

El desarrollo de centros de evaluación y la formación de evaluadores¹

Robert E. Stake

Resumen: *La práctica de la evaluación, en tanto que actividad formal, se ha caracterizado y se caracteriza aún por tres elementos clave: (i) el predominio del razonamiento basado en criterios y estándares, (ii) la apuesta por la utilización de tests y otras pruebas cuantitativas y (iii) la consideración de que la calidad es un atributo intrínseco de los programas. En la misma línea, la formación que se imparte desde los centros de evaluación se ha centrado principalmente, en la enseñanza de modelos y enfoques metodológicos. Esta perspectiva presenta importantes limitaciones a la hora de determinar la eficacia de un sistema educativo. Se hace, pues, necesario recuperar otra clase de razonamiento que, con frecuencia ha sido denostado: el conocimiento episódico, que insiste en que la calidad de un programa está íntimamente vinculada a las percepciones y a las experiencias del evaluador acerca del mismo y que sitúa al programa en su contexto. Ambos tipos de razonamiento son necesarios; su combinación e integración como si de las dos caras de una moneda se tratase, la “visión binocular”, vendría a convertirse en la mejor apuesta para el evaluador y para la evaluación.*

Palabras clave: *Evaluación basada en estándares, evaluación comprensiva, centros de evaluación, formación en evaluación, visión binocular.*

1. LOS CENTROS DE EVALUACIÓN

El primer tema que abordaré es el desarrollo de los centros de evaluación. A continuación, la formación de los evaluadores y, finalmente, hablaré sobre el concepto de calidad en la escuela. Al principio, creí que se trataba de un tema meramente técnico pero, tras reflexionar, me di cuenta de que pensar en el desarrollo de los

¹ Texto de la conferencia pronunciada por el autor en el *Simposio Internacional de Evaluación* que tuvo lugar en la Universidad Nacional Autónoma de México (México D.F.) el 25 de octubre de 2007. El texto ha sido revisado por evaluadores anónimos a los que se informó de que se trataba de una conferencia.

Revisión científica de la traducción a cargo de Ester García Sánchez.

centros nos hace pensar en todo el ámbito de la evaluación educativa. Dicho ámbito incluye:

- La búsqueda de la calidad y las diferencias en el modo de entenderla.
- La valoración de resultados, en especial los resultados del rendimiento de los alumnos.
- La naturaleza política de la evaluación.
- El clima de competitividad de las instituciones, que se esfuerzan por desarrollar una imagen de calidad.
- El carácter global de la educación, incluido el trabajo a través de internet.

Antes de adentrarme en ese vasto territorio, debo añadir unas palabras sobre mi persona. Procedo del sector cualitativo de la evaluación. Fui director adjunto del *Illinois State Testing Program* pero, decepcionado con la evaluación de programas basada en tests, decidí buscar otras cosas. Llegué a la conclusión de que los evaluadores necesitan prestar atención a las experiencias de alumnos, profesores y de otras personas para comprender la dinámica, la complejidad y la calidad de la educación. Así que cuando enseño y cuando recojo información, doy preferencia a los diálogos, a las situaciones y a los contextos. Esto es un ejemplo de ello:

Madre de gemelos: Ha enseñado a Pepe a usar la cámara, pero no a Pepita.

Profesor: Pero, Sra. Martín, Pepita tenía que reforzar las matemáticas.

Madre: Eso no es justo.

Profesor: Aprecio a los dos, no quiero ser injusto.

Madre: Pepita odia las matemáticas. ¿Es que no puede hacer que a Pepita le guste el cole tanto como a Pepe?

Profesor: Pero son niños muy distintos.

Madre: Su obligación es lograr lo mismo con los dos.

Al escuchar esta breve conversación, nos paramos a pensar tanto en la filosofía educativa como en la práctica docente. ¿Se puede entender mejor la igualdad de oportunidades en educación a través de una serie de conversaciones como ésta? ¿No es obligación de los evaluadores educativos examinar cuestiones tan complejas como la igualdad de oportunidades?

Entiendo la evaluación como la tarea de búsqueda de la calidad (lo que también implica detectar la falta de calidad, claro). Al final de este trabajo hablaré de la cali-

dad en la escuela. Pero resida donde resida, me gustaría examinar algunas de las dimensiones de la calidad, a veces para representarlas recurriendo a mediciones, pero también para hablar de los episodios que revelan calidad. Dimensiones y situaciones, cuantitativo y cualitativo, son métodos que se mezclan, sí, pero, en mi opinión, los métodos cualitativos son especialmente valiosos para ayudarnos a entender la calidad educativa. Esta manía mía será una idea recurrente en todo lo que hoy exponga.

