Así, podemos decir que programas como el SNI o el PROMEP han tenido aceptación, en parte, gracias a su capacidad para que sus tipificaciones y clasificaciones sean adoptadas por los propios académicos para autodefinirse y presentarse ante sus pares. Es común, por ejemplo, que etiquetas como “SNI I” o “Perfil deseable” aparezcan de manera prominente en la primera página de los *curricula vitae*, en solapas de libros y en páginas personales de Internet. Éste es uno de los tantos modos en que los agentes contribuyen a producir la eficacia que determina aquello que los domina.

Una postura crítica ante la evaluación del trabajo académico en educación superior en México debe incluir un cuestionamiento a las posturas que conciben a los académicos evaluados como actores pasivos, victimizados por una política torpe y cruel. Los académicos son agentes y, como tales, con sus actos perpetúan activamente las condiciones que contribuyen a restringir sus propias actividades. Hay muchas maneras en las que los estudiosos del fenómeno educativo, que critican el sistema de evaluación, les niegan agencia a los académicos evaluados. Una de ellas es el reduccionismo económico, algo que puede verse tanto en la literatura especializada como en textos de coyuntura publicados en los medios de comunicación.

Este reduccionismo puede tomar la forma de sostener que los académicos sólo participan en los programas de estímulos y evaluación por sus intereses monetarios; esto implica reducirlos a meros agentes económicos (e.g., Ibarra y Porter, 2007).²⁸ Otra forma de reduccionismo es el que explica los métodos de evaluación como producto de necesidades burocráticas de control o de nuevas formas gerenciales de administración (e.g., Lomnitz, 2016; Acosta, 2015). Finalmente, hay quienes sostienen que los modelos macroeconómicos son responsables de por qué y cómo se evalúa en todos los niveles del sistema educativo (e.g., Toledo, 2015; Aboites, 2012).

²⁷ Entrevista de investigación, 9 de diciembre de 2016.

²⁸ En palabras de Ibarra y Porter: ahora “se trabaja cada vez más por la obtención de un ingreso creciente, que se sitúa silenciosamente como ‘fin’ en sí mismo y como razón fundamental del quehacer académico” (2007: 35).

Por supuesto que los académicos tienen intereses económicos y que son constantemente empujados por aparatos burocráticos y por fuerzas económicas a las que tienen que adaptarse. Los académicos viven en el mundo. Y en el mundo hay pasiones egoístas, aparatos políticos y estructuras económicas. Todo eso pesa sobre ellos. Pero los agentes no son títeres que sólo actúan cuando alguien jala sus hilos; tampoco son una cera blanda que las estructuras económicas y administrativas moldean a su antojo. Los académicos tienen otros intereses, además de los económicos, que los impulsan a tomar decisiones sobre cómo realizar su trabajo y hacia dónde dirigir sus carreras. Y poseen también la habilidad de resistir, negociar y transformar las estructuras que constituyen su ámbito de acción.

Ni el apetito pecuniario de los profesores, ni el “sadismo” de los burócratas, ni los modelos gerenciales que se han introducido en las IES, ni el credo de los tecnócratas, ni las políticas neoliberales pueden ser tomadas como *causa suficiente* para explicar la forma actual del sistema de evaluación, o por qué los distintos actores involucrados participan en su reproducción. Apelar al neoliberalismo (o a cualquier otra fórmula reduccionista) como explicación omnímoda es una de las maneras más comunes de despojar a los actores de su agencia, y se ha convertido en una respuesta cómoda que propicia más la abulia intelectual que una postura legítimamente crítica. Hipótesis alternas o complementarias deben ser exploradas antes de conformarnos con esa visión que presenta a una política económica o administrativa de gran escala como una suerte de *deus ex machina* que causa todos los males y contesta todas nuestras preguntas. Debemos abrir rutas explicativas de por qué se gestiona y se evalúa la labor académica de la manera en que se hace ahora; esto incluirá explorar las instituciones y convenciones que nacieron endógenamente en el mundo académico, y que apoyaron la creación y mantenimiento de las políticas de evaluación.

