

Tratamiento y Usos de la Información en Evaluación

Alejandro Tiana Ferrer

U.N.E.D. (España)
1997

Introducción

De unos años a esta parte se está extendiendo entre los administradores públicos el interés por la evaluación, siendo cada vez más quienes la consideran un valioso instrumento para la mejora cualitativa de los servicios que gestionan. Si hasta hace algún tiempo los conceptos utilizados en la administración pública procedían fundamentalmente de campos disciplinares como el derecho o la ciencia política, en los últimos años se han ido adoptando nuevas orientaciones a partir de las aportaciones realizadas por las diversas ciencias de la administración. En este nuevo contexto, la evaluación ocupa un lugar específico, proporcionando conceptos, enfoques y técnicas para la gestión de sistemas complejos, como el sanitario o el educativo.

En el ámbito de la educación, la evaluación ha ocupado tradicionalmente un lugar destacado, aunque adoptando generalmente otras denominaciones (examen, calificación, certificación) y centrándose en la valoración de los aprendizajes logrados por los alumnos. Desde este punto de vista puede afirmarse que su creciente utilización no ha constituido una novedad radical para los educadores. Sin embargo, sí resultan nuevos su uso sistemático y su aplicación a otras parcelas de la realidad educativa tales como la concepción y el desarrollo del currículo, el profesorado, los centros docentes o la propia administración. Así, la evaluación se encuentra actualmente en el centro del debate educativo, tras ampliar considerablemente su perspectiva y su campo de aplicación (Tiana, 1996).

Entre los símiles utilizados para referirse a las funciones que desempeña actualmente la evaluación en relación con los sistemas educativos, uno de los más extendidos y felices es el de la *conducción* (De Landsheere, 1994; Kogan, 1996; Michel, 1996). El uso de dicho término sugiere una imagen del funcionamiento de la administración educativa muy diferente a la que ha predominado hasta ahora. Frente a la regulación mediante normas y el control jerárquico del cumplimiento de las mismas, se pone el acento en la recepción y el tratamiento de una información actualizada, capaz de permitir intervenciones más ágiles y mejor adaptadas a la diversidad de situaciones específicas. En efecto, el conductor debe estar atento al tablero de instrumentos e indicadores para recibir e interpretar una multiplicidad de señales y reaccionar del modo más adecuado en cada situación. Del mismo modo, el administrador de un sistema

complejo debe recurrir a informaciones diversas y lo más ricas posibles para orientar su acción. Entre tales instrumentos de información, la evaluación ocupa un lugar privilegiado, ya que no sólo proporciona datos descriptivos, sino también interpretaciones y juicios valorativos. El recurso a instrumentos con denominaciones tan significativas como la de *indicadores* no hace sino reforzar dicho símil.

Pero, si la imagen de la evaluación como *conducción* resulta valiosa, también nos obliga a reflexionar acerca de sus implicaciones. Una de las principales tiene que ver con el tratamiento de la información recibida, con el fin de utilizarla del modo más fructífero posible. Al igual que le ocurre a quien está sentado a los mandos de cualquier móvil, el *conductor* de un sistema complejo está obligado a buscar y seleccionar datos relevantes entre toda la información disponible, a establecer prioridades entre ellos, a integrarlos en conjuntos significativos, a valorar las posibilidades de respuesta que se le ofrecen y finalmente a tomar decisiones y ponerlas en práctica, siguiendo ciclos sucesivos de reflexión y acción.

Este modelo de actuación supone la existencia de sistemas de información suficientemente sofisticados, procedentes de diversas fuentes y capaces de combinar diferentes perspectivas, así como el desarrollo de una serie de habilidades relacionadas con el tratamiento, el uso y la interpretación de la información por parte de quienes deben utilizarla para tomar cualquier tipo de decisiones. Sin embargo, dada la novedad relativa del recurso sistemático a la evaluación en educación, aún queda mucho por hacer en ese sentido. En este como en otros campos, las ideas ingenuas siguen predominando sobre los análisis rigurosos.

El propósito de este documento consiste precisamente en aportar algunas reflexiones acerca del tratamiento y los usos de la información procedente de la evaluación. En opinión de muchos autores, se trata de una cuestión tan compleja como fundamental con vistas a obtener los mejores frutos de la evaluación educativa.

Entre todos los aspectos abarcados por ese tema general, se han seleccionado en esta ocasión cuatro que presentan especial interés y relevancia. En primer lugar, el documento aborda de modo general la cuestión de la utilidad de la evaluación, deteniéndose sobre todo en las distintas funciones que ésta puede desempeñar. Este primer apartado tiene un carácter fundamentalmente conceptual y reflexivo, con la intención de ofrecer un marco global en el que poder integrar los tres siguientes. Como se hace a lo largo de todo el documento, los análisis realizados están basados en la consideración de diversos tipos de actividades de evaluación, adoptando una perspectiva lo más comprensiva posible.

En segundo lugar, el trabajo analiza cómo se utiliza la información procedente de la evaluación. Por una parte, considera el flujo de información como un proceso continuo, que debe ser abordado y planificado de manera global. Los momentos de diseño del proyecto de evaluación y las fases intermedias de su desarrollo son objeto de especial atención en dicho proceso. Por otra parte, estas consideraciones generales no pueden hacer olvidar la relevancia de la fase final de la evaluación, en la que la redacción del informe de síntesis cobra especial importancia. A esta cuestión se dedica un apartado específico.

En tercer lugar, el documento aborda el uso comparativo de la evaluación, uno de los más interesantes y discutidos de la misma. Si bien la comparación constituye parte de la propia esencia de la evaluación, no deja de plantear problemas diversos y de gran interés. Entre ellos destaca el de las condiciones de una comparación justa, asunto estrechamente ligado al debate actual acerca de lo que se ha dado en llamar el *valor añadido* de la educación, asuntos a los que se dedica cierta atención.