1.1. El centro de evaluación típico

Consideremos los centros de evaluación universitarios. Éstas son cinco posibles características:

- El personal del centro contará con uno o dos profesores universitarios y la ayuda de estudiantes de doctorado.
- Es más una tienda de artesanía que un centro de informática.
- El director ha de ser una persona dispuesta a trabajar muchísimo.
- Rara vez la financiación es suficiente como para que la calidad de la investigación esté a la altura de la que se realiza en otros centros de la universidad.
- El centro se ajustará a las peculiaridades de su entorno, de las instalaciones, el mercado, la entidad financiadora y las estructuras educativas de la zona.

Algunos centros no presentarán estas características, por supuesto.

Pensemos en una de las formas habituales de concebir el trabajo de un centro de evaluación. Todos conocemos el principio universitario que considera tres dimensiones en el trabajo de los profesores universitarios: docencia, investigación y gestión. Gestión entendida como colaboración con su propio departamento, con la profesión y con la comunidad. Veamos lo que hace un centro de evaluación:

- Los centros participan en la docencia de los estudiantes de su universidad, tanto en materia de destrezas evaluadoras como sobre la profesión de la evaluación y, a veces, también ofrecen formación al público en general.
- Los centros investigan, en ocasiones a través de contratos, tanto sobre la teoría como sobre la práctica de la evaluación educativa.
- Los centros ofrecen servicios de evaluación, algunas veces de forma gratuita, a través de consultas, asistencia en proyectos, mejora y desarrollo de programas.

En definitiva, los centros tienen más trabajo del que pueden asumir. Cada centro es diferente y elige entre las oportunidades que se le presentan; su tarea es encontrar su señal de identidad.

1.2. Todos los organismos evalúan

Según explicó magistralmente Michael Scriven (1996), la evaluación es un proceso humano fundamental. La mayor parte de la evaluación es informal y utiliza la intuición y el sentido común, pero muchas veces es formal. No existe una línea divisoria clara entre formal e informal. La evaluación formal es reflexiva, se basa más en la recogida de datos, es más disciplinada.

En el entorno universitario y en todos los organismos se dan muchas situaciones en las que hay evaluación. La mayor parte del trabajo evaluativo no se denomina *evaluación*, así que la mayoría de la gente no aprovecha la experiencia de los centros que evalúan. Independientemente de cómo lo denomine, el evaluador/a busca la calidad incorporando, de un modo u otro, definiciones tanto particulares como universales. Consideremos algunos entornos de trabajo:

- En facultades e instituciones u organismos educativos, se evalúa el éxito de alumnos y profesores.
- En las facultades de medicina y en los hospitales, se evalúan la salud y los tratamientos.
- En las facultades de derecho, los alumnos aprenden a evaluar lo que sucede en función de si es legal o ilegal.
- En las escuelas y bases militares, los profesionales de todos los niveles evalúan las estrategias y las tácticas.
- En las agencias de inteligencia, se evalúan las amenazas a la seguridad y los riesgos.
- En las agencias reguladoras, se evalúan los riesgos, las medidas de protección y el cumplimiento de normas y reglamentos.
- En las escuelas de negocios y en las empresas, se evalúan los procesos, la gestión, las ventas y la publicidad.
- En las facultades de periodismo, se enseña periodismo de investigación, una tarea con un componente evaluativo muy elevado (aunque los medios rara vez los contratan).

Hay muchos entornos laborales, todos ellos impregnados de evaluación: la ingeniería, los ministerios, las artes y la música; entornos de todo tipo.

La mayoría da por hecho que el término *evaluación* se refiere a algo técnico o relacionado con la gestión. De algún modo, la mayoría de la gente que trabaja en las universidades ha llegado a la conclusión de que una formación general en evaluación carece de importancia. Sin embargo, es necesario trabajar en todos estos ámbitos, si los que forman los centros de evaluación desean ampliar su campo de actuación.

Muchos ámbitos necesitan de la evaluación. Los centros evaluadores compiten con otras empresas de servicios a la hora de conseguir contratos. En ocasiones, la competencia es feroz. Algunos competidores ofrecen precios muy bajos a cambio de muchas promesas, incluso de la promesa de que las evaluaciones² serán más rigurosas que las que han llevado a cabo anteriormente. A veces, he perdido contratos contra otras empresas que luego han tenido que recurrir a mí para averiguar cómo hacer el trabajo.

