

LA EVALUACIÓN EXTERNA Y SUS IMPLICACIONES. ASPECTOS TÉCNICOS, PRÁCTICOS Y ÉTICOS

Escrito por Ramón Pérez Juste

RESUMEN

El presente artículo, centrado en la evaluación externa, se escribe por invitación de los responsables de la revista.

Para el autor, la evaluación en contextos pedagógicos debe ser eminentemente formativa y, por tanto, continua, y orientada a la mejora de los destinatarios de todos los esfuerzos que realizan los poderes públicos, los profesionales de la educación (directivos, profesores y especialistas) las familias, el personal de los centros y hasta los propios educandos a partir del momento en que toman conciencia de su necesaria implicación en la propia formación.

Sin embargo, la evaluación debe servir, también, a funciones de control, de rendición de cuentas y de servicio a los responsables de los sistemas educativos y de las organizaciones educativas. Ahí cobra protagonismo la evaluación externa, en principio una evaluación realizada por profesionales expertos, cuyo hacer debe estar presidido por el cumplimiento de determinadas exigencias técnicas y éticas, muchas de las cuales están contenidas en los estándares de calidad de las evaluaciones.

El autor pasa revista a la problemática apuntada, sugiriendo la conveniencia de formativizar, en lo posible, la evaluación externa, de orientarla a lograr organizaciones que aprenden, de buscar, hasta donde sea posible, la convergencia de los modelos externos y, todavía más, de estos con los modelos evaluativos de organizaciones educativas, facilitando así la optimización de medios y recursos y el logro de contribuciones relevantes a la mejora de las personas, de las instituciones y del sistema educativo.

ABSTRACT

EXTERNAL EVALUATION AND ITS IMPLICATIONS. TECHNICAL, PRACTICAL AND ETHICAL FEATURES.

By Ramón Pérez Juste, Professor at the National University for Non-Attendant Students and President of the Spanish Society of Pedagogy.

The present article, focused on external evaluation, has been written on special request of the magazine Managing Board.

The author's outlook upon evaluation in pedagogical contexts regards evaluation as an essentially formative procedure, and, consequently, a type of permanent formation, focused on the improvement of every human segment with a part to play in Education, such as teachers, managing boards, families, staff at school and even students themselves as they realize their personal concern in their own qualification process.

External evaluation must facilitate functions of control, accountability and permanent educational benefits to people with responsibility in educational systems or educational organizations. When so, external evaluation gains importance and defines itself as a kind of evaluation carried out by highly-experienced professionals whose duties must start by checking out the

fulfilment of professional ethics and technical requirements which, in most cases are included in evaluation quality standards.

The author suggests the convenience of setting a frame for external evaluation, with the aim of producing in the first place organizations committed to a permanent learning process, secondly, a tendency to unify external models, whenever it is possible, and finally, an approach between the latter and evaluating frames of educational organizations. According to him, this would facilitate an optimisation of funds and human resources, which would eventually contribute to the improvement of people, institutions and the educational system.

INTRODUCCIÓN

Creo no exagerar si afirmo que el tema de la evaluación es uno de los que más trabajos ha dado a la luz en las últimas décadas -en nuestro país, a partir de la Ley General de Educación de 1970- y de los que más preocupan, por diversas razones, al profesorado, al alumnado y sus familias, a los responsables de los centros educativos y, desde luego, de los sistemas educativos en sus diversos niveles y competencias.

La evaluación en la ordenación de nuestro sistema educativo

El protagonismo que viene asumiendo la evaluación en los últimos años, en nuestro país y en el ámbito internacional en que España se sitúa, puede explicarse en alguna medida por su presencia en todas las grandes leyes de la democracia, desde la LODE [entre otros en los arts. 6. C, 42.1.f), 45.1. c)], a la reciente LOE, pasando por la LOGSE, donde ocupa un lugar destacado en el Título cuarto, dedicado a la Calidad de la Enseñanza, la LOPEG, en el Título III, y la LOCE, donde aparece entre los cinco ejes que la orientan, en los diferentes niveles educativos y en su Título VI.

También en la actual Ley de Educación, L.O.E, (LEY ORGÁNICA 2/2006, de 3 de mayo, de Educación, BOE de 4 de mayo) la evaluación tiene una fuerte presencia; podemos encontrarla en la exposición de motivos, en los diversos niveles educativos, en las responsabilidades del profesorado, en el capítulo III, centrado en los órganos colegiados de los centros y, en particular, en el Título VI, dedicado a la Evaluación del Sistema educativo[1]

La evaluación en la educación superior

Es interesante resaltar al respecto que la importancia de la evaluación ha traspasado los límites de los niveles obligatorios, y hasta de los post obligatorios no universitarios, llegando a la Universidad. En la Ley Orgánica de Universidades[2] también la evaluación ocupa un lugar muy relevante, siendo citada en 44 ocasiones; es más, también aparece reiteradamente en Ley Orgánica 4/2007, de 12 de abril, que la modifica en algunos de sus artículos.

Pero es que esta tendencia está muy generalizada en los grandes documentos internacionales. En dos de ellos, de gran relevancia, como son la *Declaración Mundial sobre la educación superior en el siglo XXI: Visión y acción*, y el *Marco de acción prioritaria para el cambio y el desarrollo de la Educación superior*, aprobados por la Conferencia Mundial sobre Educación Superior, organizada por UNESCO en París, en octubre de 1998, la evaluación tiene presencia en los artículos 9, d) y 11, en el primer caso, y en el apartado II.6 en el segundo.

Nada digamos sobre la importancia que cobra la evaluación en el marco de la denominada convergencia europea y, en concreto, en lo relativo al Espacio europeo de Educación superior (EEES), en concreto en las declaraciones de la Sorbona

(1998), Bolonia (1999), Praga (2001), Barcelona (2002), Berlín (2003) o Bergen (2005).

Los modelos de calidad y la evaluación

Otro aspecto que explica la atención dedicada a la evaluación deriva de su importancia en la actual preocupación social por la calidad, en este caso entendida como *gestión de la calidad*, que se abre camino a pasos agigantados debido al interés de las organizaciones educativas por la *certificación*. La preocupación por la sana competitividad, por la movilidad de estudiantes y profesores, por su fidelización y por la visibilidad de la calidad viene ganando adeptos y son variados los modelos disponibles y las agencias dispuestas a entrar en estos temas previa su *acreditación* como entidades *certificadoras*.

Modelos como ISO (*International Organization for Standardization*) o EFQM (*European Foundation for Quality Management, Modelo Europeo de Excelencia Empresarial*) están plagados de referencias a la evaluación. Es más, si ISO le dedica todo el apartado 8, la aplicación de EFQM implica la autoevaluación de la organización de que se trate.

Repercusiones sociales de la evaluación

Por otra parte, los resultados de la evaluación tienen una enorme repercusión social; afectan al alumnado, por supuesto; pero, también, al profesorado, a los responsables de los centros educativos y a las familias. La aparición de los resultados de ciertas evaluaciones ocupa las portadas de la prensa y demás medios de comunicación y se constituye en arma arrojada entre los partidos políticos, con argumentos diferentes según estén en el gobierno o en la oposición.

Sin embargo, y con más frecuencia de lo deseable, quienes hablan o escriben sobre tales informes lo hacen sin tener otra referencia que los propios resultados, sin analizar los contenidos evaluados, los contextos de referencia, las metodologías utilizadas y, en ocasiones, ni siquiera los propios resultados en evaluaciones anteriores.

Estos hechos justifican la necesidad de analizar la temática, de reflexionar sobre ella y de plantear algunos criterios que sitúen las cosas en su sitio.

Una de las primeras cuestiones a considerar es la relativa al estatuto de la evaluación, bien como herramienta, instrumento o medio, bien como finalidad. Ya sé que es difícil encontrar alguien que mantenga que la evaluación tiene sentido en sí misma; pero no lo es, sino todo lo contrario, la existencia de planteamientos y comportamientos perfectamente compatibles con tal posición.

Para quien escribe, clarificar el estatuto de la evaluación exige encuadrarla en el marco de la naturaleza de los actos educativos y, en los tiempos que corren, en los que tanto se habla de calidad en general, y de la educación en particular, enmarcarla en las actuaciones a su servicio.

Esa y no otra debe ser la perspectiva de lo que se ha dado en denominar "cultura de la evaluación" tan frecuente entre nosotros en los últimos tiempos. Con esta afirmación no nos negamos a otros planteamientos, pero sí queremos dejar clara constancia de que, en términos pedagógicos, ese es su verdadero sentido.

1. LA EVALUACIÓN EN CONTEXTOS EDUCATIVOS

Si bien toda evaluación, por su propia naturaleza, debe consistir en un proceso de valoración y en su consiguiente resultado, parece razonable admitir que cada contexto tiene sus propias exigencias, derivadas, justamente, de sus características.

1.1. Naturaleza de la evaluación

Como acabamos de señalar, evaluar equivale a valorar, a atribuir valor a la realidad evaluada. Sin embargo, y aunque cabe pensar en evaluar por evaluar, o en profesionales de la evaluación que concluyen su tarea con la emisión del juicio de valor haciéndolo llegar a quienes le encargaron la evaluación^[3] el verdadero sentido de esta es el de servir de base a la toma de las decisiones que correspondan: comprar un piso o una joya, leer un libro, ir a ver determinada película u obra teatral..., aunque, en estos casos, la decisión la tomen personas diferentes del "evaluador".