Por último, el trabajo se centra en la difusión de los resultados de la evaluación. Lejos de la idea que reduce la difusión a la entrega y eventual publicación del informe final, hoy en día se tiende a considerarla una actividad planificada, que exige la utilización de diversas estrategias y el recurso a diferentes medios de comunicación. El último epígrafe se centra precisamente en estas cuestiones.

1. La utilidad de la evaluación

Quizás no sea exagerado afirmar que la evaluación es una tarea cargada tanto de promesas como de eventuales frustraciones. Cuando se leen declaraciones entusiastas de algunos de sus más acérrimos defensores, uno estaría tentado de esperar de ella los más benéficos frutos. Pero al tropezar con no pocos comentarios escépticos o desengañados acerca de su escaso valor práctico, duda si la realidad a que unos y otros autores se refieren es la misma. Y es que la evaluación constituye una actividad compleja y ambivalente. Por una parte, constituye uno de los modelos paradigmáticos de la acción racional humana; por otra, es una actividad sometida a un riesgo continuo de desnaturalización.

Uno de los motivos principales de esa discrepancia de opiniones acerca de la utilidad real de la evaluación se encuentra precisamente en la diversidad de concepciones existentes sobre las funciones que cumple y los usos que de ella se hacen. Dicho de otro modo, las expectativas que uno deposita en la actividad evaluadora, el papel que le otorga y la perspectiva que adopta para llevar a cabo sus análisis condicionan el juicio de utilidad que finalmente emite. Por ese motivo, merece la pena comenzar este documento dedicando algún espacio a reflexionar sobre estas cuestiones, con objeto de identificar los motivos de tal discrepancia y clarificar así los elementos del debate.

1.1. Evaluación, toma de decisiones e información pública

El interés por conocer y controlar cómo se utiliza la información suministrada por la evaluación está presente en todos los evaluadores, de una u otra forma y con mayor o menor intensidad. En efecto, si siempre resulta difícil sustraerse a la valoración del impacto de la propia actividad profesional, aún lo es más cuando el objetivo último de ésta consiste precisamente en orientar la acción en alguna parcela de la realidad, a partir de su análisis detenido y riguroso. Por eso, aun cuando algunos modelos teóricos insistan más que otros en la vertiente utilitaria de la actividad evaluadora, ninguno llega a ignorarla por completo.

Una manifestación evidente de ese interés generalizado se encuentra en las “Normas para la evaluación de programas, proyectos y materiales educativos”, elaboradas hace ya más de quince años por doce organizaciones profesionales de Estados Unidos (Joint Committee on Standards for Educational Evaluation, 1981). Dicho documento, que recoge las opiniones más comúnmente aceptadas entre los evaluadores acerca de los criterios que deben regir su práctica profesional, propone treinta normas, agrupándolas en cuatro categorías básicas: utilidad, viabilidad, adecuación y precisión. Como puede apreciarse, la primera categoría enunciada es precisamente la de utilidad, poniendo así implícitamente de manifiesto el carácter central que la misma posee. Los ocho criterios agrupados bajo tal denominación, a algunos de los cuales se hará referencia más adelante, están encaminados a asegurar que la evaluación proporciona y facilita la información práctica que necesitan las audiencias interesadas en la misma.

El énfasis que los evaluadores ponen en la utilización de los resultados de su trabajo es, por otra parte, uno de los principales aspectos que distingue la evaluación de la investigación. Aunque el debate acerca de las diferencias que existen entre las dos actividades fue intenso en las primeras etapas del desarrollo histórico de la evaluación, hoy en día está ampliamente aceptada la distinción entre ambas. En opinión de algunos autores como Popham (1993), las diferencias se sitúan en tres ámbitos: la generalización de sus resultados, los valores que orientan la actividad y el propósito fundamental de la misma. Así, mientras que la investigación pretende obtener resultados con un alto grado de generalización, la evaluación se preocupa más por la adecuación de los mismos a la situación y características específicas de su objeto de estudio. Por otra parte, la investigación sitúa como valor central el logro de la verdad científica, mientras que la evaluación se preocupa por la estimación del mérito de una realidad concreta. Por fin, la investigación intenta llegar a conclusiones demostrables y verificables, mientras que el objetivo de la evaluación consiste en informar y orientar la toma de decisiones.

De acuerdo con este análisis, que con ligeras modificaciones es compartido por otros muchos autores, los evaluadores no pueden olvidar que su trabajo está encaminado a emitir juicios de valor acerca de una realidad determinada, basándose para ello en una información recogida y tratada de manera sistemática. Dichos juicios constituyen generalmente el soporte de una actuación

posterior, aunque la conexión entre ambas no sea lineal ni necesariamente inmediata. Esta vertiente valorativa, que la mayor parte de los autores aceptan como elemento esencial de un proceso evaluador y que muchos de ellos ligan con la toma de decisiones, sitúa en un lugar central la reflexión sobre la utilización de la evaluación.

Ahora bien, el hecho de aceptar que toda evaluación tiene una dimensión práctica y aplicada no implica que se deba concebir su utilidad de manera unívoca. De hecho, existen diversas posiciones a la hora de explicar cómo pueden y deben utilizarse los resultados de la evaluación, en concordancia con las diferentes funciones que a ésta se le asignan. Dicha asignación de funciones ha ido variando con el paso del tiempo, evolucionando en el sentido de una creciente complejidad y de una progresiva ampliación.