Hace cuarenta años, en Estados Unidos la mayoría de los programas educativos que contaban con financiación externa eran evaluados por centros pertenecientes a universidades. El personal de estos centros era gente amable y competente. Sin embargo, al aumentar los presupuestos destinados a evaluación y al cobrar mayor importancia las consecuencias la competencia aumentó y se perdieron las formas.

Veamos quiénes ofrecen, en la actualidad, servicios de evaluación en muchos países. Figuran aquí seis ámbitos en los que los evaluadores hacen evaluación formal:

- Grandes empresas de evaluación tales como American Institutes for Research, Abt Associates o Rand Corporation.
- Departamentos o unidades que llevan a cabo programas de evaluación interna en organismos gubernamentales y grandes empresas, tales como el General Accounting Office o los departamentos de evaluación del Banco Mundial.
- Pequeñas empresas de evaluación, a menudo en manos de alguien que antes trabajaba para una de las grandes.
- Pequeños centros de evaluación. Uno de los pequeños se hizo grande, el de Dan Stufflebeam en la Western Michigan University.
- Muchas personas, con poca o ninguna formación en evaluación, pero que son referentes en su profesión y están dispuestas a ser evaluadores en su campo.

2 En el original *assessments*.

- Y, por último, otro grupo de personas de cierto prestigio y dispuestas a participar en paneles de expertos, grupos de crisis o comités especiales de investigación, gente que no ha oído en su vida el término “objeto de evaluación” y para los que es un honor que se les pida ayuda para estudiar un determinado problema y evaluar la situación.

Así que, resumiendo, todo el mundo evalúa, y muchos a cambio de dinero. Supuestamente, los centros de evaluación deberían estar vinculados con muchos de ellos y ofrecer formación y asesoramiento a muchos otros.

Actualmente, en los campus técnicamente más avanzados de los países ricos existen, a mi modo de ver, posibilidades de establecer centros de evaluación. Sin embargo, no tengo noticia de que haya mucha demanda o de que se advierta su necesidad.

Generalmente, se supone que estos centros universitarios tendrán valores similares a los de las disciplinas académicas, es decir, las ciencias sociales y las físicas; facultades que buscan la verdad. No se espera que sean especialmente comprensivas con la gente corriente que tiene problemas en su vida cotidiana.

La gran mayoría de los responsables de los programas, cuando encargan a un centro la realización de una evaluación, buscan más un aval a lo que están haciendo que comprenderlo. Buscan el reconocimiento más que el conocimiento.

Y la mayor parte de nosotros, como evaluadores, advertimos que necesitan un refrendo y ofrecemos nuestra ayuda, sobre todo si nos pagan por ello. Y hacemos lo posible por encontrar aspectos positivos; pero también queremos señalar las deficiencias o los problemas que hallamos. La cuestión es, ¿tendrá el centro de evaluación el coraje necesario para señalar los problemas? ¿evaluaremos correctamente la falta de calidad? Creo que la gran mayoría de los centros de evaluación de las universidades se sienten menos seguros que los evaluadores independientes y, por ello, son reacios a señalar defectos.

Parte del negocio de los centros educativos está en impartir cursos, formar evaluadores y ayudar a las diferentes disciplinas académicas, a los profesionales y al público en general a comprender que la evaluación formal es una disciplina en sí misma. La evaluación tiene teorías, métodos, prácticas y un código ético. La evaluación informal es una actividad peculiar. En cambio, la evaluación formal es una disciplina.

2. LA FORMACIÓN DE EVALUADORES

Del mismo modo que los centros evaluadores son diferentes y se adaptan a sus circunstancias, la formación de evaluadores se ha amoldado y debe seguir amoldándose a las necesidades de los estudiantes, de los profesores y de las instituciones. Desde mi punto de vista, no sería bueno para nuestra profesión estandarizar más la

práctica de la evaluación. Podemos imaginarnos otro mundo en el que la evaluación sea tan variada, opaca y adaptada a cada contexto que confunda a nuestros clientes y al público; pero no creo que sea el caso en la actualidad. Así que la formación también debe diversificarse.

El *mentoring* y la formación en proyectos deberían servir para diversificar la formación³ de los evaluadores, al igual que la enseñanza en el aula. Se debe dar oportunidad a quien se está formando en evaluación para que adapte los conceptos y la metodología a su ámbito de trabajo.