Este carácter instrumental de la evaluación se aprecia de manera más nítida en todas aquellas profesiones en las que la evaluación precede a las actuaciones específicamente profesionales, como puede ser el caso de la Medicina, donde el diagnóstico -verdadera evaluación- es una etapa previa a la auténtica actividad del médico: tratar de curar.

Parece claro, en consecuencia, que la naturaleza de la evaluación radica en el acto de valorar, que debe servir de base y fundamento para las decisiones que procedan.

1.2. La educación como contexto para la evaluación

Los contextos educativos se caracterizan por su orientación a la mejora de las personas; todas las sociedades y, en particular, las avanzadas, han hecho de la educación un derecho de las personas en las primeras etapas de la vida y un deber de aquellas que son responsables de su correcto desarrollo.

Dado el carácter instrumental de la evaluación, su sentido primigenio en los contextos educativos no puede ser otro que el de proporcionar la información necesaria y adecuada para facilitar las decisiones que hagan posible tal mejora en las condiciones más favorables.

Sin embargo, dada la complejidad de la educación, en la que confluyen personas en funciones muy diferentes -padres, profesores y los propios alumnos-, medios, recursos y organizaciones en las que tales personas y medios se ordenan hacia la mejora, no debe caber duda de que la evaluación debe incidir, también, sobre las propias organizaciones y su personal pues su mejora o perfeccionamiento redundará en la mejora de las personas a educar.

No cabe duda de que el verdadero sentido de la enseñanza no es la propia enseñanza sino el de facilitar el aprendizaje y la formación de los alumnos; como no debe caber duda de que la mejora del ambiente en las aulas, de las relaciones humanas en los centros educativos, de la gestión, de la coordinación o de la participación, aunque pueda tener otras repercusiones, encuentra en el perfeccionamiento del alumnado todo su sentido y utilidad.

Por tanto, los programas de los profesores -tanto su concepción como su diseño, implantación, implementación y resultados- y los propios centros educativos, en la medida en que, en el sentido más amplio posible, son auténticos medios y recursos al servicio de la mejora de las personas, deben ser objeto de evaluación. Renunciar a ello es hacerlo a disponer de la información precisa para promover las decisiones de mejora que correspondan, tanto rectificando lo que no va bien como introduciendo cambios que potencien lo que va razonablemente bien o afianzando lo que funciona correctamente.

Los planteamientos anteriores nos conducen a una importante conclusión: aunque la evaluación puede tener, y las tiene, otras importantes funciones o finalidades, la fundamental en los contextos educativos no puede ser otra que la conocida desde Scriven como *formativa*, esto es, aquella que se lleva a cabo con el ánimo de favorecer y hacer posible la mejora de la realidad evaluada.

1.3. El sentido de las evaluaciones externas

Acabamos de señalar que, si bien la función fundamental de la evaluación en contextos educativos no puede ser sino la formativa, también puede servir a funciones diferentes. Y así es.

En efecto: si la educación es un derecho de los ciudadanos y un deber y responsabilidad de quienes deben asegurarla desde las instituciones sociales, en particular desde las autoridades y desde las familias, también es un hecho que la sociedad, en cumplimiento de tal deber y responsabilidad, invierte cuantiosas sumas en hacerla realidad.

No debería extrañar, por tanto, que la sociedad asuma, en consecuencia, funciones de *control* sobre el sistema educativo; asegurarse de que se respetan las normas, que el personal cumple sus obligaciones, que los recursos invertidos se utilizan con eficiencia y eficacia es, por tanto, otra importante función de la evaluación; o mejor, dos importantes funciones, las conocidas desde Nevo como *administrativa* y *sociopolítica*.

Por lo general, este tipo de evaluaciones son realizadas por personas expertas, con o sin responsabilidades en el sistema educativo; y ello, tanto para asegurar su independencia en relación con los protagonistas de la realidad evaluada, tratando de asegurar así la imparcialidad, como para llevarlas a cabo con la debida calidad técnica.

1.3.1. Del puro control a la formativización de las evaluaciones sumativas

Valdría la pena, no obstante, que, puesto que se llevan a cabo, al menos de forma indirecta, en un contexto educativo, se planteara una intencionalidad formativa aun en el caso de aquellas evaluaciones con decidida función de control.

Esta "formativización" de las funciones no estrictamente educativas de la evaluación no es sólo cuestión de intencionalidad, que también, sino que implica actuaciones técnicas a las que nos referiremos más adelante[4] Cuando las evaluaciones externas, previa la intencionalidad formativa, se diseñan adecuadamente, pueden ofrecer no sólo los resultados de la misma sino facilitar las decisiones adecuadas para mejorar la realidad evaluada.

2. LA EVALUACIÓN EXTERNA

Como hemos reseñado, la evaluación puede servir a funciones claramente diferentes, siempre sobre la base de su naturaleza o esencia: valoración, mediante la aplicación de determinados criterios preespecificados, orientada a la toma de decisiones.

2.1. Las grandes funciones de la evaluación

En el ámbito educativo las funciones clásicas han venido siendo las conocidas a partir de Michel Scriven como sumativa y formativa. Sin embargo, y al margen de otras menos comunes[5] cabe destacar la importancia de las denominadas "administrativa", orientada al control, y "sociopolítica", centrada en el logro de apoyo público para determinados proyectos.

La necesidad de control por parte de los responsables de las organizaciones educativas y, desde luego, de aquellos a quienes se encomienda la dirección del sistema educativo, no parece exigir demasiada argumentación.

Tanto desde la perspectiva de asegurar el cumplimiento de la normativa como desde la necesidad de disponer de información adecuada, precisa y relevante sobre diversos aspectos, entre ellos la eficiencia en el uso de los recursos o la eficacia del conjunto de medidas -proyectos, planes, programas- la evaluación representa una fuente

importantísima de información que permite a los responsables ajustar las decisiones y tomar las medidas que fueren precisas para corregir la situación, introducir nuevos elementos o potenciar la realidad existente.

Por otra parte, en un mundo globalizado como el actual, donde la movilidad del ser humano, tanto voluntaria como forzosa, es una realidad en continuo crecimiento, se hace evidente la necesidad de establecer procedimientos que permitan situar los resultados del sistema en el conjunto de otros sistemas educativos, disponer de referencias para la valoración de los propios resultados e, incluso, facilitar la toma de decisiones de mejora mediante procedimientos comparativos y de emulación (*benchmarking*). Es más, la deseable movilidad de las personas requiere procedimientos que faciliten la comparabilidad de la formación recibida, evitando trabas innecesarias y obstáculos sin sentido.

2.2. Evaluación externa y apoyo a la organización.

Considero suficientemente justificada una evaluación externa orientada al control tanto de las organizaciones como de los programas e, incluso, de los aprendizajes. Pero, entiendo, también queda convenientemente argumentado que esta finalidad puede, y si puede, debe, compatibilizarse con la que es inherente a toda evaluación pedagógica.

Por tanto, en la medida de lo posible, las evaluaciones externas, ligadas en lo fundamental al control y a la rendición de cuentas, deberían orientarse a la mejora de las organizaciones educativas, para lo cual es imprescindible que la información que en ellas se recoja resulte relevante a la hora de facilitar la toma de decisiones de mejora.

En este sentido, puede resultar de extraordinaria importancia, valor y utilidad[6] ligar la evaluación a uno de los enfoques más prometedores para la mejora de los centros educativos y de su personal: el de las *organizaciones que aprenden*. En efecto: en la medida en que la información recogida lo permita, las organizaciones educativas -escuelas, colegios, institutos- pueden acumular información valiosa, por organizada y sistemática, que pueda dar lugar a una rica experiencia, fundamental para crecer y mejorar sobre la base del conocimiento acumulado. Lo contrario representa algo muy común entre nosotros pero sin justificación alguna: tener que abordar todas las situaciones como si fueran nuevas, afrontar los problemas como si no se hubieran dado antes, tomar decisiones como si no se dispusiera de experiencia sobre las respuestas adecuadas e inadecuadas[7]

Esta posibilidad requiere determinadas cualidades técnicas en el contenido de los instrumentos de recogida de datos, entre las que cabe reseñar las relativas a la validez, ligada a la representatividad de su contenido y de las muestras en las que se recoge la información, algo difícil en el primer caso como consecuencia de la tendencia generalizada a dar respuesta, a través del currículo, a la diversidad social, cultural, organizativa y personal. Junto a la validez, la fiabilidad, que reduzca al mínimo los denominados errores de medida. Pero merece la pena destacar una cualidad que vengo denominando *valía* de la información, algo que ligo en lo esencial a su riqueza, definida en términos de matización, concreción y especificidad[8]

Un aspecto importante para que las evaluaciones externas permitan la mejora de las organizaciones educativas tiene que ver con la convergencia de los modelos, técnicas e instrumentos utilizados en la evaluación de los sistemas educativos (caso de los indicadores) y los que se aplican en los centros educativos[9]

En efecto: una razonable convergencia de ambos permitiría a los centros educativos, que es donde se encarnan las decisiones en pro de la calidad, disponer de una doble referencia: la idiosincrásica, en la que cada organización educativa pueda contrastar

su avance, estancamiento o retroceso en relación con evaluaciones previas, y la normativa, al apreciar si sus resultados evidencian progreso o mejora en relación con otras organizaciones. Cabe reseñar, por otra parte, que tal convergencia no es una utopía en muchos de los indicadores que se incluyen en los diferentes sistemas.