Así, la primera respuesta inmediata a la pregunta acerca de la utilidad de una evaluación suele ser que sirve ante todo para proporcionar información y elementos de juicio a quienes deben tomar decisiones. Esta idea está presente en muchas personas y también en no pocos políticos y administradores públicos, como se pone habitualmente de manifiesto en las palabras de muchos de ellos. De acuerdo con esta concepción, los evaluadores deberían proporcionar una información rigurosa, válida y fiable a los responsables de una determinada institución o programa de intervención, a partir de la cual éstos tomarían las decisiones más oportunas acerca de su mantenimiento, supresión o modificación, adoptarían medidas para su mejora o llevarían a cabo cualquiera otra acción que se considere conveniente.

Dicha respuesta se apoya en la idea de una *utilización instrumental* de los resultados de la evaluación, que fue la concepción predominante durante las primeras etapas de su desarrollo histórico. Como afirma W.J. Popham de manera elocuente:

“a comienzos de los setenta [...] existía la creencia dominante de que las evaluaciones educativas bien conducidas podían, y *debían*, constituir el elemento singular más importante para la toma de decisiones. Los evaluadores educativos que iniciaban algún proyecto de importancia soñaban con el momento en que los responsables políticos, después de consultar diligentemente su informe, tomarían decisiones esencialmente concordantes con los resultados de la evaluación” (Popham, 1993:5).

Aunque la mayoría de autores actuales no dudan en calificar de ingenua dicha concepción, continúa estando presente en las mentes de muchas personas. Sin duda, subyace en muchas de las críticas que denuncian la escasa incidencia de la evaluación en los procesos de toma de decisiones.

A pesar de la pervivencia de este tipo de ideas, actualmente se tiende a considerar que la relación existente entre evaluación y toma de decisiones es más compleja de lo que sugiere su concepción instrumental. Cualquier persona que se haya enfrentado con la experiencia directa de tener que tomar decisiones en un campo cualquiera de actuación sabe que los elementos sobre los que se apoya ese proceso son muy variados y de diverso orden. Por una parte, se suele contar con información directa o indirecta, procedente de diversas fuentes y con mayor o menor pertinencia y fiabilidad en cada caso. Por otra parte, se han de tener en cuenta ciertos condicionamientos, de carácter normativo, institucional o personal, que determinan el margen existente de maniobra. Se ha de tomar así mismo en consideración el contexto de relaciones interpersonales, anticipando las reacciones que puedan producirse. Por último, se han de establecer negociaciones multilaterales con el fin de lograr que todos los actores implicados se sientan copartícipes de la decisión adoptada. En suma, se trata de una tarea delicada en la que el conocimiento y la información juegan un papel relevante, pero también la capacidad de negociación y de movilización de voluntades, el sentido de la oportunidad e incluso la intuición.

Siendo la toma de decisiones una tarea compleja, no debe extrañar que la evaluación sea uno de los elementos que contribuyen a ella, pero en modo alguno el único. Frente a quienes creen, de manera ingenua, que la evaluación constituye el único soporte de la decisión, o al menos el principal, son muchos los autores que desde una perspectiva más desapasionada subrayan el carácter limitado de su contribución. El hecho de reconocer su importancia no implica mantener su carácter exclusivo.

Siguiendo esa línea argumental, Shadish, Cook y Leviton (1995:448-449) afirman que "la evaluación es un acto político en un contexto en el cual el poder, la ideología y los intereses son primordiales e influyen sobre las decisiones más que la información procedente de la evaluación" y que "incluso cuando los resultados de la evaluación y las decisiones de los gobiernos coinciden, los primeros sirven muchas veces para justificar decisiones tomadas a partir de otros criterios. Los ejemplos de un uso instrumental, inmediato y frecuente de la evaluación son todavía bastante raros". Por su parte, Popham (1993:5-6) afirma que "la mayor parte de las decisiones educativas de cierta importancia se toman en un entorno claramente político e interpersonal en el que la evidencia desempeña un papel secundario", además de que "las evaluaciones educativas raramente arrojan evidencias indiscutibles".

Aunque algunos puedan considerar que este tipo de afirmaciones encierra cierto derrotismo, e incluso cinismo, más bien hay que reconocer que responden a una actitud realista. Cuando un director de una escuela o instituto, un inspector, un responsable de un distrito escolar o un dirigente ministerial han de tomar una decisión complicada, es muy probable que recurran a cuantas orientaciones y criterios les aporten los estudios de evaluación de que dispongan, pero es bastante dudoso que se basen únicamente en ellos para realizar dicha tarea.

Incluso puede afirmarse que no actuarían correctamente si no tuviesen en cuenta otros elementos de información.

Debido a este tipo de objeciones, la concepción puramente instrumental de la evaluación ha ido debilitándose con el transcurso del tiempo, abriéndose paso otra que podríamos denominar *conceptual o iluminativa*. De acuerdo con la misma, la evaluación no tendría como función única o principal la de ofrecer elementos para la toma de decisiones, sino que también cumpliría un papel importante contribuyendo a mejorar el conocimiento de los procesos educativos y arrojando luz sobre los mismos. Su función sería, pues, la de *iluminar* la concepción que una determinada comunidad tiene de una cierta parcela de la realidad (educativa, en este caso), aportando información relevante y elementos de reflexión acerca de la misma.

De acuerdo con esta concepción que hemos denominado conceptual o iluminativa, la evaluación puede realizar varias aportaciones destacadas. En primer lugar, permite sensibilizar a una comunidad acerca de la existencia de determinados problemas o de las características que los mismos presentan en un momento dado. En segundo lugar, contribuye a movilizar la conciencia social en torno a ciertos asuntos de interés general. En tercer lugar, pone de manifiesto la ineficacia de ciertas prácticas o enfoques, por más extendidas y arraigadas que se encuentren. En cuarto lugar, permite arrojar luz acerca del impacto real de determinadas actuaciones o políticas, más allá de la propaganda difundida sobre las mismas (Rossi y Freeman, 1993).