Muchos cursos de evaluación hacen hincapié en los distintos modelos o enfoques metodológicos. Creo que pensar en estos modelos distrae la atención de quienes se están iniciando en la evaluación, que debiera dirigirse a las grandes cuestiones de la evaluación. Tal y como yo lo entiendo, estas grandes cuestiones son:

- ¿Se hacen siempre en evaluación comparaciones con estándares?
- ¿Debería la evaluación limitarse a las cuestiones planteadas por el cliente?
- ¿Qué es la calidad?
- ¿Se puede vincular estrechamente la evaluación con la mejora de un programa?
- ¿Cuál es la influencia de la política sobre la evaluación?
- ¿Son siempre útiles los estándares profesionales de evaluación?

La formación de evaluadores siempre debería incluir tales cuestiones. Sin embargo, cuando escribí un libro sobre evaluación (*Standards-Based and Responsive Evaluation*) que se ocupaba de estas cuestiones, no tuvo ningún éxito. No entiendo por qué.

En el libro argumentaba también que explicitar los criterios no es un requisito imprescindible de una buena evaluación ya que, a menudo, hemos de dejar que sean las descripciones de los comportamientos y de las actividades, más que las palabras mismas, las que proporcionen las definiciones. En mi opinión, parte del desarrollo de la disciplina consiste en no abusar tanto de las explicaciones y las nomenclaturas.

2.1. La sistematización de la evaluación informal

Una de mis estrategias formativas es recalcar la importancia de la sistematización continua de la evaluación informal más que su sustitución. Si la evaluación infor-

3 *Education* en el original.

mal pudiese hablar, nos diría que es mejor que la evaluación formal porque, gracias a ella, la humanidad ha sobrevivido miles de años. Hasta hace poco, la humanidad sobrevivió sin la evaluación formal. Probablemente intervino la suerte, pero lo que es seguro es que la supervivencia fue posible porque se tomaron decisiones correctas, y acertar en las decisiones es una cuestión evaluativa. Dado que la evaluación formal tal y como la conocemos hoy no existió prácticamente hasta 1950, y que hasta el sistema judicial, los hospitales, las comunidades religiosas y las ciencias tienen unos cuantos siglos de antigüedad, podemos llegar a la conclusión de que ha sido la evaluación informal la que nos ha traído hasta aquí, no la formal.

Para que la evaluación formal sea mejor que la informal, tenemos que estar convencidos de que nos ayuda a reconocer mejor la “bondad” y, desde luego, sin cambiar nuestra idea acerca de qué es lo bueno y sin incurrir en otros efectos colaterales.

Sin embargo, aunque nos convenciéramos de que, habitualmente, sería mejor evaluar nuestros programas de un modo informal, no formal, la gran mayoría de la gente en la actualidad piensa que la evaluación formal es más fiable, que es más probable que sea ésta la que nos proporcione las evidencias en las que basarnos acerca de la calidad y las deficiencias de los programas. Es aconsejable reflexionar sobre dónde es mejor poner el énfasis, si en la evaluación informal o en la formal. Sin embargo, si en un contrato se establece que es necesario evaluar un programa, se está aludiendo a la evaluación formal. Aun así, dentro de la evaluación formal hay cabida para los datos informales, los juicios y las interpretaciones.

Sigo hablando, para mis adentros, de la creación de centros de evaluación y de la formación de evaluadores.

2.2. Funciones y estilos de evaluación

La evaluación pretende obtener conocimiento sobre el valor de una cosa. Buena parte de ese conocimiento proviene de la experiencia personal, a menudo de nuestra propia conciencia. Sé que el Allopurinol es un medicamento bueno para mí porque si no lo tomo se me forman piedras en el riñón. Lo que me dice el médico es importante y puedo leerme la *Enciclopedia de la Salud* pero, para mí, mi propia experiencia es fundamental. No puedo fiarme siempre de la experiencia pero la utilizo constantemente; la de las personas que conozco y la experiencia indirecta, la reflexión y el análisis de personas a las que no conozco. Reúno todas esas experiencias y conocimientos y, combinando intuición y precisión, hago una valoración de conjunto. También me ayuda beber mucha agua...

La evaluación trata siempre de determinar el mérito y las debilidades. A veces la evaluación hace mucho más, pero su función esencial consiste en determinar el mérito. Ése es su objetivo primordial, su definición, su razón de ser.

En la sala de espera escuché a una señora contarle a un conocido que era “evaluadora”:

—¿A qué te dedicas?

—Ayudo a la gente a perder peso.

—¿Por qué te haces llamar evaluadora?

—Porque la gente te presta más atención si eres evaluadora en vez de dietista.