De Vries y Álvarez (2015) se han hecho la pertinente pregunta de por qué las políticas de evaluación persisten a pesar de que los resultados que se desea obtener con ellas son insatisfactorios. Esa permanencia no sólo se debe a que las políticas se han institucionalizado y “encerrado” (*locked-in*), como ellos sostienen. Aunque las políticas de evaluación han fracasado en cumplir sus metas manifiestas, persisten, en parte, porque han tenido éxito en solucionar problemas que no necesariamente buscaban resolver en primera instancia, pero que son igualmente cruciales. En particular, han tenido éxito en legitimar la repartición desigual de recursos económicos y simbólicos.

El bizantino sistema de evaluación en las IES existe en parte porque la sombra —y en algunas instituciones la dolorosa presencia cotidiana— de una academia clientelista lleva a los académicos a abrazar el productivismo y la numerología. Si no consideramos la presencia de esa sombra, el sistema de evaluación parecería como algo únicamente impuesto desde afuera de la academia (proveniente del neoliberalismo, de la tecnocracia, la racionalidad gerencial, etcétera). Es crucial entender que los evaluados reconocen en las evaluaciones ciertos valores que ellos aceptan (implícita y explícitamente) como legítimos. Su imagen y posición como académicos se forman, en

parte, a través de los resultados de las evaluaciones mismas. Las evaluaciones crean al académico. Los académicos necesitan tener éxito en los procesos de evaluación (para ser contratados y para avanzar en el escalafón, por ejemplo). Al mismo tiempo, las evaluaciones, para existir, requieren del “aval” de los académicos; necesitan que éstos las reconozcan como legítimas (algo que sucede tácitamente, por ejemplo, al aceptar someterse a una evaluación).

Futuras investigaciones tendrán que poner mayor atención a las fuerzas, intereses y valores internos de los agentes académicos para poder determinar cuándo los cambios llegaron desde fuera del campo, cuándo llegaron desde dentro, y cuándo hubo afinidades y coincidencias que actuaron en conjunto.

HACIA NUEVOS TIPOS DE EVALUACIÓN

El sexenio de Peña Nieto terminó con una elección presidencial donde el partido opositor ganó con amplia ventaja. La promesa de un cambio en la política educativa en México fue uno de los temas recurrentes de la campaña del candidato ganador. Qué tipo de cambio se proponían no fue algo muy claro, más allá de algunos puntos generales, entre ellos, eliminar el elemento punitivo de las evaluaciones a profesores de educación obligatoria. Sin embargo, la idea de que las evaluaciones a los profesores deberían continuar siguió estando presente y fue, de hecho, un precepto que defendieron todos los candidatos presidenciales y sus partidos.

Si las evaluaciones van a continuar, ¿qué tipo de evaluación se debería promover ahora? Hay una serie de requisitos generales que han sido identificados por varios especialistas como esenciales en los procesos de evaluación académica (Buendía *et al.*, 2017):

- Se espera que la evaluación tenga por finalidad promover la mejoría de los evaluados.
- Los evaluadores deben dar recomendaciones sobre cómo los evaluados pueden mejorar su desempeño sustancial y cómo podrían mejorar sus resultados en evaluaciones futuras.
- El fin primordial de la evaluación no debe ser premiar o castigar a los evaluados.
- La evaluación debe ayudar a que se evite un divorcio entre los objetivos de las trayectorias individuales y las metas de las instituciones en las que laboran.
- Los criterios de evaluación deben considerar las particularidades de lo que se evalúa, y tomar en cuenta las múltiples dimensiones de lo evaluado (no debe reducirse a medir un aspecto aislado).
- Las reglas y procedimientos de evaluación deben ser explícitos y claros para todos los participantes en el proceso de evaluación.
- Los resultados y el proceso de evaluación deben ser transparentes.
- Dado que cualquier proceso de evaluación es falible, deben existir recursos de revisión.

Cómo deberían ponerse en práctica estos principios tendrá que variar de acuerdo a cada contexto institucional y a cada nivel educativo.

Estas recomendaciones sobre evaluación no son un secreto que este-mos haciendo público en este artículo. Muchos especialistas que trabajan en las instituciones donde se conciben y ejecutan las políticas de evaluación —incluidos los del INEE— las conocen bien. El *quid* para reformar las evaluaciones no es “crear conciencia” en las personas que diseñan los programas de evaluación sobre la existencia de estos lineamientos; el problema de fondo está en que, pese a que se conocen las fuertes limitaciones que tienen las formas actuales de evaluación, no se intenta —o no se puede— transformarlas.