3. DIMENSIÓN TÉCNICA DE LA EVALUACIÓN EXTERNA

Una de las acciones más comunes entre los seres humanos es la valoración de cuanto ocurre a su alrededor. De hecho, podríamos afirmar que en una gran mayoría de nuestras conversaciones estamos emitiendo juicios de valor sobre las personas con quienes nos relacionamos, sobre sus comportamientos, sobre las cosas que nos rodean...

Pero es evidente que la mayor parte de tales valoraciones carecen de las características propias de una evaluación en sentido técnico, en general y en el ámbito educativo en particular. Suelen estar atravesadas de prejuicios, basarse en informaciones o datos incompletos o sesgados y ser utilizadas para los fines más diversos, entre los que no suele ser el principal el de mejorar la realidad evaluada, sea esta una persona, un comportamiento o un objeto.

Por tanto, no estamos hablando de este tipo de evaluaciones sino de aquellas que, por cumplir determinados requisitos técnicos, permiten a quienes corresponda tomar decisiones adecuadas y pertinentes que, de forma directa o indirecta, puedan contribuir a mejorar la realidad evaluada.

3.1. El objeto de la evaluación y su función

A la hora de establecer tales requisitos es preciso introducir un elemento nuevo: la naturaleza del objeto evaluado. No es lo mismo evaluar un centro educativo que los centros educativos de una comunidad autónoma; no es lo mismo evaluar el programa de un profesor que los programas de una determinada materia en un ámbito territorial concreto o en un nivel o etapa educativa; no es lo mismo evaluar el aprendizaje de los alumnos en general que hacerlo de determinadas competencias; y no es lo mismo, por cerrar este punto, abordar la evaluación externa, o interna, de centros educativos que de programas o de los aprendizajes, haciéndolo con la intención conocer el funcionamiento del sistema o la situación concreta de un centro educativo.

Por motivos obvios, debemos centrar nuestra atención en algún aspecto concreto ya que un artículo no permite abordar una problemática tan amplia como diversa. En consecuencia, vamos a limitar nuestro foco de atención a la evaluación externa del aprendizaje, aunque algunas consideraciones sean de aplicación a algunos de los ámbitos citados.

La evaluación del aprendizaje de los alumnos en un ámbito concreto -ciudad, zona de inspección, provincia, comunidad autónoma o sistema educativo nacional- implica algunos requisitos clave, tanto más importantes cuanto más amplitud y extensión tenga la realidad evaluada. Evaluar el sistema educativo español y hacerlo sobre los niveles obligatorios exige un cumplimiento exquisito de tales requisitos, entre los que cabe destacar los relativos al muestreo, cuya cualidad fundamental no es otra que la representatividad de las muestras; la validez, que asegura la coherencia tan plena como sea posible entre la realidad y la información que la representa, y la fijación de ciertas puntuaciones y niveles como exigibles para considerar que los resultados son suficientes y satisfactorios. Veamos, a continuación, estos tres aspectos de carácter técnico.

3.2. Problemas de muestreo

Una evaluación externa del aprendizaje de los alumnos pretende, por lo general, conocer el grado en que el sistema funciona, el grado en que los objetivos básicos -en

la forma que se formulen: capacidades, competencias, contenidos conceptuales, procedimentales o actitudinales- se alcanzan por todo el alumnado o por los alumnos de diferentes cursos, niveles o etapas.

Poder hacer afirmaciones de esta naturaleza implica asegurar la representatividad de las muestras utilizadas para la recogida de datos y ello comporta acudir a procedimientos que permitan disponer de muestras con el tamaño necesario^[10] y que gocen de la debida imparcialidad en el proceso de selección^[11]

3.3. Problemas de validez

Sin embargo, una muestra imparcial y representativa no es suficiente, aunque sí sea imprescindible.

La otra gran cualidad que debe reunir la evaluación tiene que ver con la validez de la información recogida, de los datos proporcionados por los instrumentos de medida utilizados.

Esta cualidad es esencial, pero asegurar que tal información representa el dominio de los objetivos a evaluar es una tarea ímproba por diversas razones, entre las que no tienen menor importancia las relativas a la amplitud y profundidad de la realidad evaluada o a la diversidad de objetivos del currículo.

En efecto: no es lo mismo evaluar el programa de Lengua castellana de cuarto de primaria que el currículum de la educación obligatoria, como tampoco lo es en el caso de una comunidad autónoma que de todo el sistema educativo, en el que, en función de sus competencias, el currículo puede presentar diferencias notables. Es más: en función de las prescripciones legales, el currículo de cada centro educativo puede presentar peculiaridades que, de no tenerse en cuenta, pueden dar lugar a resultados que no cumplan con la condición de validez.

Por otra parte, dada a diversidad de objetivos -conceptuales, procedimentales y actitudinales- así como la irrupción de las competencias en el concepto de currículo (art. 6 de la LOE) lograr la validez implica una serie de actuaciones técnicas difíciles de implementar.

Sea como sea, y no es este el lugar para abordar con la mínima profundidad estas exigencias, quede constancia de que la validez de los datos es fundamental a la hora de evaluar los niveles de logro del alumnado. Dejemos constancia, eso sí, de la conveniencia de acudir a enfoques evaluativos como las pruebas referidas a criterio, en lugar de las clásicas pruebas de referencia normativa, como una estrategia técnica fundamental para lograr la calidad -validez- de la información.

3.4. Establecimiento de estándares y puntos de corte^[12]

Pero asegurar la validez de la información recogida no evita otro problema de gran entidad, cual es el de la interpretación de los resultados.

Debemos tener en cuenta que ciertos valores numéricos, como las calificaciones escolares, nos resultan fácilmente interpretables porque, aparte de la familiaridad que hemos alcanzado en su manejo, hemos asumido determinadas equivalencias entre los números y la calificación que llevan aparejada: 3, insuficiente; 6, bien; 8, notable; 9, sobresaliente; 10, excelente. Sin embargo, es cierto que si profundizáramos en tales números podríamos encontrarnos que su asignación no es clara, hasta el punto de que las investigaciones ponen de relieve que un mismo evaluador puede asignar valores claramente diferentes -del insuficiente al sobresaliente- a unas mismas pruebas en momentos diferentes y que, por supuesto, este hecho se da claramente cuando los evaluadores son personas distintas.

Este hecho pone de relieve una cierta arbitrariedad en las calificaciones, bien porque diferentes evaluadores se fijan en aspectos distintos de la realidad evaluada, bien porque conceden más o menos importancia a determinados componentes de la misma, bien porque unos son más rigurosos que otros al calificar, bien porque...

La gran cuestión, pues, es establecer con claridad las reglas que permiten establecer los niveles que se consideran suficientes porque representan el dominio de la materia evaluada. Y, en ese punto, debemos reconocer que las reglas no existen y que, en consecuencia, deben establecerse por consenso entre personas expertas. Pero deben establecerse, aunque sea difícil, porque, si no fuera así, unos mismos resultados podrían alcanzar diferentes valoraciones lo que, además de injusto para las personas, sería confuso para la sociedad a la que se pretende informar con este tipo de evaluaciones[13]

La fijación de estándares, esto es, de niveles de dominio claramente definidos que deben ser establecidos como norma, patrón o referencia a lograr en determinados aprendizajes es una condición básica para ordenar la enseñanza y, por supuesto, la evaluación del aprendizaje. Insisto: la enseñanza, y su evaluación, porque la condición básica para el logro de la validez es la coherencia entre lo evaluado y aquello que debe ser objeto de enseñanza / aprendizaje, siendo este último aspecto el elemento directriz y nunca al contrario.

Por último, hay una cuestión técnica que, también, debe resolverse: la puntuación a partir de la cual se considera que se da el dominio de la realidad evaluada. Una decisión de esta naturaleza puede tomarse por dos procedimientos fundamentales: a partir del juicio de expertos en las diversas materias y en las características del alumnado evaluado, o de forma empírica, estableciendo tales puntuaciones -aunque pueda parecer un contrasentido- a partir de los resultados obtenidos por muestras de sujetos tomados como norma o patrón[14]

Si el primer procedimiento se justifica, precisamente, por su carácter de personal especializado, experto, con un elevado nivel de dominio de la materia y de las características evolutivas de los alumnos cuyo aprendizaje se va a evaluar e, incluso, con la suficiente experiencia en materia de enseñanza / aprendizaje, el segundo encuentra su apoyo en la realidad objetiva. Parodiando las palabras de M. Bunge, referidas a la necesidad de doble contrastación, teórica y experimental de los procedimientos o técnicas experimentales[15] podríamos afirmar que es necesario comprobar en la realidad lo que la teoría -el juicio de los expertos en este caso- mantiene, dado que un amplio abanico de circunstancias -contextuales, sociales, familiares...- pueden alterar lo que, en condiciones normales se considera razonable.

Para terminar, recordemos que, por motivos de espacio, no cabe presentar la problemática anterior aplicada a los diferentes objetos de evaluación, en concreto, al menos, los relativos a la evaluación externa de programas y de centros educativos, cada una de ellas con obvias peculiaridades.