Los dietistas también son evaluadores, pero si esta señora se dedica a ayudar a las personas a adelgazar, está confundiendo a la gente al hacerse pasar por evaluadora. ¿Puede cada uno llamarse lo que quiera? Claro que sí. Pero nos vamos a entender mejor si consideramos “evaluador/a” a quien, dentro de un campo determinado de actividad, averigua e informa sobre los méritos y carencias del objeto de evaluación. El valor que se le asigne sirve para muchos propósitos, tales como la mejora de un proceso, la concesión de un premio, la aplicación de medidas especiales de refuerzo educativo o comprobar el cumplimiento de un contrato. Estos son algunos de los muchos modos en que se emplea la evaluación (Scriven, 1967).

2.3. Razonamiento basado en criterios y razonamiento episódico

El diccionario no lo recoge, pero los evaluadores normalmente entendemos por *criterio* un descriptor del objeto de evaluación, como su eficacia, su duración, su coste o su valor estético. Y por *estándar* entendemos la “cantidad” de ese criterio con la que clasificamos al objeto de evaluación en un determinado nivel de calidad. Más jerga, pero útil.

Hace 48 años yo era estudiante de doctorado en un Departamento de Psicología. Un día, sentado a mi mesa de trabajo, me di cuenta de que existen las Ciencias (Sociales) de la Educación porque los psicólogos y sociólogos de la educación lograron reestructurar los fenómenos educativos en forma de variables. Habían inventado los constructos, los “ladrillos” para el análisis sistemático de la educación, y llamaron a esos constructos *variables*. A veces, llamaban a estos mismos constructos descriptivos *atributos, propiedades, cualidades, características, facetas o dimensiones*. En muchas ocasiones los llamaban *criterios*. Al reducir a variables los fenómenos complejos -que se dan en el aula, en una sala de juntas, en la historia o en la comunidad- se podían manejar las cosas. Traducir los hechos a variables se denomina *razonamiento basado en criterios*.

Una variable es un atributo que varía. Puede hacerlo de diferentes maneras, pero los científicos sociales decidieron poner el énfasis en la variación en la cantidad. Se considera que ésta (la cantidad) varía a lo largo de una escala. Así que una vez

identificado el constructo, es decir, la escala, lo importante era medir la cantidad. Podemos utilizar estas cantidades para describir, distribuir, comparar, incluso para que parezca que averiguamos las causas, e interpretar dichas causas como base para el control, la mejora o la reforma. Hace 48 años, me parecía que había descubierto la pólvora. Con criterios, con el razonamiento basado en criterios, podíamos medir, y con escalas de medida, podíamos mover montañas (Stake, 2004).

Gracias al muestreo y al razonamiento basado en criterios, el estudio de la educación podía ser algo preciso, permitir que se obtuvieran generalizaciones y ser útil. Todas mis dudas se disiparon. Me apunté a la ciencia de los tests. Me dediqué en cuerpo y alma a convertirme en un “hombre que mide”, y aún lo soy. Soy un “hombre que mide”, trabajo en evaluación de programas. Intento medir la cantidad y la calidad de la educación o de la formación, de los servicios sociales, como también los puntos fuertes y los fallos, los criterios vagos y difusos de la docencia y el aprendizaje.

La psicología me dice que la alternativa al razonamiento basado en criterios es el razonamiento episódico. Los fenómenos educativos se conocen a través de episodios, sucesos, prototipos, actividades, acontecimientos. Estos episodios tienen una base temporal y contextual. Los habitan personas con su propia personalidad, su historia, sus aspiraciones y sus debilidades. A veces hablamos acerca de la personalidad y de las debilidades, de los contextos y de los episodios, en términos de variables. Casi cualquier cosa puede convertirse en una variable; sin embargo, el discurso basado en variables a menudo simplifica en exceso. Conseguimos las herramientas pero perdemos las situaciones.

Para planificar la formación en evaluación y para establecer los principios éticos de un centro deberíamos ser conscientes de que ambos tipos de razonamiento, el basado en criterios y el episódico, coexisten tanto en nuestra cultura como en nuestro pensamiento. Con una especie de visión binocular, a veces podemos reducir el *gap* y, sin darnos cuenta, conseguir una imagen a la que no se llega únicamente con uno de los dos criterios: logramos profundizar combinando ambos tipos de análisis. Un ejemplo de esta combinación de análisis se encuentra en la evaluación de la calidad escolar.