Pongamos, como ejemplo, el primero de los puntos arriba citados: “se espera que la evaluación tenga por finalidad promover la mejoría de los evaluados”. Esto es algo que las autoridades educativas repitieron hasta la saciedad durante los últimos seis años. “Evaluar para mejorar” fue un eslogan que se repitió en discursos, medios impresos y anuncios de radio y televisión de la SEP y el INEE. No obstante, en la práctica, todo el peso estuvo en “evaluar ...”; pero la parte de “... para mejorar” nunca se dijo cómo se lograría. También se hablaba de “evaluación para la calidad”, pero se evaluó mucho y no hubo mejoras evidentes. Esto fue así, en parte, porque no se usaron adecuadamente los resultados como guías para identificar áreas en las que hay que optimizar los procesos educativos; y no se usaron para mejorar, entre otras cosas porque el tipo de evaluaciones mismas no parecían haber estado diseñadas para ese fin.

En el futuro se deberían evaluar las realidades educativas partiendo del hecho de que se trata de procesos complejos donde operan, simultáneamente, múltiples dimensiones interconectadas. Por eso hay que evitar la cosificación de magnitudes abstractas cuando se habla de calidad. Ésta —lo mismo que otros fenómenos complejos, como la inteligencia, para la que también se han creado mediciones altamente imperfectas (Enzensberger, 2009)— no es un fenómeno compacto o perfectamente delimitado. Los *fenómenos complejos no se deben medir en una escala unidimensional*, como un agregado numérico o un promedio para luego situarlo en una serie o un orden de prelación. Ese tipo de medición no sirve para fenómenos multidimensionales (como la calidad académica de un docente o de una institución de educación superior). Los resultados de exámenes de opción múltiple, como los que se usan para evaluar a los profesores de educación básica, o los *rankings* numéricos con los que se compara y clasifica a las universidades de todo el mundo, no son instrumentos propicios para determinar la “calidad” o “idoneidad” de lo medido. Y, por extensión, tampoco deberían ser utilizados como criterios de selección laboral, en lo que respecta a las personas, o para la repartición de presupuesto o estímulos, en lo que respecta a las instituciones.

Además de pensar en evaluaciones multidimensionales, hay otras áreas que sería recomendable cambiar. En las evaluaciones para profesores de educación obligatoria se debe explorar la idea —sugerida por varios

especialistas— de realizar evaluaciones entre pares, como sucede usualmente en educación superior. Un grupo de profesores del mismo nivel e institución (o región) puede evaluar el trabajo de sus colegas, en vez de tener que responder exámenes estandarizados de opción múltiple diseñados por instituciones centralizadas. Una combinación de ambas modalidades también es posible. Con alguna fórmula de este tipo se podría avanzar en subsanar dos necesidades importantes para la evaluación de docentes y directivos: 1) ajustar las evaluaciones a las condiciones de los evaluados, y no al revés (*i.e.*, no ajustar a los evaluados a las condiciones de las evaluaciones, como sucede ahora); y 2) evitar la tentación de usar una evaluación única y homogénea para analizar condiciones, instituciones y procesos educativos diversos (Rueda, 2011).

Mencionaremos otro par de recomendaciones generales. En primer lugar, en vez de continuar con la tendencia propia de la “obsesión métrica” de introducir más evaluaciones para subsanar las carencias de las evaluaciones presentes, se podría optar por hacer *menos y mejores evaluaciones*. Una evaluación que aquilata algo relevante es más valiosa que tres o cuatro evaluaciones que consideran cualidades periféricas. Menos evaluaciones, además, reducirían el elevado costo de la burocracia evaluadora, además de que permitirán a los evaluados concentrarse más en su trabajo y perder menos tiempo en afrontar la interminable marejada de evaluaciones que no para de crecer. Junto a esto, es importante subrayar que las evaluaciones se deben utilizar para lo que fueron diseñadas, y no para fines ulteriores que sacan de contexto sus resultados; en el caso de la Evaluación Nacional de Logro Académico en Centros Escolares (prueba ENLACE), por ejemplo, cuyo propósito manifiesto es tener información “de los conocimientos y habilidades que tienen los estudiantes en los temas evaluados”, con el fin de “proporcionar elementos para facilitar la planeación de la enseñanza en el aula”,²⁹ sus resultados se han utilizado para comparar escuelas, de manera que se ha creado un nuevo mercado de prestigio de las instituciones de educación básica.