El énfasis en la concepción iluminativa de la evaluación, frente a la meramente instrumental, lleva también asociada la superación de la idea de un uso a corto plazo y únicamente por parte de los responsables de los programas e instituciones educativas. En vez de pensar en un impacto inmediato y directo, la aceptación de la función iluminativa obliga a considerar los efectos diferidos e indirectos de la evaluación, en plazos de tiempo más largos que los habituales. Así mismo, como tendremos ocasión de ver más adelante, obliga a analizar el tipo de uso que de ella hacen las diversas audiencias interesadas en la misma.

Siguiendo esta línea de pensamiento, hoy en día se acepta cada vez más que la evaluación cumple una importante función de información pública acerca de los asuntos de interés general y de objetivación de los fenómenos sometidos a debate social. Su contribución en este sentido es múltiple. Por una parte, permite fomentar el funcionamiento realmente democrático de nuestras sociedades, aportando elementos y criterios a partir de los cuales sea posible una discusión abierta e informada acerca de los asuntos sometidos a evaluación. Por otra parte, contribuye a modificar los marcos cognitivos de los ciudadanos y a facilitar la comprensión individual y colectiva de los fenómenos analizados. Por último, no se puede olvidar que también aporta elementos para la toma de decisiones, aunque deban integrarse éstos en un conjunto más amplio de datos y criterios.

Así pues, si la evaluación tendió en sus primeras etapas evolutivas a buscar su justificación social en la utilidad instrumental que prometía, en la actualidad encuentra su principal legitimación en una combinación de sus dos funciones principales, instrumental e iluminativa. Es necesario señalar, además, que ambas concepciones no suelen considerarse ya como alternativas, sino más bien como complementarias.

Junto a estas dos concepciones básicas acerca del uso de la evaluación, algunos autores añaden una tercera. Es el caso de Rossi y Freeman (1993), quienes hablan de un tipo posible de utilización que denominan *persuasivo*. Su ámbito específico sería el de la política, donde la evaluación desempeñaría una función eminentemente argumentativa al servicio del discurso político, sea éste de gobierno o de oposición. Su misión sería similar a la que cumple la inclusión de citas de literatos relevantes en los discursos parlamentarios o en los mítines, por ejemplo. En su opinión, este tipo de uso desborda los estrictos límites de la evaluación y de su incidencia práctica y no constituye, en consecuencia, objeto de análisis específico por parte de los evaluadores.

En este mismo sentido, Nigel Norris nos ha expuesto un caso concreto en que la propia evaluación se insertaría en una estrategia de imagen con la finalidad de realzar la calidad de una determinada actividad. Para los promotores del programa educativo que utiliza como referencia, el hecho mismo de someterse a una evaluación externa enfatizaría la racionalidad subyacente a su concepción y su proceso de desarrollo, reforzando así su credibilidad. Su ejemplo es muy elocuente, ya que los responsables del programa sometido a evaluación invirtieron unos recursos apreciables en la elaboración de folletos atractivos tendentes a difundir su estrategia de evaluación, con el propósito de crear una imagen positiva basada en la existencia de una planificación y un control racionales (Norris, 1990:72).

Sin duda no sería difícil encontrar otros ejemplos semejantes a los sugeridos o expuestos por los autores citados, en los que el uso de la evaluación sobrepasaría ampliamente las funciones instrumental o iluminativa, para adentrarse en otra persuasiva, cuya finalidad sería la de generar credibilidad, movilizar voluntades, lograr motivación o influir en una audiencia determinada. Aun reconociendo el interés de analizar esta perspectiva, no nos extenderemos en este momento más en ella, puesto que habría que integrarla en un análisis más amplio acerca de los modos de actuación política, que desborda ampliamente el objetivo de estas páginas. Quizás deba señalarse que este tipo de uso es el que más fácilmente se desliza hacia diversos abusos, como algunos de los que acertadamente expone Miguel Angel Santos: convertir la evaluación en un elogio a quien la patrocina o la realiza; convertir la evaluación en un instrumento de dominación, control y opresión; o hacer públicas sólo aquellas partes del informe que tienen un carácter halagador (Santos, 1993).

Para terminar este epígrafe, podría recurrirse a una cita de Shadish, Cook y Leviton, quienes ilustran muy adecuadamente la progresiva expansión del concepto de *utilización* de la evaluación, hasta llegar a englobar una gran diversidad de nuevas ideas y prácticas:

“Hoy en día, el concepto de uso de la evaluación abarca su utilización (de manera independiente o conjuntamente con otros factores) para la toma de decisiones, la inclusión de referencias o citas suyas en los debates públicos, su empleo en cursos de formación continua de docentes y de formación inicial de futuros profesionales, o su utilización con objeto de reconceptualizar programas y problemas sociales. El principal medio a través del cual ejerce su influencia ya no es simplemente la presentación de un informe a quienes deben tomar decisiones [...] Los medios actualmente utilizados para la difusión del conocimiento incluyen no solamente los informes y resúmenes dirigidos a sus promotores o patrocinadores, sino también la elaboración de otros informes dirigidos a los profesionales, diversas presentaciones efectuadas a través de los medios de comunicación y conversaciones imprevistas en encuentros informales” (Shadish, Cook y Leviton, 1993:451).