Dediquemos unos minutos al concepto de *calidad*. Se utiliza habitualmente en dos sentidos. Puede referirse a las características de algo, como por ejemplo de una obra musical o de la forma de enseñar de una determinada persona. Hablar de cualidades es una cuestión más descriptiva que valorativa. En música, se referiría a variables como el lirismo, la sonoridad, el clasicismo o el casticismo. En educación, recurrimos a términos como creativa, convencional, centrada en el niño, dispersa. Para John Dewey, “el propio interés y la simpatía tienen c(u)alidades contrarias”. Estos usos de la palabra *calidad* se refieren a la naturaleza o a los ingredientes de las cosas, no a si son buenas o malas.

La acepción que interesa al evaluador *sí* tiene que ver con la bondad de las cosas. La calidad de la interpretación de una obra musical es su grado de excelencia. La cali-

dad de un episodio educativo son sus méritos y carencias. Éste es el sentido en el que normalmente utilizamos el término en el trabajo de evaluación. Buscamos la calidad de los programas, en el sentido de su mérito y su valor. Pero eso deja muchas cosas en el tintero. Con frecuencia, es difícil concretar y llegar a un acuerdo sobre qué es la calidad, el mérito o la excelencia. Los estándares que reflejamos en nuestras palabras o en indicadores son a menudo más simples, menos complejos que lo que experimentamos personalmente. Cuando decimos que la calidad de una redacción de un alumno es mediocre, tenemos en cuenta muchos aspectos: su coherencia, el tema, la gramática, la creatividad, la originalidad, el hilo temporal, el uso de las palabras, el cumplimiento de la tarea, incluso algunas características en las que no pensamos de antemano (aunque no siempre son las mismas, ni siempre las valoramos de la misma manera en todas las ocasiones). El razonamiento basado en criterios es importante, pero la interpretación también lo es.

Como evaluador, no encuentro necesario explicitar qué calidad busco. Algunos evaluadores hacen lo posible por ser explícitos. Pero yo estoy cansado de no utilizar más que uno o unos cuantos criterios. Pretendo enseñar a mis alumnos a que se familiaricen a través de la experiencia con una serie de aspectos del programa. Algunos evaluadores prefieren dedicar sus recursos a medir el mejor o los mejores criterios.

Otra cosa que nos diferencia a los evaluadores es nuestra idea respecto a dónde reside la calidad. Cuando piensas en la calidad de un melón francés, puedes pensar en la bondad como una de sus propiedades. O puedes pensar en su calidad, definida a partir de las opiniones de quienes lo comen. ¿La calidad pertenece al melón o a la experiencia de comérselo? Obsérvese que esto último exige prestar una mayor atención a quien se lo está comiendo. Los evaluadores no se ponen de acuerdo en valorar cuánta atención hay que prestar al “destinatario” del objeto evaluado. Cuando se piensa en la calidad en los tests de rendimiento, se puede considerar como una cualidad de quien se somete a la prueba o como el resultado de la interacción entre quien realiza la prueba y los examinadores. Es parte de la formación.

No hay por qué coincidir conmigo, pero me gusta pensar que la calidad tiene su origen en la experiencia humana. El concepto de calidad no sirve para nada si no se tiene en cuenta la experiencia humana como referente clave. La calidad existe porque las personas la experimentan. Y a través de los años, las experiencias positivas que han contado más han sido quizás la comodidad, la felicidad y el bienestar. Y, desde luego, las experiencias negativas de la falta de comodidades, la ira y miedo. Lo que reconocemos enseguida como una enseñanza de calidad o un melón de calidad tiene sus raíces, en mi opinión, en nuestras experiencias previas. A estas alturas, podemos haber desarrollado estándares formales o informales, convenciones y tradiciones, para valorar los melones y la música, pero las raíces de la idea de calidad están en las experiencias que vamos recordando a lo largo del tiempo. Un reloj es de buena calidad en parte porque marca bien las horas y tiene una maquinaria precisa, pero también porque a la gente le parece muy bueno y superior a otros relojes que hayan conocido. Esto hace que los evaluadores, a medida que nos vamos metiendo en el

negocio de evaluar las cosas, nos damos cuenta de que la calidad depende en gran medida de quién la esté experimentando.

2.4. La calidad en la escuela

Los evaluadores educativos no se ponen de acuerdo acerca de lo que es un buen indicador de la calidad en la escuela. Muchos en Estados Unidos se conforman con las puntuaciones obtenidas por los alumnos. Muchos otros prefieren los sistemas de acreditación e inspección educativa. Todos están de acuerdo en afirmar que la escuela es un entorno complejo y que cualquier indicador por sí solo simplificaría en gran medida el concepto de calidad en la escuela. Sin embargo, existe una gran presión a nivel mundial para considerar sólo una de las dimensiones de la calidad.