En lo que respecta a las evaluaciones en educación superior, hay que considerar otras dimensiones. Una investigación reciente de Vasen (2018) devela un asunto importante de las evaluaciones y de la actividad científica en las IES muy conocido por los profesores-investigadores, pero del que se habla poco abiertamente: la notoria discrepancia entre lo que se enuncia oficialmente como el objetivo de las investigaciones realizadas con algún tipo de apoyo o financiamiento público, y las prioridades que se toman en cuenta a la hora de evaluar a los investigadores. Al contrastar el discurso presente en documentos oficiales de las instituciones públicas con los criterios de evaluación y las prácticas de los propios evaluadores, Vasen encontró una desarticulación entre los objetivos manifiestos de las políticas científicas y las conductas efectivamente premiadas en las evaluaciones. Así, “mientras a nivel político se impulsa un modelo de científico comprometido con el medio y las actividades de vinculación y movilización, en la evaluación

29 Puede consultarse en: <https://goo.gl/GBzXN8> (consulta: 17 de octubre de 2018).

académica se incentiva un perfil clásico” (2018: 1); es decir, se premia, principalmente, que un investigador tenga publicaciones internacionales.

Tal estado de las cosas manifiesta una tensión sobre la que sería útil discutir más en la comunidad académica. Por un lado, las instituciones públicas que otorgan financiamiento a las investigaciones (o que otorgan sobresueldos a los investigadores, como el SNI), tienen derecho a pedir que sus inversiones pecuniarias estén en sintonía con los objetivos generales del gobierno. Eso es lo que ha pasado recientemente y lo que, previsiblemente, seguirá sucediendo con el próximo gobierno federal. Por ejemplo, Elena Álvarez-Buylla Roces, quien fungirá como la próxima directora del CONACyT, ha dicho que en el próximo sexenio se “priorizará la ciencia orientada a la atención de las necesidades de los más pobres” y que se buscará una ciencia orientada “a la comprensión profunda, prevención, previsión y solución de problemáticas de salud, alimentación, ambiente, inequidad, exclusión y violencia” (Enciso, 2018: s/p). Dado que la principal institución de financiamiento para la ciencia en México —el CONACyT— se sostiene con fondos públicos, y es una dependencia pública cuyos dirigentes y objetivos son nombrados y delineados desde el ejecutivo federal, resulta comprensible que dé prioridad a los intereses del gobierno en turno.

Por otra parte, el campo académico y científico es un microcosmos que valora y defiende la independencia de sus fines y su capacidad para asentar, tan autónomamente como sea posible, los principios con los cuales se deben evaluar sus productos. La preciada independencia y autonomía académico-científica va de la mano con el rechazo a criterios heterónomos de evaluación, es decir, ajenos a su propia naturaleza. Muchos científicos se preocupan por resolver problemas científicos que pueden, o no, coincidir con los que se consideran “problemas sociales”. Las investigaciones de un especialista en geofísica podrían servir mucho a la sociedad, al ayudar a prevenir desastres causados por terremotos; si bien los trabajos de un especialista en lógica matemática no necesariamente encontrarán una traducción igual de directa para resolver problemas “reales”, la ciencia, en general, necesita de fases altamente especulativas que no tienen ninguna conexión directa con la vida práctica pero que, a largo plazo, pueden ser la base de adelantos significativos. Cualquier sistema de evaluación que menosprecie cualquiera de estas dos caras de la investigación tendrá problemas para ser efectiva.

La tensión entre académicos que defienden los atributos de aquello que consideran importante (el reconocimiento de sus pares, el prestigio de las publicaciones internacionales, la libertad de cátedra, el valor de la “ciencia por la ciencia”, etc.), y el interés del gobierno en que su inversión en ciencia y tecnología se traduzca en frutos palpables y socialmente presumibles, no va a desaparecer pronto ni es fácil de resolver. Muchas IES públicas tienen que lidiar, pues, con la discrepancia entre: 1) la expectativa de que la educación superior y la investigación pública son bienes públicos que deben servir como instrumentos para el desarrollo y el bienestar colectivo; y 2) los criterios internos del campo académico que valoran la “ciencia por la ciencia”, la “libertad de cátedra e investigación” y los sistemas de evaluación

que premian, sobre todo, los productos de investigación estandarizados y descontextualizados (principalmente los artículos de investigación).