4. IMPLICACIONES DE LA EVALUACIÓN EXTERNA

Si, como venimos manteniendo, el sentido último de toda evaluación en contextos pedagógicos debe ser la mejora, bien directamente cuando la evaluación toma funciones *formativas*, bien de modo indirecto en el caso de evaluaciones orientadas al control, administrativas y sociopolíticas, lo cierto es que toda evaluación, por el mero hecho de realizarse, conlleva consecuencias, esto es: influye, alterándola para bien o para mal, la realidad evaluada.

Las consecuencias deberían siempre ser positivas; sin embargo, si no se toman las medidas oportunas, lejos de dar lugar a la mejora -sea ésta directa o por vías

indirectas- puede producir efectos contraproducentes, algunos de los cuales se conocen técnicamente como *efectos no planeados*, deseados o no.

Señalaremos a continuación algunos de los efectos que deberían evitarse, destacando genéricamente el influjo, buscado o no, en el currículo de los centros educativos.

4.1. Evaluación y calidad

Las investigaciones realizadas en relación con la eficacia de la Escuela resaltan como uno de los factores que contribuyen a su *eficacia* el contar con un adecuado sistema de evaluación[16]. En la medida en que la eficacia es uno de los componentes de la calidad, podremos mantener que un buen sistema de evaluación contribuye a la calidad.

Y lo es porque la evaluación en funciones formativas hace extraordinarias aportaciones para la mejora tanto del aprendizaje como de la enseñanza, esto es, de los programas de los profesores, siempre que demos a término "programa" el significado profundo que supone integrar su objeto -contenidos y objetivos- con su dimensión de *plan*, ordenado y sistemático, a su servicio.

La aportación fundamental desde esa perspectiva es, o puede ser si la evaluación está bien concebida y realizada, una información útil y valiosa no ya de los resultados o logros, que también, sino de lo que va bien -puntos fuertes- y lo que deja que desear -aspectos mejorables; pero, sobre todo, tal información orienta sobre las razones de lo que va bien y permite identificar las de aquellas que están detrás de lo que no va tan bien, facilitando así las decisiones adecuadas para la mejora. Este hecho, unido a su oportunidad -la información se recoge en momentos próximos a los procesos educativos- potencia su contribución a la mejora.

Por otra parte, y en una dimensión no tanto técnica como, tal vez, ética, estos planteamientos evaluativos implican en quienes los emprenden una cierta actitud de humildad: asumir su corresponsabilidad, junto al alumnado, en los resultados del aprendizaje, y hacer suyo un principio fundamental de la mejora continua: por muy bien que se hagan las cosas, siempre pueden mejorarse.

Pues bien, por la propia naturaleza de las cosas, todos estos aspectos no suelen estar presentes en las evaluaciones externas, razón por la que, de no tomar determinadas precauciones, la contribución a la calidad puede no darse e, incluso, producirse un efecto contraproducente, derivado de alterarse el sentido instrumental de la evaluación -medio al servicio de los objetivos- para convertirse ella misma en objetivo o finalidad con sentido propio.

En efecto: del mismo modo que un alumno puede tener como meta aprobar en lugar de saber, un centro educativo puede plantearse quedar bien en un determinado *ranking* puede ordenar sus esfuerzos y recursos a superar las pruebas de evaluación sin preocuparse por aquellos objetivos que no suelen formar parte de las mismas o que lo hacen con menor presencia o menor peso[17]

Precisamente para evitarlo, es necesario partir de proyectos educativos de calidad -de conformidad con los planteamientos establecidos en la teoría y en las leyes- y diseñar instrumentos de recogida de información coherentes, al máximo, con aquellos. De no ser así, se estará trasladando el mensaje a los responsables de los centros educativos -directivos y profesorado- de que algunos contenidos / objetivos tienen poca relevancia por lo que, aunque no se trabajen, no importa o importa menos que si se dejan de trabajar otros[18]

En la misma dirección, y con un efecto concluyente, pueden ir en muchas ocasiones determinadas evaluaciones externas ligadas a procesos de *certificación*. La

certificación no debería ser sino el resultado de hacer bien las cosas, lo que lleva consigo, como exigencia, una buena planificación, organización y gestión, unos procesos bien diseñados, implantados e implementados, y unos logros fruto de tales acciones. Sin embargo, cuando la certificación no es el resultado natural de tales acciones sino la meta a lograr sea como sea, por el camino se pueden sacrificar objetivos y procesos por considerarlos subsidiarios en lugar de esenciales.

4.2. La reorientación de la actividad docente

Una de las consecuencias más negativas derivadas de esa alteración de la función instrumental de la evaluación es la modificación del verdadero currículo de centros y aulas, trastocando de hecho el proyecto educativo y curricular y hasta el programa de los profesores[19]

No hay nada de malo en que el profesor adiestre a su alumnado en el logro de determinadas destrezas, estrategias y habilidades, en la capacitación en competencias, en el dominio de conceptos. Pero sí lo hay si tales actuaciones no se corresponden con la totalidad de los contenidos / objetivos de su programa dado que estos son más ambiciosos que los que, de ordinario, suelen formar parte de las pruebas externas.

En definitiva: estamos manteniendo la necesidad de que el proyecto curricular y el programa del profesorado, en perfecta coherencia con un proyecto educativo de calidad, sean los que guíen la elaboración de las pruebas externas y no al contrario. Y para ello, debe quedar claro que los componentes de aquéllos, todos, pueden llegar a formar parte de éstas, con la debida presencia y ponderación. Cualquier alteración de este principio tendrá repercusiones negativas para la calidad de la educación, por muy buenos resultados que los alumnos de esos centros educativos puedan presentar y hasta presumir de ellos ante el cuerpo social.

4.3. Evaluación y forzamiento de las pautas de calificación del profesorado

Otro de los riesgos derivados de la evaluación externa tiene que ver con ciertas consecuencias de la misma para el profesorado.

Es verdad que, con más frecuencia de lo deseable, algunos profesores se sirven de la evaluación como un medio para doblegar al alumnado; como lo es que las pautas de evaluación de muchos profesores no se mantienen por una manifiesta incoherencia entre los objetivos explícitos de sus materias, su actuación docente y el contenido y criterios de evaluación aplicados a sus pruebas.

Son más frecuentes de lo deseable, en especial en bachillerato y en la universidad, los casos de profesores que crean auténticos cuellos de botella, con la acumulación de alumnos que no consiguen superar su asignatura si no es después intentarlo durante varios cursos académicos.

Pero no es menos cierto que no se puede poner al profesorado exigente en el brete de perder su trabajo o determinados reconocimientos -sociales o económicos- si no logra determinados niveles de éxito, medidos por lo general en términos de alumnos que superan la materia[20]. Porque si así se hace, se estará forzando su voluntad y llevándole a la conclusión de que podrá evitar este tipo de situaciones simplemente siendo más laxo a la hora de calificar.

Habrà, por tanto, que buscar soluciones equilibradas para evitar ambas situaciones indeseables. En el primer caso, estableciendo procedimientos previamente regulados para evitar enquistamientos indeseables. En el segundo, ayudando a los profesores a plantear programas razonables en función de la duración y extensión de las materias [21]. Y, en consecuencia, evaluaciones y pruebas plenamente coherentes con una definición adecuada de los programas.

5. ÉTICA Y EVALUACIÓN EXTERNA

Si desde una perspectiva pedagógica la evaluación debe tener un fuerte componente técnico para que puedan hacerse realidad las grandes aportaciones a la mejora a las que nos hemos referido, a las evaluaciones externas cabe exigirles, además, un conjunto de cualidades éticas dadas las repercusiones que de ella se pueden derivar para las organizaciones y para la comunidad educativa[22]

La aparición espuria de listas ordenadas de mayor o menor calidad (ranking) de las organizaciones educativas puede tener serias consecuencias para aquellas que aparecen en los peores puestos. Se podría señalar que si hay diferencias no hay nada malo en hacerlas públicas. El problema es que, por lo general, tales clasificaciones son muy discutibles tanto por los criterios utilizados para su elaboración[23] como por la forma en que se aplican las pruebas[24] y, sobre todo, en la integración de los resultados de esos diferentes criterios en un único valor que represente, con la debida ponderación, a todos ellos[25]

5.1. La imprescindible cualificación de los evaluadores

Aunque pueda parecer una obviedad, conviene reseñar como una cuestión ética la necesidad de que quienes lleven a cabo evaluaciones deban contar con la necesaria capacitación. El hecho de que, como hemos dejado constancia, la actividad evaluativa es común entre los seres humanos, lleva a muchas personas a asumir que pueden evaluar sin preparación profesional alguna.

Por otra parte, el que la evaluación confiera a los evaluadores un cierto estatus de autoridad, nunca debería llevar a comportamientos prepotentes para con aquellas personas cuya actividad sea objeto de evaluación.

Por último, conviene reseñar que, ante la importancia de la evaluación y sus consecuencias, se han elaborado códigos de comportamiento cuyo respeto debe ser exigido a todos aquellos que realicen evaluaciones, bien en forma profesional bien como aficionados o amateurs.

5.1.1. Cualificación profesional y técnica[26]

Uno de los problemas fundamentales para llevar a cabo la denominada "cultura de la evaluación" es la de disponer de profesionales debidamente cualificados.