La medición de la calidad se vuelve extraordinariamente difícil por el hecho de que distintas personas buscan cosas diferentes en las escuelas, incluso algunas contradictorias, como que los estudiantes sean más conformistas y que sean más reivindicativos. Muchas políticas escolares consideran que estas diferencias son errores de comunicación más que mecanismos democráticos. La calidad es muy complicada.

Los profesores saben mucho de calidad en la escuela, normalmente más que los gestores, que los evaluadores externos y que los miembros del gobierno. Pero los juicios del profesorado y los episodios más significativos son infravalorados. Y a menudo a aquéllos se les acusa de servir a sus propios intereses individuales, que a veces sucede. Pero un buen proceso de evaluación recurre al razonamiento basado en criterios y al razonamiento episódico de los profesores.

¡Un momento! Estábamos hablando de formación. La formación de los evaluadores debería incorporar una reflexión sobre la necesidad de conocer la calidad general de la escuela. El pensamiento moderno y la ética de la gestión dicen que ello es necesario sencillamente para poder comparar las escuelas. Sin embargo, como evaluador profesional, he llegado al convencimiento de que no se suelen emplear bien los indicadores de calidad institucional. Evaluamos la calidad escolar por cuestiones de control, competitividad, castigo y orgullo. Rara vez nos cuestionamos su necesidad.

Es importante para todos en la escuela reconocer la calidad, ver los problemas y buscar la mejora, pero la mejora viene del estudio de problemas concretos, no haciendo un seguimiento de los resultados de la escuela a lo largo del tiempo. La mejora se consigue a partir del trabajo individual y la investigación-acción y, a veces, del estudio de casos o de la evaluación de los participantes. Quien se esté formando en evaluación debe aprender que no existe una única visión de la calidad en la escuela. La escuela puede muy bien estar formada por personas que tienen ideas muy distintas sobre la calidad escolar. La calidad en la escuela y los logros de los alumnos se pueden entender mejor a través de prolongadas observaciones de la actividad escolar. Consideremos este episodio.

La clase de Ciencias comenzó con la observación del crecimiento de una semilla. Cada estudiante coge un par de envases de cartón de leche, cada uno los suyos, del poyete de la ventana. Hoy tienen que registrar el crecimiento vertical de uno o más brotes. Habían plantado nomeolvides, cosmos, y algo llamado mezcla de plantas perennes. Cada alumno lo registra en un diario que ya han comenzado con anterioridad. El diario está hecho con folios blancos doblados por la mitad para hacer un librito.

Un alumno pregunta: *¿Las plantas crecen más rápido si les hablas?*

Profesora: *Buena pregunta, esto lo han estudiado los científicos. Quizás han considerado que, para que crezcan, es mejor la música que hablarles (pausa). Kevin, la estás regando demasiado.*

Un alumno: *¿Cómo hacen las plantas?*

Profesora: *Eso es algo de lo que hablaremos hoy. Primero comprobemos su crecimiento. Levantad la mano si tenéis un tallo de al menos cuatro centímetros o más... ¿y quién tiene seis centímetros o más? ¿nueve centímetros? (se van levantando varias manos, al final sólo una).*

—*Sylvia, tus brotes no han crecido tanto.*

—*Sí que han crecido.*

—*Vale, ¿cuántos habéis notado cambios desde el lunes pasado? ¿Qué cambios?*

—*Dos brotes más.*

—*Uno ha crecido, antes era más pequeño.*

—*El mío estaba saliendo torcido y ahora está derecho.*

—*A lo mejor tenía poca agua, si ahora lo estás regando más...*

La clase continúa. La profesora les hace hablar sobre comparar el crecimiento. Hablan de medidas. La profesora hace que piensen en un experimento. Diseña uno. Hablan de comparar medidas y de medias. La profesora decide enseñarles a calcular la media, y lo hace mal. Cuando el supervisor de la profesora advierte el error, se disgusta mucho. Un evaluador formularía esta pregunta: “¿Lo que les ha enseñado con la realización del experimento queda invalidado por haberles enseñado mal cómo calcular la media?”.

2.5. La aptitud académica

La formación en evaluación debe adaptarse a cada individuo. Pero en educación, la mayoría de quienes se están formando ha de saber que evaluar las escuelas a partir

de tests de rendimiento estandarizados carece de validez. La mayoría de las evaluaciones escolares en Estados Unidos emplean la aptitud académica como indicador de la calidad de la escuela. Esto es, la aptitud académica de sus estudiantes.