1.2. Niveles de aplicación y de análisis, usos y audiencias de la evaluación

Hasta aquí hemos venido adoptando una perspectiva general para analizar los usos y las funciones de la evaluación. Sin embargo, no se puede olvidar que estamos hablando de un instrumento susceptible de aplicación a muy diversas realidades y capaz de responder a distintas finalidades. Por ceñirnos al ámbito educativo, sabemos que resulta posible evaluar, por ejemplo, un centro docente, un programa de intervención psicopedagógica, el trabajo de un profesor o una determinada parcela del sistema educativo, pero también que los planteamientos, las técnicas a utilizar y las posibles aplicaciones de cada una de esas actuaciones son muy distintas entre sí. Para abordar con mayor profundidad y rigor la cuestión de los usos de la evaluación, resulta imprescindible diferenciar los diversos niveles o planos en que ésta puede ser abordada.

En una primera aproximación, podemos distinguir tres *niveles de aplicación* de la actividad evaluadora, correspondientes a otros tantos ámbitos o planos en los cuales la evaluación puede básicamente efectuarse. El primero tiene por objeto la valoración de la actividad de los distintos *individuos* que participan en el proceso educativo. Es el nivel en que se ha utilizado tradicionalmente con mayor frecuencia, sobre todo en su vertiente de calificación y certificación de los resultados logrados por los estudiantes. Como sabemos, la evaluación de los alumnos constituye una parte importante del trabajo docente, aunque a veces haya recibido otras denominaciones. Pero los alumnos no son los únicos individuos objeto de la actividad evaluadora. Desde hace ya décadas, los

profesores de muchos países también han sido sometidos a algún tipo de valoración periódica, con consecuencias a menudo sobre su carrera profesional o incluso sobre su salario, y otro tanto podría decirse de los directores escolares o de ciertos cargos de responsabilidad administrativa. En términos generales, puede afirmarse que las evaluaciones desarrolladas en este nivel individual son las que cuentan con mayor experiencia y tradición, aunque quepan algunas matizaciones en función de los agentes concretos a los que hagamos referencia.

Un segundo nivel de aplicación corresponde a la valoración de las *instituciones* o entidades singulares que componen una determinada parcela del sistema educativo. El ejemplo más característico de este tipo de unidades es sin duda un centro docente, de cualquier carácter y nivel de que se trate. En una institución como ésta desarrollan su actividad numerosos individuos, enmarcados en unas estructuras organizativas y unos patrones determinados de funcionamiento, cuya evaluación exige desbordar el ámbito puramente individual. El resultado de la actuación de esas personas y de sus relaciones mutuas produce unos determinados efectos educativos. El estudio y la valoración de tales actuaciones, relaciones y efectos constituye uno de los ámbitos más prometedores de aplicación de la evaluación educativa actual.

Pero en este segundo nivel de aplicación que hemos denominado institucional también podrían incluirse otro tipo de entidades singulares, quizás no tan visibles como los centros docentes, aunque no por ello menos interesantes. Es el caso de lo que genéricamente se han denominado *programas* educativos, rótulo que abarca una gran diversidad de actividades de intervención, con dimensiones, características y cobertura muy variables. Si bien es cierto que su carácter institucional es de naturaleza diferente al de los centros docentes, no cabe duda de que comparten con ellos algunas características comunes. Desde el punto de vista que aquí nos interesa la principal es que se trata de unidades singulares que demandan atención específica. Al igual que en el caso de los centros, su evaluación requiere atender a los procesos internos de organización y funcionamiento desde una perspectiva supraindividual. En este ámbito también contamos con una larga experiencia, iniciada hace ya varias décadas en Estados Unidos y continuada posteriormente en otros países.

El tercer nivel de aplicación es el que aborda la valoración del conjunto de un *sistema educativo* o de alguna parcela significativa del mismo. Frente a los dos casos anteriores, en éste desaparece el énfasis en la singularidad, para adoptar una perspectiva global. Aunque la evaluación realizada en este nivel esté basada en información correspondiente a unidades singulares, la pretensión de globalidad que la guía determina el rango y selección de los datos (recurriendo frecuentemente a muestras, datos agregados, experiencias piloto o casos ejemplares), así como el tipo de tratamiento que de ellos se realiza. Es necesario insistir en que evaluar un sistema educativo (o una parcela del mismo) no significa necesariamente hacerlo con cada una de sus unidades constitutivas. La generalidad del nivel de aplicación condiciona la perspectiva global adoptada.

Así pues, la evaluación puede aplicarse o llevarse a cabo en tres niveles o planos, progresivamente más alejados del ámbito individual y con una perspectiva cada vez más amplia. Como puede fácilmente inferirse, aunque no nos extendamos por el momento en esta apreciación, el hecho de ubicarnos en uno u otro determina nuestro planteamiento general, nuestros enfoques metodológicos, la selección de datos que realicemos y el tratamiento que hagamos de los mismos.

Pero la identificación de estos tres ámbitos genéricos resulta útil no sólo para distinguir los ámbitos de aplicación de la actividad evaluadora, sino también para diferenciar los tipos de análisis que la evaluación permite realizar. Así, del mismo modo que hemos hablado de tres *niveles de aplicación* de la evaluación (individual, institucional y sistémico), de acuerdo con el tipo de realidades que se evalúan, se podrían distinguir otros tres niveles homólogos en los que ésta puede utilizarse para tomar decisiones o arrojar luz. Así, podríamos hablar de tres *niveles de análisis*, de acuerdo con el tipo de conclusiones que pueden extraerse de una determinada evaluación para los individuos (sean alumnos o profesores), las instituciones (centros docentes, programas) o el sistema educativo en su conjunto, e independientemente del ámbito concreto en que aquélla se efectuó. Como veremos, los niveles de aplicación y de análisis pueden o no coincidir, produciéndose una serie de posibilidades cruzadas que merece la pena considerar con cierto detenimiento.