La competencia debe referirse no sólo a la formación genérica sobre la teoría, modelos, técnicas e instrumentos para la evaluación, sino al dominio experto de las características específicas de los modelos de contraste. Parece una obviedad, pero no tiene sentido evaluar contra ISO, EFQM o cualquier otro modelo de calidad sin el dominio profundo de tales modelos.

Conviene hacer notar que una de las condiciones necesarias, aunque no suficiente, para que se puedan obtener consecuencias positivas de la evaluación es la relativa a la *credibilidad* del evaluador[27] y ésta no es posible sino a partir de personas que, en su hacer profesional, ponen claramente de relieve su capacitación y dominio.

5.1.2. Calidad humana

La evaluación puede llegar a representar para los evaluados, en sus actividades o productos o en sus características personales y profesionales, situaciones de desasosiego, inquietud, ansiedad o estados emocionales de naturaleza similar.

Las relaciones humanas, inspiradas en el respeto a la dignidad e integridad de todos los que intervienen en la evaluación, pueden reducir la tensión y la ansiedad, favoreciendo así incluso el que los resultados sean más válidos, en la medida en que no se han visto mediatizados por un contexto amenazador y bloqueante.

5.1.3. Respeto de los estándares de calidad

Como señalábamos, muchas profesiones han elaborado códigos de conducta para regular la actuación de los profesionales.

En nuestro país no se puede decir que exista la profesión de evaluador como tal, pero sí es cierto que son muchos los profesionales que llevan a cabo evaluaciones.

En 1974, un la Asociación Americana de Investigación en Educación y el Consejo Nacional de Medición en Educación, dio a conocer unos estándares para los Manuales y los tests de Psicología, revisando la edición de 1966. En 1975 el *Joint Committee on Standards for Educational Evaluation*, ampliado con la incorporación de doce nuevos miembros, comenzó un trabajo que dio a conocer en 1981; se trataba de la publicación *Standards for Evaluations of Educational Programs, Proyects and Materials*, en la que se propuso un total de 30 normas, agrupadas en cuatro grandes categorías, exigencias o criterios (cuadro 1).

Años más tarde, en 1994, tomando como base los principios de la publicación de 1981, se dio a conocer una revisión de los mismos bajo el título *The Program Evaluation Standards*, que incluye algunas modificaciones.

Los estándares, que pueden ser utilizados como instrumento de formación de evaluadores, han sido concebidos como un medio útil tanto para los evaluadores como para los usuarios de los programas; con ello se pretende que estos sean capaces de valorar las evaluaciones a las que puedan acceder.

El Joint Committee define un estándar como un principio acordado por las personas comprometidas en una práctica profesional, destinado a mejorar la calidad de la misma. Considera que los estándares "*proporcionan una guía para evaluar los programas de educación y formación, proyectos y materiales*"; en conjunto, y a su juicio, proporcionan una estructura para diseñar y asesorar evaluaciones.

Los estándares están organizados en torno a cuatro grandes criterios, cada uno de ellos con un conjunto de normas asociadas. Los criterios: *utilidad, viabilidad, honradez y precisión* representan exigencias de diferente entidad para los evaluadores; algunos de ellos son de difícil aplicación en nuestro país donde, como hemos señalado, no se da una verdadera profesión de evaluador.

Conviene dejar constancia en este punto de la importancia de las ocho normas ligadas al criterio de honradez o probidad, destinadas a asegurar una evaluación realizada con planteamientos legales y éticos, respetando a quienes intervienen en la evaluación y están interesados en las conclusiones que integran el informe, que debe ser equitativo.

5.2. Actuaciones de los poderes públicos y demás entes evaluadores

Uno de los problemas éticos ligados a la evaluación externa tiene que ver con la tendencia de quienes las encargan, por estar en el poder[28] a considerar que los informes son suyos. Tal apropiación puede llevarles a divulgar aquella parte que les beneficia, reservándose información que les perjudica o que pueda reportarles ventajas en la contienda política o profesional, o a desfigurar los resultados.

Al respecto, son clarificadoras las palabras de Stufflebeam y Shinkfield (1987); a partir de su conocida definición de evaluación[29] hablan de evaluaciones políticamente orientadas, "*que son dirigidas o utilizadas subrepticamente para provocar un punto de vista determinado, positivo o negativo, sobre un objeto, independientemente de la valoración objetiva de su valor o mérito*", a las que denomina pseudoevaluaciones. Pues bien, en relación con ellas se expresan con enorme dureza cuando afirman:

"Puede que recopilen la información rigurosamente, pero no revelan las verdaderas conclusiones, o lo hacen de modo selectivo, o incluso las falsifican, engañando tanto a la gente como a sí mismos". (págs. 67 s).

Los autores diferencian, dentro de ese grupo, entre "investigaciones encubiertas" y "estudios basados en las relaciones públicas". En relación con las primeras, afirman:

"Por lo general, el cliente quiere que la información obtenida sea tan sólida técnicamente como sea posible, pero también que se le garantice la posibilidad de controlar la revelación de la información".

Estos comportamientos son censurables, y así ha quedado establecido al incluir entre los estándares un conjunto amplio de normas, pertenecientes a los cuatro grandes criterios, como las siguientes:

***U1. Descripción de la audiencia.* Las personas implicadas o afectadas por la evaluación deben ser identificadas a fin de poder atender sus necesidades.**

***U3. Alcance y selección de la información.* La información recogida debe ser tan amplia y adecuadamente seleccionada como para incluir cuantas cuestiones se consideren pertinentes y relevantes en relación con el objeto de evaluación y para poder dar la debida respuesta a los intereses y necesidades de las diferentes partes interesadas.**

***U6. Difusión y oportunidad del informe.* Los hallazgos significativos y el informe de la evaluación deben darse a conocer a las partes afectadas e implicadas a fin de que puedan ser debidamente utilizados.**

***F2. Viabilidad política.* La planificación y dirección de la evaluación debe procurar la cooperación de los grupos interesados así como evitar cualquier intento de intervenir en su desarrollo o en sus resultados.**

***P1. Orientación al servicio.* Las evaluaciones deben diseñarse con la intención de ayudar a las organizaciones a atender y satisfacer las necesidades de los participantes.**

***P5. Informes claros y completos.* Los informes en que se reflejan las evaluaciones deben ser claros y completos, incluyendo en sus conclusiones tanto los aspectos positivos y favorables como los negativos y desfavorables, de forma que los primeros se conviertan en base de nuevas acciones y los segundos en elementos para las actuaciones de mejora.**

***P6. Derecho a saber.* Evaluador y cliente deben asegurar que los resultados de la evaluación, con sus correspondientes limitaciones, serán accesibles a las personas o grupos afectados, algunas de las cuales tienen derecho a recibir los resultados.**

***P7. Conflicto de intereses.* Los intereses en conflicto deben afrontarse de forma abierta y honesta, de modo que no interfieran en el proceso y en los resultados de la evaluación.**

***A10. Conclusiones justificadas.* Las conclusiones a las que llega la evaluación deben justificarse explícitamente a fin de que las diferentes partes implicadas y afectadas puedan someterlas a valoración.**

***A11. Objetividad del informe.* Los informes deben reflejar con claridad los resultados de la evaluación, evitándose toda distorsión debida a las posiciones personales de cualesquiera de las partes interesadas.**

6. PROBLEMAS PENDIENTES

Como se ha podido constatar, la evaluación es una actividad compleja, en la que se entrecruzan elementos de muy diversa naturaleza, lo que conduce a soluciones muy diferentes y, en ocasiones, muy alejadas unas de otras.

Tal *complejidad* da lugar a una serie de problemas sin resolver, sobre los que habrá que tomar decisiones en cada momento. Reseñamos algunos de los fundamentales, sin pretender en modo alguno haber recogido todos, aunque si los más importantes, al menos a juicio del autor, aunque reconociendo que ese calificativo depende, en ocasiones, de las circunstancias de la evaluación y de la naturaleza de la realidad evaluada.

6.1. La existencia de expertos

Si bien en la autoevaluación y, por lo general, en las evaluaciones formativas, no es necesario acudir a personal profesionalmente preparado, aunque siempre será conveniente una mínima capacitación, la evaluación externa es, o debe ser, una actividad técnica, necesitada de profesionales expertos que la conduzcan a buen puerto.

Con frecuencia, el diseño mismo de la evaluación lo realizan expertos, y personas expertas son las encargadas de recoger, organizar y elaborar la información como base para la configuración de un juicio técnico y hasta para la propuesta de medidas de mejora.

En nuestro país, diversas actuaciones de evaluación externa, como puede ser el caso del denominado *Plan EVA*, las evaluaciones del INECSE[30] o el Plan de Evaluación de la Calidad de las Universidades, de carácter mixto -primero, autoevaluación de la institución; después, evaluación externa por expertos- asumen la necesidad de expertos.

En este punto el problema es grave ya que no contamos ni de lejos con el número de *expertos* necesarios para llevar a cabo tal labor. Pero las consecuencias de una evaluación externa, propia de expertos, llevada a cabo por personas con una formación apresurada, concentrada en unas sesiones de fin de semana, son fáciles de imaginar.

No debemos olvidar que si, ya de por sí, la evaluación levanta recelos en las personas e instituciones evaluadas, la situación puede ser mucho más seria si a ello se une que quienes deban llevar a cabo la tarea no sean reconocidos por los evaluados como dotados de la competencia y de la experiencia necesarias y sí, simplemente, como personas que ocupan un puesto de autoridad.