La aptitud académica puede definirse como la predisposición a aprender en la escuela, es decir, a beneficiarse de la experiencia. La aptitud es, en parte, una función de la inteligencia “natural”, de las enormes capacidades que provienen de nuestro código genético, cuando tiene ocasión de desarrollarse en su plenitud. Hay muchas cosas en la vida que contribuyen al desarrollo de las aptitudes, sobre todo en los primeros años de vida. Fomentar las aptitudes es, en gran medida, una tarea de madres y padres, pero los hermanos, los compañeros, los parientes, luego los profesores, orientadores y tutores, además de los grupos sociales y de los compañeros de trabajo y todos los demás, ayudan también al desarrollo de las aptitudes académicas. Nuestros futuros evaluadores lo saben.

Desde 1900 hemos utilizado los tests de inteligencia, creados para medir las capacidades, según su desarrollo en nuestra cultura. En parte porque decir que un determinado grupo es más inteligente que otro puede resultar hiriente, se modificó el nombre de test de inteligencia a test de aptitud académica. Los ítems se modificaron mínimamente, sólo se cambió la etiqueta. La inteligencia es algo real. La aptitud es real. Los logros son reales. Pero las aptitudes y los logros son dos cosas bien distintas.

La presión política y los contratos de evaluación forzaron a los evaluadores a poner el énfasis en la medición de la calidad en la escuela. Estas presiones crean dudas en la opinión pública en cuanto a los logros de las escuelas: las escuelas son caras, no cumplen todo lo que prometen y no satisfacen nuestras expectativas. Se les presiona para que “rindan cuenta de sus actos”. Para medir esto, los evaluadores en los Estados Unidos tomaron los ítems de los tests de aptitudes académicas (antes tests de inteligencia) y empezaron a llamarlos ítems de rendimiento del alumno. Lo que en su día medía las aptitudes ahora mide el rendimiento.

En Estados Unidos hemos empleado tests con millones de estudiantes. Era obligatorio en todos los estados. Se ignoraba que los tests de rendimiento tienen poco valor diagnóstico y poquísima validez externa. A veces, las puntuaciones medias se analizan en algunos debates sobre política educativa. En ocasiones, se sanciona a los que obtienen los peores resultados. Es un mundo de fantasía. Sam Wineburg lo explicaba muy bien en su trabajo *Crazy for history* (2004). También lo hizo Neil Postman en su libro *The End of Education* (1995). Los tests no sirven para determinar la eficacia educativa y, sin embargo, la política federal estadounidense en materia de educación y las de la mayoría de los 50 estados están basadas en gran medida en los resultados (puntuaciones) de las pruebas.

Estas mediciones aluden a los logros, pero no los miden. Incluso así, quizás sirvan para motivar a los niños. No sabemos cuánto bien o cuánto daño les están cau-

sando a los niños pero nosotros, como evaluadores, no podemos hablar como si de ese modo estuviéramos midiendo la calidad escolar o los logros de los alumnos. No lo estamos haciendo.

3. A MODO DE CONCLUSION

La formación de los evaluadores no tiene que ver únicamente con determinar qué rendimiento satisface ciertos estándares de calidad, sino de “experimentar” situaciones de éxito y de fracaso. Más que centrarse en buenos ejemplos de evaluación, los cursos de evaluación de programas, creo, deberían prestar mucha atención a los errores del evaluador, las dificultades y la resolución de problemas.

Los buenos centros de evaluación también deben localizar los problemas, los errores y el comportamiento imitativo. Necesitan desarrollar vías de comunicación para hablar de los problemas, y de la mala calidad, con un público que no quiere hablar de esos temas. Para los que creamos centros de evaluación y formamos evaluadores el reto consiste en hacer que nuestra comunicación esté contextualizada, sea sensible con los problemas y se base en la experiencia.

REFERENCIAS BIBLIOGRÁFICAS Y DOCUMENTALES

- Margin, S. (1992), *The Wellness Encyclopedia of Food and Nutrition*. Berkeley: University of California.
- Postman, N. (1995), *The End of Education*. Nueva York: Vintage.
- Scriven, M. (1967), “The methodology of evaluation”, en R.E. Stake, ed., *Perspectives of curriculum evaluation*. Chicago: Rand McNally.
- (1996), “The skeleton in the disciplinary closet”, Millercom Lecture, Universidad de Illinois, 23 de abril.
- Stake, R. E. (2004), *Standards-Based and Responsive Evaluation*. Thousand Oaks, Ca: Sage.
- Wineburg, S. (2004), “Crazy for history”, *The Journal of American History*, 90(4): 1401-1414.