Se trata de un caso típico de “ambivalencia sociológica” (Merton, 1976): expectativas normativas incompatibles sobre actitudes, creencias y comportamientos de ciertas personas o instituciones. Los investigadores tienen que lidiar con la incompatibilidad que producen expectativas encontradas de proveer bienes intelectuales que produzcan “impacto social” y ser autores de conocimientos que se rijan por las exigencias internas y esotéricas del campo académico. Futuras evaluaciones del trabajo académico que deseen distender esta ambivalencia tendrían que aceptar que la investigación se rige por dinámicas propias y que puede ser artificial e inconveniente imponerle criterios externos (por ejemplo, obligarla a ser obvia e inmediatamente útil para combatir la pobreza, la exclusión, etcétera). No reconocerlo abre la puerta a numerosas simulaciones y engaños.

Éstos son, claro está, temas complejos donde chocan los intereses y concepciones del mundo de muchos actores. Reconocer esta complejidad debería sensibilizarnos ante los severos límites que tienen las visiones voluntaristas: aquéllas que fundan sus previsiones en el deseo de que se cumplan, más que en las posibilidades reales de concretarlas. El cambio de las instituciones —las educativas incluidas— no se limita al simple deseo de que las cosas cambien. Constantemente se dice que los cambios requieren que haya “voluntad política”; pero esa opinión refleja un diagnóstico chato y limitado del problema. Hay muchos problemas que no se pueden resolver, aunque se desee resolverlos, o haya voluntad para solucionarlos.

Se debe comprender que la inercia histórica de las evaluaciones —y el complicado entramado de instituciones que componen el ecosistema de la evaluación— hace que las soluciones voluntaristas resulten poco factibles. Las evaluaciones siguen una trayectoria dependiente (*path dependency*), como han mostrado De Vries y Álvarez (2015). Ese tipo de trayectoria hace que la resistencia al cambio sea pertinaz y provenga de muchos frentes. Es difícil desgastar la herencia del pasado una vez que ésta se ha establecido. Y mientras más se prolonga la existencia de un programa de evaluación unido a un sistema de recompensas y castigos, más complicado es eliminarlo. En parte por eso es más probable que veamos en el futuro próximo un cambio sustancial en las evaluaciones que nacieron con la reforma educativa de 2013; pero que programas como el SNI permanezcan sin transformaciones profundas.

Igualmente, debemos tener en cuenta que la gran variedad de mediciones y evaluaciones cuantitativas pueden estar bien o mal hechas, y pueden ser empleadas para numerosos fines. Hay que recordar que no hay nada en ellas que las haga intrínsecamente despreciables y maliciosas. Este tipo de evaluaciones puede ser útil para asir fenómenos masivos que de otro modo sólo podrían conocerse precaria y anecdóticamente; pero tampoco son infalibles, y tienen limitaciones que no se pueden ignorar. Tienden, como toda medición, a ser reduccionistas; la cuantificación, para ser efectiva, requiere concentrarse en unas pocas de las múltiples cualidades del objeto medido. También hay

que evadir la idea de que lo que estamos viviendo es simplemente producto de una arbitrariedad burocrática o el capricho de algún funcionario. Nos encontramos frente un fenómeno de época, una corriente internacional que, si bien no es inevitable, tampoco va a desvanecerse fácilmente; tenemos que entenderla y estudiarla para usarla a nuestro favor (Vera, 2016).

Quisiéramos enfatizar una última idea. Como ya dijimos, pese a las demandas para echar atrás la reforma educativa, esencialmente todos los actores involucrados (incluidos los maestros) insistieron en que las evaluaciones deben permanecer. Quienes se opusieron a la reforma declararon en muchas ocasiones que no objetaban las evaluaciones por sí mismas, sino sólo su carácter punitivo. Si esto se cumpliera, se tendrían mecanismos de evaluación de los profesores sin que los resultados pongan en peligro su trabajo. Tal escenario significaría una clara victoria a corto plazo para el magisterio. Sin embargo, si el instrumento evaluador permanece, probablemente se naturalizará y se convertirá en un elemento no cuestionado del paisaje institucional educativo. Eso acarrearía sus propios peligros, como reactivar, años más adelante, la parte punitiva en un instrumento de evaluación que estará más enraizado de lo que está ahora.