La necesidad de competencia de los evaluadores, ya citada, es imprescindible; de no darse, lo más probable es que la evaluación pierda una gran parte de su efectividad y hasta pueda llegar a ser la gran ocasión perdida: en efecto, cuando los resultados de una evaluación defraudan y son insatisfactorios, resulta muy difícil intentarlo de nuevo.

6.2. La factibilidad

Entre los riesgos ligados a la creación e implementación de una cultura de la evaluación se encuentran los relacionados con los intentos de perfeccionismo.

En muchos casos los modelos suelen ser muy ambiciosos, planteando actuaciones evaluativas difícilmente implementables por los medios y recursos necesarios; puede ocurrir que exijan un personal no disponible, dedicación excesiva[31] materiales y recursos más allá de las posibilidades.

Conviene reseñar en este punto que uno de los criterios incluidos en los estándares de calidad de las evaluaciones es el relativo a la *viabilidad o factibilidad* de las mismas, integrado por tres normas, orientadas a asegurar que las evaluaciones sean realistas, prudentes y adecuadas. En concreto son las siguientes: a) *Procedimientos prácticos*; b) *Viabilidad política*; c) *Eficacia en los costes*.

La primera de ellas pretende que las evaluaciones ofrezcan la información necesaria pero creando los menores problemas posibles a quienes han de proporcionarla; evaluaciones perfectamente diseñadas pueden llegar a ser imposibles de llevarse a la práctica. La viabilidad reclama un ejercicio de búsqueda de evaluaciones que, sin renunciar a los aspectos esenciales, no fracasen por su nivel de utopía. La segunda norma implica la colaboración de las partes interesadas, algo difícil de lograr sin la debida credibilidad de los evaluadores y sin unos planteamientos prudentes y asequibles. La tercera plantea la necesidad de un costo -en tiempo, personal y medios- asequible y adecuado a las aportaciones esperables.

Algo de lo que es necesario huir a toda costa es de la mera acumulación de información que, luego, no es utilizada o que es claramente infrautilizada. El reto se puede plantear en los siguientes términos: conjugar una evaluación que merezca la pena por la aportación de información que pueda representar, y el tiempo, los recursos y el personal necesarios para lograrlo. Estamos hablando, por tanto, de eficiencia.

Para lograr estas metas es necesario investigar; en efecto, la investigación debería orientarse a identificar los contenidos mínimos necesarios para una evaluación que permita ser valorada como *suficiente*, dejando fuera todos aquellos otros aspectos que podríamos denominar *superfluos*. Se trata, en definitiva, de elaborar propuestas centradas en lo esencial, suficientes y factibles.

La exigencia de factibilidad resulta especialmente necesaria si se quiere crear una cultura autoevaluativa, la que, como sabe el lector, viene considerando el autor como la que representa mayores posibilidades de mejora.

6.3. ¿Juicios expertos o juicios democráticos?

Un aspecto especialmente delicado es el relativo a la toma de decisiones: ¿quién debe tomarlas? ¿Se acude al principio de autoridad, dejándolas en manos de quienes tienen la responsabilidad (directivos, poderes públicos)? ¿Se asignan al personal de las organizaciones? ¿Las deben tomar los expertos que han llevado a cabo la evaluación?

Debemos reconocer que el verdadero sentido de la evaluación no es el de llegar a los juicios valorativos, algo esencial pero que no tiene sentido en sí mismo, sino el de conducir a la toma de decisiones de mejora de los aspectos evaluados.

Pues bien: para hacer de ese planteamiento una realidad es necesario, y hasta imprescindible, contar con el asentimiento, la colaboración y hasta el compromiso del personal de las organizaciones que ha de llevar a cabo, de implementar, de hacer realidad, las decisiones derivadas de la evaluación. Y para lograr tal compromiso, nada mejor que plantear su participación en la interpretación de la evaluación, en el diagnóstico sobre la realidad evaluada.

Lo ideal sería que tanto unos -los juicios- como otras -las decisiones- fueran el resultado de *sesiones de evaluación* en las que, con una adecuada dinámica de grupo, los responsables de la institución alcanzaran el consenso sobre los aspectos fundamentales o, al menos, la aceptación y asunción como propios de los acuerdos mayoritarios. Pero si no fuera así, resolver las cuestiones en forma de votación no parece ser la respuesta más correcta.

Si se acepta que la *participación* es una cuestión de *grado*[32] y que, en consecuencia, admite diversas manifestaciones, desde la simple recepción y aportación de información hasta la co-gestión de las instituciones, pasando por la delegación de competencias, parece adecuado pensar que, en casos de falta de acuerdo final, sea en el diagnóstico de los puntos fuertes y débiles, sea en la selección de las medidas de mejora a aplicar, sea en la elaboración de los correspondientes planes tanto de mejora de la institución o del programa como de su misma evaluación, las posiciones últimas corresponden a los responsables de su gobierno, y ello como algo inherente a su responsabilidad directiva. No obstante, nunca se insistirá lo suficiente en la importancia de conseguir, mediante el diálogo y las sesiones de grupo debidamente conducidas, un acuerdo básico tanto en el diagnóstico de los puntos débiles como en las medidas a tomar: el apoyo a los planes de mejora depende en gran medida de este importante hecho.

6.4. La comparabilidad de los datos

En unos tiempos caracterizados, entre otras notas, por la globalización o mundialización, no debería extrañarnos que la comparabilidad de los resultados, sobre todo cuando se trata de programas a gran escala o de evaluaciones institucionales amplias que afectan al sistema educativo, a las universidades públicas (o a las privadas), a un conjunto de colegios de una misma institución o titularidad... sea una cuestión a abordar y resolver.

En ocasiones, este hecho viene impulsado por la función evaluativa de rendición de cuentas o por el cumplimiento del principio de derecho a la información para facilitar las decisiones de los ciudadanos.

El problema de la comparabilidad es de elevada complejidad desde una perspectiva puramente técnica/[33] pero se complica todavía más por algunas de las exigencias de nuestro tiempo en lo relativo al *currículum* de las instituciones o por los derechos consagrados tanto a la libertad de cátedra como a la autonomía institucional.

En efecto: la tendencia cada vez más acusada a que los *currícula* reduzcan su grado de centralización en los diferentes países, dejando autonomía a las diversas comunidades y regiones en que se estructura cada nación[34] y hasta a los propios centros para acomodar y completar las líneas básicas del currículo nacional, es una dificultad a salvar.

El caso de las Universidades puede ser todavía más agudo. Por principio constitucional, las universidades gozan de un elevado grado de autonomía en su organización, gestión y desarrollo de las actividades académicas. ¿Cómo comparar resultados de universidades con tamaño y complejidad muy diferentes, planes de estudio con notable diversidad, número de alumnos, profesores y personal distintos, *ratios* muy alejadas unas de otras...?

Un problema añadido es el relativo a las dificultades que este hecho representa para la movilidad estudiantil, tanto para las convalidaciones dentro del propio sistema cuanto para el reconocimiento de títulos entre diferentes países.

Una solución, al menos parcial, siempre que la comparación se oriente a la *mejora*, puede encontrarse en seleccionar como referencia no los estándares nacionales o internacionales sino los datos de la propia institución en una perspectiva evaluativa diacrónica. Estaríamos hablando en tal caso de *superación*, *avance*, *mejora*, *progreso* en lugar de *superioridad* y *competitividad*. Esta perspectiva podría hacerse compatible con una evaluación criterial: la referencia serían los niveles de eficacia, eficiencia, satisfacción... que la propia institución, tras el análisis de su situación, considerara razonablemente alcanzables. Las comparaciones entre los indicadores de los países quedarían limitadas a aquellos aspectos estructurales, funcionales y organizativos en

que los datos brutos pueden ponerse en relación con datos contextuales. Por ejemplo: % de alumnos escolarizados en los diferentes niveles; *ratios* de éxito (superación de los diferentes cursos) por etapas educativas; % del producto interior bruto invertido en educación; % de alumnos del Centro educativo que superan las pruebas de selectividad y cuota de estos que culminan, después, los estudios universitarios en cada una de las carreras...

Conviene señalar que la comparabilidad bien entendida debiera afrontarse a partir de bancos de ítems, debidamente calibrados, que permitan la construcción de pruebas de naturaleza, estructura, composición y dificultad equiparadas; de este modo, los resultados de cada serie admiten una razonable comparación con los de series sucesivas para establecer el grado en que se da estancamiento, avances o retrocesos entre unas y otras[35]

A MODO DE SÍNTESIS FINAL

La evaluación es un instrumento poderoso al servicio de la calidad. En los países avanzados la evaluación está presente en los diversos modelos de gestión de la calidad así como en las actuaciones orientadas a la acreditación y certificación de bienes, productos y servicios.

Como no podía ser de otro modo, la evaluación es un poderoso instrumento al servicio de la calidad de la educación por sus aportaciones a la mejora de la realidad evaluada, sea el aprendizaje del alumnado, la actividad docente del profesorado, las organizaciones educativas, los medios y recursos disponibles, el ambiente de las instituciones, la participación del personal o cualquier otro aspecto de los que conforman la compleja realidad de los sistemas educativos modernos.