Al entrecruzar los tres niveles de aplicación identificados más arriba con los correspondientes niveles de análisis se dibuja una tabla de doble entrada que nos permite explorar los usos más adecuados de la evaluación en cada situación concreta. Como es lógico, la utilización más fructífera es la que se basa en análisis realizados en el mismo nivel que el de aplicación. Así, la evaluación individualizada de un alumno o profesor es la más adecuada para valorar su actuación y adoptar decisiones acerca de su promoción, certificación o progreso. Otro tanto podría decirse sobre la evaluación singular de un centro docente o de un programa de intervención, cuando se trata de decidir acerca de cómo mejorarlos o de la conveniencia de clausurarlos, o sobre la evaluación del sistema educativo cuando se pretende valorar su estado y situación antes de adoptar medidas de reforma. Siguiendo mentalmente el esquema gráfico que antes se avanzaba, las cuadrículas situadas en la diagonal de esa tabla de doble entrada representarían las situaciones en que la evaluación resulta más adecuada y fructífera.

En cada una de esas tres situaciones óptimas (aplicación y análisis en el nivel individual, institucional o sistémico, respectivamente) se pueden identificar algunos usos específicos de la evaluación. En primer lugar, si nos situamos en el ámbito individual, la evaluación constituye un instrumento que puede ser utilizado con fines de *acreditación* y de *desarrollo*. En efecto, la evaluación realizada sobre individuos (sean alumnos, profesores, directores o administradores) con la intención de adoptar decisiones que les puedan afectar

de una u otra manera ha presentado siempre esta doble dimensión. Por una parte, ha estado encaminada a emitir juicios de valor conducentes a la atribución de grados académicos, a la promoción en sus estudios o carrera o a la certificación de haber superado un determinado escalón en su progreso académico o profesional; es lo que se denomina *acreditación*. Por otra parte, también ha incluido entre sus objetivos la elaboración de planes de desarrollo personal, entiéndanse éstos como actividades de recuperación, decisiones de reorientación vocacional o profesional, o acomodación de actividades a las características individuales de los que en ellas participan; es a lo que aquí se hace referencia con el término de *desarrollo*. En otras palabras, la evaluación en el ámbito individual ha tenido y sigue teniendo la doble vertiente, sumativa y formativa, que popularizara Scriven hace ya tres décadas.

En segundo lugar, en el ámbito institucional también se identifican esas dos vertientes, aunque utilicemos distintas denominaciones para referirnos a ellas. Por una parte, la evaluación permite un uso de carácter sumativo, esencialmente orientado hacia el exterior de la institución, que podríamos calificar de *rendición de cuentas* o de *supervisión*. Por otra parte, también ofrece un uso formativo, orientado más bien hacia el interior de la propia institución, que podría denominarse de *mejora y desarrollo*. El primero hace hincapié en la valoración de los logros o las carencias del centro docente o el programa de que se trate, con la intención de aumentar su transparencia y proporcionar información acerca de su situación. El segundo se centra sobre todo en los puntos débiles y en las posibilidades que existen para su mejora, con el propósito de elaborar planes de desarrollo. Lógicamente, ambas finalidades pueden combinarse en proporciones variables, sin que debamos hacer demasiado hincapié en su diferenciación. En la mayor parte de los casos, el propósito de rendición de cuentas y el de desarrollo van íntimamente unidos, sin que resulte fácil separarlos. Y otro tanto podría decirse, paralelamente, de la conexión existente entre cada tipo de uso y su mayor o menor apertura al exterior.

En tercer lugar, también pueden identificarse dos usos principales de la evaluación en el ámbito sistémico, aquí denominados de *conducción* y de *información pública*, respectivamente. El primero de ellos fue comentado en la introducción del documento y tiene que ver con el nuevo estilo de gestión pública que se extiende progresivamente en la administración educativa. Como allí se decía, se trata de una imagen sugerente que permite concebir e interpretar de manera distinta las tareas clásicas de la administración y la dirección. El segundo uso corresponde a la función iluminativa de la evaluación, aplicada a este ámbito específico. Como podrá apreciarse un poco más adelante, tiene que ver con la contribución que ésta puede realizar al debate público y al funcionamiento democrático de una sociedad.

Como se indicaba más arriba, los usos más adecuados y fructíferos de la evaluación son aquellos en que el nivel de aplicación coincide con el de análisis. No obstante, no pueden ni deben descartarse otros en los que ambos niveles no

sean idénticos. Pero estas traslaciones de nivel deben ser consideradas con atención, pues en ellas puede encontrarse el origen de parte de los posibles abusos de la evaluación. Por ejemplo, Miguel Angel Santos señala el riesgo de “elegir sesgadamente para la evaluación algunas parcelas o experiencias que favorezcan una realidad o una visión sobre la misma” (Santos, 1993:71) como uno de los riesgos reales de utilización incorrecta de la misma.

Entre los usos legítimos de un análisis realizado en un nivel diferente al de aplicación merece la pena destacar dos. El primero es aquel en que el análisis se realiza en un nivel de mayor generalidad que el de aplicación, con una finalidad de *ejemplificación o iluminación*. El segundo se caracteriza porque el análisis tiene lugar en un nivel más particularizado que el de aplicación, con una finalidad de *reflexión por inferencia*.

En el primer caso, ciertas evaluaciones llevadas a cabo sobre individuos o instituciones son utilizadas para ejemplificar determinados aspectos de una realidad institucional o sistémica más amplia que la evaluada. Así, por ejemplo, la valoración de los procesos y relaciones que tienen lugar en uno o más centros docentes puede dar pistas para interpretar fenómenos que se producen de una manera más general en un sistema educativo. Del mismo modo, el estudio de las reacciones de determinados individuos en un contexto institucional puede arrojar luz sobre las características de este último.