REFERENCIAS

- ABOITES, Hugo (2012), *La medida de una nación. Los primeros años de la evaluación en México. Historia de poder y resistencia (1982-2010)*, Buenos Aires, Itaca-UAM-CLACSO.
- ACOSTA Silva, Adrián (2015, 26 de noviembre), “La evaluación como ingeniería”, *Campus Milenio*, en: <https://bit.ly/2Da0jEx> (consulta: 12 de septiembre de 2018).
- ÁLVAREZ Mendiola, Germán (2011), “El fin de la bonanza. La educación privada en México en la primera década del siglo XXI”, *Reencuentro*, núm. 60, pp. 10-29.
- ARNAUT Salgado, Alberto (2017, 21 de julio), “¿Tantos millones de hombres y mujeres hablaremos inglés?”, *Profelandia*, en: <https://goo.gl/PMfCRi> (consulta: 17 de octubre de 2018).
- ARNAUT Salgado, Alberto (2018), “Leyes, gobernabilidad y reforma educativa”, ponencia presentada en el encuentro “Balance del sexenio en materia educativa”, México, UNAM-IISUE, 16 de mayo, <https://goo.gl/vZZ5rB> (consulta: 17 de octubre de 2018).
- BACKHOFF, Eduardo, Norma Larrazolo y Martín Rosas (2000), “Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)”, *Revista Electrónica de Investigación Educativa*, vol. 2, núm. 1, pp. 11-28.
- BOBBIO, Luigi, Gianfranco Pomatto y Stefania Ravazzi (2017), *Le Politiche Pubbliche. Problemi, soluzioni, incertezze, conflitti*, Milano, Mondadori (edición Kindle).
- BOURDIEU, Pierre (1995), *Respuestas. Por una antropología reflexiva*, México, Grijalbo.
- BRACHO González, Teresa (2018), “Hacia un concepto de calidad”, *Gaceta de la Política Nacional de Evaluación Educativa en México*, vol. 4, núm. 10, pp. 23-27.
- BUENDÍA, Angélica (2013), “Genealogía de la evaluación y acreditación de instituciones en México”, *Perfiles Educativos*, vol. 35, número especial, pp. 17-32.
- BUENDÍA, Angélica, Susana García Salord, Rocío Grediaga, Monique Landesmann, Roberto Rodríguez-Gómez, Norma Rondero, Mario Rueda y Héctor Vera (2017), “Queríamos evaluar y terminamos contando. Alternativas para la evaluación del trabajo académico”, *Perfiles Educativos*, vol. 39, núm. 157, pp. 200-219.
- CAMPBELL, Donald T. (1976), “Assessing the Impact of Planned Social Change”, *Occasional Paper Series*, núm. 8, en: <https://goo.gl/tircBx> (consulta: 17 de octubre de 2018).