En contextos educativos, el sentido primigenio de la evaluación no puede ser otro que la mejora de las personas, algo ligado directamente a la función formativa de la evaluación, por lo general evaluaciones continuas llevadas a cabo por el personal de las organizaciones y, con frecuencia, a enfoques autoevaluativos, a juicio de quien escribe los de mayor potencial de mejora personal, grupal e institucional.

Sin embargo, los sistemas educativos tienen una evidente dimensión social, y ello no sólo en el sentido de que la persona integra las dimensiones individual y social y, por consiguiente, se orientan a la capacitación de las nuevas generaciones para que sean capaces de desenvolverse con éxito en la vida tanto en sus dimensiones profesionales y cívicas como de relaciones humanas, sino porque a su servicio se ponen ingentes medios -por más que siempre sean considerados insuficientes- y recursos cuya eficacia, eficiencia y efectividad es preciso promover y asegurar.

Es en este contexto donde cobra especial relieve e importancia la evaluación externa, en funciones sumativas, administrativas o sociopolíticas, en terminología de D. Nevo.

A mi juicio, con el fin de mantener la necesaria coherencia con la naturaleza de las actuaciones educadoras, conviene que tales evaluaciones, necesarias y potencialmente útiles, se "formativicen", esto es, se diseñen y se ejecuten de conformidad con ciertas exigencias técnicas y éticas que permitan extraer de ellas información valiosa para la mejora de las personas y de las organizaciones educativas.

Por otra parte, conviene dejar constancia del valor de los estándares de calidad de las evaluaciones, que ofrecen una pauta que debe ser respetada para que no se altere el sentido primigenio de la evaluación, pasando de ser un instrumento, herramienta o medio al servicio de la educación a convertirse en un fin en sí misma, prostituyendo así su sentido, alterando el currículo oficial de las organizaciones, modificando los objetivos de las materias, forzando las pautas evaluadoras del profesorado y hasta

estimulando en los estudiantes la cultura utilitarista de que lo importante no es saber sino aprobar.

Si se respetan estos principios, la evaluación, la cultura evaluativa, representará una aportación de extraordinario relieve a la calidad de la educación, desde el nivel del sistema educativo en su conjunto a la formación de cada una de las personas que se dedican a la crucial tarea educativa: educadores, padres y el propio alumnado.

[1]El término "evaluación" aparece recogido en el texto de la Ley en no menos de 120 ocasiones.

[2]Ley Orgánica 6/2001, de 21 de diciembre

[3]Un tasador a la persona que le encarga la tasación; un crítico literario, cinematográfico o teatral, al editor de la revista o medio de difusión; un juez, a la sociedad, a través de la publicación de la sentencia...

[4]Valga señalar, por el momento, la necesidad de utilizar técnicas de recogida de información que puedan ofrecer datos relevantes para tomar las decisiones no sólo sumativas sino, también, formativas. En efecto: una prueba en funciones estrictamente sumativas no tiene por qué ser representativa -validez de contenido- de la realidad evaluada, mientras que debe poseerla necesariamente si se pretende, aunque sólo sea indirectamente, esa "formativización".

[5]Tal es el caso de la función de creación de teoría, citada por la profesora R. Fernández Ballesteros. Vid. FERNÁNDEZ BALLESTEROS, R. (1996) *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis, pp. 30-32.

[6]Aunque este punto resulta especialmente adecuado para la autoevaluación de las organizaciones y programas educativos.

[7]En este sentido es de destacar que una de las características del sistema ISO es el de documentar los procesos, precisamente para evitar esta forma de proceder. Si bien es cierto que tal actividad puede ser laboriosa para los profesores y responsables de los centros educativos, no lo es menos su extraordinario valor y utilidad para afrontar con mayores posibilidades de éxito la problemática del día a día de los mismos. De esta manera se podrá disponer de conocimiento sólido sobre las causas de la violencia escolar, sobre la forma de abordar las faltas de disciplina, sobre el modo de motivar a los alumnos que repiten curso, sobre procedimientos para personalizar las relaciones humanas, sobre sistemas para lograr la participación de los padres...

[8]Con un ejemplo creo poder aclarar el concepto. Preguntar al alumnado, o a las familias, por la calidad de los materiales didácticos en general, puede conducir a una valoración intermedia, dado que unos pueden ser de gran calidad y otros presentar deficiencias. A partir de ahí, pocas decisiones podrán tomarse que redunden en la mejora. Sin embargo, si los consultados pueden referir sus valoraciones a los materiales de las asignaturas concretas que cursan, la información contará con la *valía* necesaria para poder mejorarlos

[9]Los diversos sistemas de indicadores no presentan entre si una convergencia notable en lo relativo a su concepción de calidad. Este hecho podría facilitar la búsqueda simultánea de convergencia tanto entre ellos como de ellos con los modelos de evaluación de centros educativos, haciendo así un gran servicio a la calidad. El lector puede encontrar algunos modelos de interés en: OCDE (1995): *Análisis del Panorama Educativo*. Los indicadores de la OCDE 1995. Madrid: Mundi-Prensa y OCDE.; OECD (1997) *Education at a Glance: OECD Indicators 1997*. Paris, OECD;

INCE (2000): Sistema Estatal de Indicadores de la Educación. Madrid: Ministerio de Educación, Cultura y Deporte (<http://www.ince.mec.es/>). La última versión, correspondiente a 2006, puede encontrarse en:
<http://www.institutodeevaluacion.mec.es/>

[10] Para la fijación del tamaño se cuenta con tablas adecuadas, además de fórmulas que permiten establecer el tamaño necesario. Contra lo que algunos puedan pensar, no es mejor acudir a todos los sujetos que forman la población que a una muestra con las debidas garantías (tamaño e imparcialidad).

[11] La deseable imparcialidad se logra acudiendo a procedimientos aleatorios.

[12] El lector interesado pueden encontrar una síntesis suficiente en el capítulo 16 del texto *Problemas y métodos de investigación en Educación personalizada*, volumen 5 del *Tratado de Educación personalizada*, publicado por Rialp bajo la dirección de García Hoz (1994). Un tratamiento más detallado y técnico puede encontrarse en tres artículos de J.M. Jornet y J.M. Suárez, recogidos en el monográfico de la revista **Bordón** dedicado al tema de la *Cuantificación y medida en educación*, (vol. 41, 2, 1989).

[13] Piénsese en una cuestión tan sencilla como la de establecer cuándo un grupo de alumnos domina la ortografía. Decidir la amplitud del texto, la cantidad de palabras con dificultad que se introducirán, la naturaleza de la dificultad de tales palabras, el contexto en que se introducirán tales palabras (lenguaje coloquial, técnico, literario) y, sobre todo, a partir de cuántas palabras -y de qué naturaleza- mal, o bien, escritas se evalúa como no dominio o dominio, son algunos de los aspectos a que nos estamos refiriendo.

[14] El INECSE, en la evaluación de la Educación Primaria (2003) utilizó la Teoría de Respuesta al Ítem (TRI). Para interpretar el rendimiento de cada alumno se crea una escala con media y desviación típica arbitrariamente establecidas. La escala que se utilizó tiene un rango de de 0 a 500, una media global de 250 y una desviación típica de 50. En ella se marcan, además de la media, los puntos de referencia que se apartan de la media en \pm una a cuatro desviaciones típicas: 50, 100, 150, 200, 250, 300, 350, 400, 450.

[15] En su extraordinario libro *La investigación científica*, editado por Ariel en 1979 (6ª edición, en el capítulo 14, dedicado al experimento, afirma (p. 849): "*Los procedimientos de contrastación o técnicas experimentales deben contrastarse o convalidarse `teóricamente y experimentalmente'; experimentalmente porque pueden no funcionar, y teóricamente porque, aunque funcionen, el funcionamiento puede deberse a alguna causa distinta de la presumida en el procedimiento. La contrastación teórica de una técnica experimental consiste en explicar, a base de teorías independientemente corroboradas, cómo funciona el procedimiento. Incluso una técnica que tenga éxito debe considerarse sospechosa mientras sus éxitos -y sus limitaciones- no se expliquen a base de leyes científicas*".

[16] Puede consultarse al respecto el trabajo de M. Rivas Navarro, Factores de eficacia escolar: una línea de investigación didáctica, publicado en un número monográfico de la revista *Bordón* sobre el tema general de *Trayectoria y perspectivas de la educación compensatoria*, número 264, septiembre - octubre de 1986, págs. 693-708.

[17] Esta forma de actuar desactiva los proyectos educativos de calidad y hace inoperantes las acciones previstas a su servicio.

[18] Piénsese en el contenido de determinados informes nacionales e internacionales y se podrá comprender que su contribución a la calidad no es precisamente muy elevada.

[19] Una experiencia común entre los profesores del antiguo COU era la conversión del programa de sus asignaturas en otro diferente consistente en la preparación para la prueba de selectividad.

[20] Nada digamos si lo que puede llegar a perder un profesor es su puesto de trabajo.

[21] Por esta línea van algunas decisiones que se están tomando en el marco del denominado Espacio Europeo de Educación Superior. Se trata de calibrar lo que es un crédito (unidad de medida) para evitar el caso frecuente de profesores que plantean asignaturas cuyo tamaño real (horas de estudio, de prácticas) supera con mucho el tamaño legal.

[22] Vale la pena señalar que la dimensión ética también debe darse en las evaluaciones internas, en especial en las orientadas al control, como son las evaluaciones sumativas del aprendizaje del alumnado: calificar con aprobado, notable o sobresaliente puede tener repercusiones posteriores serias para la obtención de una beca o para la selección para determinados puestos.