Como sabemos, dicho tipo de operaciones son aceptables desde el punto de vista científico, aunque el viejo problema de la inducción haya generado largas e interesantes polémicas a lo largo de la historia. En concreto, el estudio cualitativo de casos singulares con fines iluminativos cuenta con una ya notable tradición (Simons, 1980; Stake, 1995). Un reciente estudio de la OCDE sobre las innovaciones en el estudio de las ciencias, las matemáticas y la tecnología, basado en estudios de casos nacionales, nos ilustra convincentemente acerca de la potencia de este enfoque (Black y Atkin, 1996). La principal dificultad estriba en la identificación de la realidad concreta a estudiar y en el carácter ejemplar de la misma. Además, hay que tener en cuenta que esta traslación impone sus propias reglas de cautela en la generalización, que se deben respetar so riesgo de extraer conclusiones erróneas.

En el segundo caso mencionado, ciertas evaluaciones realizadas sobre instituciones o sistemas se utilizan para reflexionar y adoptar decisiones en un nivel individual o institucional más particularizado. Por ejemplo, la consideración de los resultados de una evaluación regional o nacional por los miembros de un centro docente, con la intención de reflexionar acerca de qué consecuencias deberían extraerse para el funcionamiento de éste, constituye una práctica habitual en lugares muy diversos. Cuando el Ministerio de Educación y Cultura de la República Argentina elaboró unos documentos dirigidos a los profesores acerca de cómo podría abordarse la enseñanza de las áreas con peores

resultados en los operativos nacionales de evaluación, no estaba sino haciendo este tipo de uso de la misma.

Como puede deducirse fácilmente, en todos los ámbitos analizados se combinan en una u otra proporción las funciones instrumental e iluminativa de la evaluación. Sin embargo, podría afirmarse que los niveles más particularizados (individual e institucional) son más aptos para un uso instrumental orientado a la toma de decisiones, mientras que en el más general (sistémico) predomina el uso iluminativo.

Llegados a este punto, conviene recordar que el uso que se realiza de una evaluación determinada no es un fenómeno abstracto, sino que se produce a través de personas o de grupos concretos. Es cierto que la evaluación es una tarea técnicamente compleja, que se lleva a cabo merced a la actuación de diversos especialistas. Pero ello no quiere decir que sea responsabilidad exclusiva de los técnicos ni que se desarrolle en el vacío. Su contexto es profundamente humano, puesto que cada evaluación concreta afecta e interesa a una diversidad de individuos y de colectivos, todos los cuales tienen algo que decir acerca de su enfoque, su proceso y sus resultados.

El reconocimiento de esta confluencia de intereses legítimos y la necesidad de tener en cuenta a todos los implicados e interesados en una actuación evaluadora concreta han llevado a la introducción del concepto de *audiencias* de una evaluación. Con esa denominación se hace referencia a todas las personas y grupos afectados o que tienen intereses legítimos en un proceso determinado de evaluación. En cada uno de los ámbitos antes abordados pueden identificarse sus diversas audiencias, que abarcan, en todo caso, a los individuos directamente concernidos por la evaluación, los que desarrollan su actividad en ese contexto objeto de análisis, los responsables de la institución concreta, los promotores del estudio y las personas ante quienes deben presentarse los resultados alcanzados. En ciertos casos, algunos de estos grupos pueden coincidir total o parcialmente, como sería el caso de una evaluación de un centro docente promovida por su dirección, consejo escolar o claustro de profesores, en que dichos grupos actúan simultáneamente como clientes y sujetos de la misma.

La introducción del concepto de audiencia significó un avance histórico importante, por cuanto sustituyó la relación bilateral entre evaluador y patrocinador por otra multilateral en la que los diversos agentes educativos ocupan un lugar específico y asumen su propia responsabilidad. Además, amplió la noción anterior de *destinatarios*, que podía entenderse reducida a los promotores y a los responsables de la actividad evaluada, incluyendo a las personas y grupos mencionados. Las preguntas, preocupaciones y necesidades de éstos se convirtieron en aspectos fundamentales en el proceso de evaluación, con el propósito de responder a las situaciones particulares, abandonando el dogmatismo metodológico (Norris, 1990). Algunos modelos de evaluación, como el elaborado por R. Stake con el nombre de “respondente”

("responsive evaluation"), hicieron especial hincapié en la atención a las necesidades de información de las audiencias naturales del programa evaluado, viniendo a llamar la atención sobre un aspecto anteriormente ignorado. De ahí se deriva su importancia, más que de la metodología concreta propuesta, objeto de revisión posterior.

Las "Normas para la evaluación de programas, proyectos y materiales educativos" a las que se hizo referencia al comienzo de este epígrafe constituyen una demostración palpable de la preocupación que se fue extendiendo entre los evaluadores por atender adecuadamente a sus diversas audiencias. Concretamente, las normas incluidas en la primera categoría, relativas a la utilidad de la evaluación, están "concebidas para asegurar que una evaluación proporcione las informaciones prácticas que necesita una audiencia determinada" (Stufflebeam y Shinkfield, 1987:27). En ellas se establecen principios tendentes a asegurar que la evaluación responde a las necesidades e intereses de las audiencias concretas, proporcionándoles una información suficientemente clara y fácilmente utilizable. En la actualidad, no es posible emprender una evaluación ignorando el papel que en ella juegan los implicados en la misma. La utilidad de una evaluación es en buena medida la que de ella extraigan sus mencionadas audiencias.