- CAPANO, Giliberto (2011), "Government Continues to Do its Job. A comparative study of governance shifts in the higher education sector", *Public Administration*, vol. 89, núm. 4, pp. 1622-1642.
- CHUAYFFET Chemor, Emilio (2015, 24 de junio), "Evaluación y reforma educativa", *El Universal*, en: <https://goo.gl/FoMC5A> (consulta: 17 de octubre de 2018).
- Comité Invisible (2017), *Ahora*, Logroño, Ed. Pepitas de Calabaza.
- DAHL, Robert y Charles Linblom (1954), "Politics, Economics and Welfare", *Journal of Politics*, vol. 16, núm. 2, pp. 386-389.
- DE VRIES, Wietse y Germán Álvarez (2015), "Can Reform Policies be Reformed? An Analysis of the Evaluation of Academics in Mexico", en Pavel Zgaga, Ulrich Teichler, Hans G. Schuetze y Andrä Wolter (coords.), *Higher Education Reform: Looking back-looking forward*, Fráncfort, Peter Lang Verlag, pp. 257-272.
- Decreto por el que se reforman los artículos 3o. en sus fracciones III, VII y VIII; y 73, fracción XXV (2013b, 26 de febrero), *Diario Oficial de la Federación*, México.
- ENCISO L., Angélica (2018, 10 de julio), "El CONACyT dará prioridad a la ciencia orientada a combatir la pobreza", *La Jornada*, en: <https://goo.gl/aBW7q7> (consulta: 17 de octubre de 2018).
- ENDERS, Jürgen (2015), "Una 'carrera armamentista' en la academia: los rankings internacionales y la competencia global para crear universidades de clase mundial", *Revista de la Educación Superior*, vol. 44, núm. 176, pp. 83-109.
- ENZENSBERGER, Hans Magnus (2009), *En el laberinto de la inteligencia*, Barcelona, Anagrama.
- ESPELAND, Wendy Nelson y Michael Sauder (2007), "Rankings and Reactivity: How public measures recreate social worlds", *American Journal of Sociology*, vol. 113, núm. 1, pp. 1-40.
- ESPELAND, Wendy Nelson y Michael Sauder (2016), *Engines of Anxiety: Academic rankings, reputation, and accountability*, Nueva York, Russell Sage Foundation.
- ESPELAND, Wendy Nelson y Mitchell L. Stevens (1998), "Commensuration as a Social Process", *Annual Review of Sociology*, vol. 24, pp. 313-343.
- FALABELLA, Alejandra (en prensa), "La seducción por la hipervigilancia: el caso de la educación escolar chilena (1973 a 2011)", en Carlos Ruiz, Francisco Herrera y Leonora Reyes (coords.), *La privatización de lo público en la educación escolar. El caso de Chile en perspectiva comparada*, Santiago, Editorial Universitaria.
- GRANADOS Roldán, Otto (2018), "Palabras de Otto Granados Roldán, secretario de Educación Pública de México, en la 204a sesión del consejo ejecutivo de la UNESCO", 9 de abril, en: <https://goo.gl/f5TBK4> (consulta: 18 de octubre de 2018).
- Gobierno de México (1996, 19 de febrero), "Programa de Desarrollo Educativo 1995-2000", *Diario Oficial de la Federación*, México.
- Gobierno de México (2013), *Plan Nacional de Desarrollo*, México, en <https://goo.gl/JxwNpv> (consulta: 18 de octubre de 2018).
- GONZÁLEZ-Ledesma, Miguel A. (2014), "New Modes of Governance of Latin American Higher Education. Chile, Argentina, Mexico", *Bordón. Revista de Pedagogía*, vol. 66, núm. 1, pp. 137-150.
- GONZÁLEZ-Ledesma, Miguel A. (2015), *The Changing Role of Governance of Latin American Higher Education*, Tesis de Doctorado, Florencia, Scuola Normale Superiore di Pisa.
- HOOD, Christopher C. (1983), *The Tools of the Government*, Londres, MacMillan.
- HOOD, Christopher C. y Helen Z. Margetts (2007), *The Tools of Government in the Digital Age*, Houndmills, Palgrave Macmillan.
- IBARRA, Eduardo y Luis Porter (2007), "El debate sobre la evaluación: del homo academicus al homo economicus", *Reencuentro*, núm. 48, pp. 34-39.
- LASWELL, Harold, D. (1936), *Politics: Who Gets What, When, How*, Nueva York, Whittlesley House.
- LOMNITZ, Claudio (2016, 12 de mayo), "Curriculismo mágico", *La Jornada*, p. 15.
- LOWI, Theodore (1964), "American Business, Public Policy, Case-Studies, and Political Theory", *World Politics*, vol. 16, núm. 4, pp. 677-715.
- MARTÍNEZ Rizo, Felipe (1992), "La desigualdad educativa en México", *Revista Latinoamericana de Estudios Educativos*, vol. 22, núm. 2, pp. 59-120.
- MERTON, Robert K. (1976), *Sociological Ambivalence and Other Essays*, Nueva York, The Free Press.