[23] En cierto estudio sobre la calidad de las universidades se valoraba como un criterio de calidad el número de mujeres profesoras en relación con el de varones profesores. Cuando menos, hay que reconocer lo discutible del criterio.

[24] Es muy significativa la queja que, sobre la seriedad en la aplicación de las pruebas, realizan las autoridades locales de Bradford (Inglaterra). Pueden consultarse en FITZ-GIBBON, C.T. (1996): *Monitoring Education. Indicators. Quality and Effectiveness*. Londres: Casell.

[25] Una anécdota al respecto. Me contaba un compañero de una gran universidad que, sin hacer nada, habían subido varios puestos en el ranking por el mero hecho de haber perdido varios miles de universitarios de un curso a otro. La razón era sencilla: *ratios* como alumnos por profesor, libros de la biblioteca por alumno, espacios deportivos por alumno, ... habían mejorado notablemente. Pero para ello no habían tenido que introducir mejora alguna; es más, hasta cabría pensar si había razones más allá de las demográficas que habían conducido a la pérdida del alumnado, en cuyo caso estaríamos ante el mundo al revés.

[26] En el marco de una obra dirigida y coordinada por mí (*Calidad en educación, calidad de la educación. Documentos para una concepción integral e integrada*, publicada en 2005 por la Asociación Española para la Calidad -AEC) se dedica su segunda parte a la *Deontología profesional de la certificación y la evaluación* (págs. 103-142. Allí se señalan como requisitos técnicos y profesionales de los evaluadores su *competencia*, su *honestidad y honradez*, su *imparcialidad* y su *integridad*.

[27] Este es uno de los estándares, a los que nos referiremos en seguida, englobados dentro del criterio de *utilidad*.

[28] *No necesariamente se trata del poder político, pero sí de personas que están al frente de organizaciones de carácter público, que pueden considerarse dueños de la información recogida por ser quienes pagan su costo, sin tener en cuenta que no son dueños de tal dinero sino sólo administradores del mismo.*

[29] *Estudio sistemático planificado, dirigido y realizado con el fin de ayudar a un grupo de clientes a juzgar y/o perfeccionar el valor y/o el mérito de algún objeto" (p. 67)*

[30] El Instituto de Evaluación es la nueva denominación establecida por la Ley Orgánica 2/2006, de 3 de mayo, de Educación (L.O.E.) para el antiguo INECSE (Instituto Nacional de Evaluación y Calidad del Sistema Educativo), en su artículo 142.

[31] Si se reconoce que la evaluación es un instrumento, se comprenderá que el tiempo necesario no deberá ir en detrimento de lo fundamental. Esto es, no debe ocurrir que por evaluar se detraiga el tiempo necesario para educar.

[32] El lector puede consultar CISCAR, C. y URÍA, M.E. (1986): *Organización escolar y acción directiva*, Madrid: Popular; o MUÑOZ, A. y ROMÁN, M. (1989): *Modelos de organización escolar*, Madrid: Cincel.

[33] De hecho, hasta en el marco de los *indicadores* resulta difícil de lograr. Así, en *Análisis del Panorama educativo. Los indicadores de la OCDE. 1995*, editado en Madrid por Mundi-Prensa, se lee lo siguiente: "*Comparabilidad: mejoras y problemas persistentes*: Como resultado de las mejoras en el instrumento e instrucciones de recogida de datos y el continuo diálogo entre la Secretaría de la OCDE y los suministradores nacionales de datos, estos son, de algún modo, comparables. (...) Sin embargo, sigue habiendo importantes problemas de comparabilidad..." (Pág. 11).

[34] Un ejemplo lo tenemos en nuestro país con la prueba de selectividad aplicada tras el Bachillerato para acceder a la Universidad, que no puede tener como base unas mismas materias debido a que se dan diferencias entre las que se estudian en los diversos cursos. En efecto, las materias que en unas Comunidades Autónomas se estudian en segundo curso de Bachillerato (sobre el que se estructura la citada prueba) en otras se cursan en primero, dándose el caso de que, en otras, se reparten entre primero y segundo.

[35] La existencia de tales *bancos de ítems* permite generar pruebas paralelas comparables sobre un mismo universo de medida.

REFERENCIAS Y BIBLIOGRAFÍA BÁSICA

AUSTIN, G. y REYNOLDS, D. (1.990): "Managing for Improved School effectiveness: an international survey", *School Organization*, 10, 2-3, 167-178.

CASTILLO ARREDONDO, S. (2003) *Evaluación de programas de intervención socioeducativa: Agentes y Ámbitos*. Madrid: Prentice Hall.

COHEN, E. y FRANCO, R. (1993) *Evaluación de Proyectos sociales*. Madrid: Siglo XXI

COMITÉ CONJUNTO DE ESTÁNDARES PARA LA EVALUACIÓN EDUCATIVA (1998) *Estándares para la evaluación de programas*. Bilbao: Ediciones Mensajero.

CONSEJO DE UNIVERSIDADES (1991) *La evaluación de las instituciones universitarias*, Madrid: Secretaría General del Consejo de Universidades

FERNÁNDEZ BALLESTERIOS, R. (Coord.) (1996) *Evaluación de Programas*. Madrid: Síntesis.

INSTITUTO NACIONAL DE CALIDAD Y EVALUACIÓN (2000) *Sistema estatal de indicadores de la educación*. Madrid: Ministerio de Educación, Cultura y Deporte.

INSTITUTO NACIONAL DE CALIDAD Y EVALUACIÓN (2003) *Evaluación de la Educación primaria*. Madrid: Ministerio de Educación y Ciencia.

JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1994) *The Program Evaluation Standards*. Beverly Hills: Sage Pub.

JORNET, J.M. y SUÁREZ, J.M. (1994) Evaluación referida al criterio. Construcción de un test criterial de clase. En V. GARCÍA HOZ, V. (Coord.) *Problemas y métodos de*

investigación en Educación personalizada, volumen 5 del *Tratado de Educación personalizada*, Madrid: Rialp.

JORNET, J.M. y SUÁREZ, J.M. (1989) Conceptualización del dominio educativo desde una perspectiva integradora en evaluación referida la criterio (ERC). En Bordón, Vol. 41.2, 217-236.

LUJÁN, y PUENTE, J. (1996) *Evaluación de Centros. El Plan EVA*, Madrid: MEC.

MARTIN, M^a T. y SARRATE, M (Coords.) (1999) *Evaluación y ámbitos emergentes en animación sociocultural*. Madrid: Sanz y Torres.

MINISTERIO DE EDUCACIÓN Y CULTURA (1997) *Modelo Europeo de Gestión de Calidad*, Madrid: MEC.

NEVO, D. (1.986) "The Conceptualization of Educational Evaluation" en E.R. HOUSE (Edit): *New Directions in Educational Evaluation*, Lewes: Falmer Press, 15-30.

NEVO, D. (1.990) "Role of the Evaluator", en H. J. WALBERG y G. D. HAERTEL (Ed.): *The International Encyclopedia of Educational Evaluation*, Oxford: Pergamon Press. (pp. 89-90).

OECD (1997) *Education at a Glance: OECD Indicators 1997*. Paris, OECD.

O.C.D.E. (1.991): *Escuelas y calidad de la enseñanza. Informe internacional*, Madrid, Paidós-M.E.C.

PÉREZ JUSTE, R. y GARCÍA RAMOS, J.M. (1989) *Diagnóstico, evaluación y toma de decisiones*. Volumen 9 del *Tratado de Educación personalizada*, Madrid: Rialp.

PÉREZ JUSTE, R. (Coord.) (2000) Evaluación de programas educativos. En Revista de Investigación Educativa. Número monográfico. Vol. 18, 2.

PÉREZ JUSTE, R. y otros (2004) *Hacia una educación de calidad. Gestión, instrumentos y evaluación*, (3^a edición), Madrid: Narcea

PÉREZ JUSTE, R. (Coord.) (2005) *Calidad en Educación, calidad de la Educación. Documentos para una concepción integral e integrada*. Madrid: Asociación Española par la Calidad (AEC)

PÉREZ JUSTE, R. (2006) *Evaluación de programas educativos*. Madrid: La Muralla.

RIVAS NAVARRO M. (1.986) Factores de eficacia escolar: una línea de investigación didáctica, Bordón, 38, (264), 693-709.

ROSSI, P.H. y FREEMAN, H.E. (1.989) *Evaluación: Un enfoque sistemático para programas sociales*, México: Trillas.

STUFFLEBEAM, D.L. y SHINKFIELD, J. (1987) *Evaluación sistemática. Guía teórica y práctica*. Buenos Aires: Paidós.

VELAZ DE MEDRANO, C. (1995) *Evaluación de programas y de centros educativos: diez años de investigación*. Madrid: Ministerio de Educación y Ciencia.

VILLA, A. y ALVAREZ, M. (Coords). (2.003) *Un modelo de calidad pedagógica para los centros educativos*, Bilbao: Mensajero.

VILLA, A. y ALVAREZ, M. (Coords). (2.003) *Ámbitos y criterios de la calidad pedagógica*, Bilbao: Mensajero.

WILSON, Jh.D. (1.992) *Cómo valorar la calidad de la enseñanza*, Madrid, Paidós-M.E.C.