- MULLER, Jerry Z. (2018), *The Tyranny of Metrics*, Princeton, Princeton University Press.
- NEAVE, Guy (2012), *The Evaluative State, Institutional Autonomy and Reengineering Higher Education in Western Europe. The prince and his pleasure*, Londres, Palgrave Macmillan.
- NUÑO Mayer, Aurelio (2016), "Mensaje del secretario de Educación Pública, Aurelio Nuño Mayer, durante la presentación de resultados de la evaluación de desempeño, ciclo 2015-2016", 29 de febrero, en: <https://goo.gl/3S1Bah> (consulta: 18 de octubre de 2018).
- O'NEIL, Cathy (2016), *Weapons of Math Destruction: How big data increases inequality and threatens democracy*, Nueva York, Crown.
- OCDE (2014), *Education Policy Outlook. Spain*, OCDE, en: <https://goo.gl/Xidu5r> (consulta: 18 de octubre de 2018).
- PARADEISE, Catherine y Jean-Claude Thoenig (2017), *En busca de la calidad académica*, México, Fondo de Cultura Económica.
- PLÁ, Sebastián (2018), *Calidad educativa. Historia de una política para la desigualdad*, México, UNAM.
- PORTER, Theodore (1995), *Trust in Numbers: The pursuit of objectivity in science and public life*, Princeton, Princeton University Press.
- RODRÍGUEZ, Roberto e Imanol Ordorika (2011), "The Chameleon's Agenda: Entrepreneurial adaptation of private higher education in Mexico", en Brian Pusser, Ken Kempner, Simon Marginson e Imanol Ordorika (eds.), *Universities and the Public Sphere: Knowledge creation and state building in the era of globalization*, Nueva York, Routledge, pp. 219-241.
- ROLDÁN, Nayeli (2018, 13 de mayo), "SEP redujo recursos para capacitar docentes al mismo tiempo que multiplicó su gasto en comunicación social", *Animal Político*, en: <https://www.animalpolitico.com/2018/05/sep-gasto-reforma-educativa-comunicacion/> (consulta: 30 de mayo de 2018).
- RUEDA, Mario (coord.) (2011), *¿Evaluar para controlar o para mejorar? Valoración del desempeño docente en las universidades*, México, UNAM.
- SALAZAR, Claudia (2018, 12 de mayo), "Excede SEP 2,700% el gasto en imagen", *Reforma*, en: <https://goo.gl/ogcxbK> (consulta: 17 de octubre de 2018).
- SCHMELKES, Sylvia (2018), "Definiciones de calidad en la educación en el Instituto Nacional para la Evaluación de la Educación", *Gaceta de la Política Nacional de Evaluación Educativa en México*, vol. 4, núm. 10, pp. 18-22.
- SCHNEIDER, Anne y Helen Ingram (1990), "Behavioral Assumptions of Policy Tools", *The Journal of Politics*, vol. 52, núm. 2, pp. 510-529.
- SUNSTEIN, Cass R. (2014), *Why Nudge? The politics of Libertarian Paternalism*, New Haven, Yale University Press.
- THALER, Richard H. y Cass R. Sunstein (2008), *Nudge: Improving decisions about health, wealth and happiness*, New Haven, Yale University Press.
- THOMPSON, William (1889), *Popular Lectures and Addresses*, Londres, Macmillan & Co.
- TOLEDO, Víctor M. (2015, 7 de julio), "Mitos neoliberales: la evaluación educativa", *La Jornada*, sección Opinión, en: <https://bit.ly/2ERF1Nu> (consulta: 12 de septiembre de 2018).
- VAN VUGT, Frans, A. y Harry de Boers (2015), "Governance Models and Policy Instruments", en Jeroen Huisman, Harry de Boer, David D. Dill y Manuel Souto (eds.), *The Palgrave International Handbook of Higher Education: Policy and Governance*, Londres, Palgrave-Macmillan, pp. 38-56.
- VASEN, Federico (2018), "La 'torre de marfil' como apuesta segura: políticas científicas y evaluación académica en México", *Archivos Analíticos de Políticas Educativas*, vol. 26, núm. 96, pp. 1-23.
- VERA, Héctor (2016, 20 de enero), "El sistema educativo y la ubicuidad de la evaluación", *Nexos*, blog de educación Distancia por Tiempos, en: <https://goo.gl/uo-GwNd> (consulta: 17 de octubre de 2018).
- VERA, Héctor (2017a), "El *homo academicus* y la máquina de sumar: profesores universitarios y la evaluación cuantitativa del mérito académico", *Perfiles Educativos*, vol. 39, núm. 155, pp. 87-106.
- VERA, Héctor (2017b), "La evaluación cuantitativa del trabajo académico: tres analogías", *Sociológica*, vol. 32, núm. 90, pp. 277-301.