Algunas de estas audiencias se han revelado especialmente importantes a la hora de hablar de los usos de la evaluación. En el ámbito institucional, por citar un caso relevante, es necesario mencionar la utilidad que pueden extraer los profesores y los directores de los procesos de evaluación de los centros docentes. En los últimos tiempos se ha insistido repetidamente en los beneficios que las familias y la administración educativa pueden obtener de una valoración objetiva del funcionamiento y los resultados de los centros. Argumentos de ese tipo han conducido en algunos países a la publicación de listas de clasificación de centros, en una tendencia que tendremos ocasión de analizar con mayor detalle en un epígrafe posterior, pero sobre la que podemos adelantar serias reservas. Pero, en cambio, se ha analizado mucho menos la influencia que pueden ejercer este tipo de procesos sobre el trabajo de los profesores y los directores. Y cuando dichos análisis se han efectuado, han olvidado casi por completo su vertiente no estrictamente utilitaria, pese a la importancia que puede tener. Veamos algún ejemplo concreto de dichos usos no convencionales.

En relación con los docentes, por hablar de una categoría de las mencionadas, son muchos quienes sostienen que existe una interconexión entre el desarrollo de procesos de evaluación interna del centro y la implicación de los profesores en los asuntos escolares. Aunque dicha relación no es tan clara en el caso de la evaluación externa, la puesta en marcha de procesos colectivamente asumidos de evaluación institucional interna puede ofrecer una valiosa oportunidad para favorecer dicha implicación y aumentar simultáneamente la moral de los docentes, contribuyendo así a mejorar el ambiente o clima escolar. Más allá de

las decisiones que se adopten al final de dicho proceso, su propio inicio produce un efecto valioso para los componentes de la institución.

En el caso de los directores, así mismo aludidos, vale la pena señalar que autores como Glasman y Nevo, después de analizar detenidamente un buen conjunto de investigaciones empíricas, llegan a la conclusión de que podrían mejorar notablemente su práctica y su estilo de toma de decisiones mediante la utilización de los conceptos y los métodos de la evaluación. De acuerdo con tal sugerencia, los directores no deberían ser simples “consumidores ilustrados de la evaluación”, sino personas capaces de utilizar los enfoques y la metodología evaluadora en sus procesos de toma de decisión. Según su propia formulación, se trataría de formar directores “mentalizados acerca de la evaluación” (“evaluation-minded principals”), lo que no significa transformarles en evaluadores profesionales, sino desarrollar en ellos las actitudes, las habilidades y los conocimientos que les faciliten un uso productivo y enriquecedor de la evaluación (Glasman y Nevo, 1988).

En el ámbito sistémico, por poner otro ejemplo diferente, hay que insistir en la función iluminadora que desempeña la evaluación, que suele ser más importante que la estrictamente utilitaria. Son muchos los autores que han puesto en conexión este tipo de usos no convencionales de la evaluación con su contribución al funcionamiento democrático de una sociedad. Así, Escudero se refiere a esta utilidad en la perspectiva de un *modelo de servicio público*, tendente a la “animación de debates públicos, pluralistas y democráticos, que, informados por los resultados de la evaluación, activen no tanto procesos de toma de decisiones instantáneas cuanto acuerdos negociados deliberativamente” (Escudero, 1995:101). La exigencia de identificar y atender a las necesidades de las audiencias es considerada por Barry MacDonald un requisito inexcusable de una evaluación democrática, como pone certeramente de manifiesto en un interesante texto que puede servir para cerrar este apartado:

"Una evaluación democrática debería procurar servir a los valores universales de las sociedades democráticas, atendiendo a los dos elementos que creemos fundamentales en la democracia. Por una parte el ejercicio del voto, que debe ser un voto informado y, por lo tanto, la necesidad de asumir como valor el deber de informar a la ciudadanía de todos los asuntos que le competen. Y por otra parte, que la democracia se caracteriza por un sistema de equilibrios y controles para prevenir el uso ilícito del poder. La evaluación debe tener en cuenta todos los intereses que se puedan identificar, es decir, nadie puede comprar la evaluación ni determinar las preguntas porque si no sería un mero instrumento de aquellos que pueden comprar una evaluación. Tener en cuenta todos los intereses significa identificar las preocupaciones de los diferentes grupos respecto de la actividad que se está analizando, y para garantizar que sus intereses estén

representados todos han de recibir información de todos los resultados" (MacDonald, 1995:15-16).

2. La utilización de la información

Como hemos podido apreciar en las páginas anteriores, no existe un criterio único para valorar la utilidad de la evaluación, debiendo aceptar su diversidad de usos, en función de cuáles sean sus ámbitos de aplicación y de análisis, sus propósitos concretos y las audiencias a las que afecta. Si en algunos casos podemos considerar útiles a las evaluaciones que permiten tomar decisiones más o menos directamente, en otras ocasiones lo serán si ayudan a mejorar el ambiente de una institución, a explicitar sus principales puntos fuertes y débiles o si contribuyen a objetivar el debate público sobre educación. Por ese motivo, la emisión de juicios acerca de la utilidad de una evaluación determinada basándose simplemente en la valoración de las decisiones inmediatas que ésta haya originado constituye una simplificación excesiva y es fuente de posibles distorsiones y conclusiones erróneas. Por no poner sino un simple ejemplo, hay que reconocer que la tremenda influencia ejercida en el mundo educativo por el conocido Informe Coleman, incluso más allá del marco estadounidense, desbordó ampliamente su traducción en decisiones concretas, que fueron muy pocas.

La conciencia de que una evaluación puede producir efectos que sobrepasan su utilización directa en términos de toma de decisiones ha llevado a algunos autores a plantearse y estudiar la cuestión del *impacto* real que esta actividad es capaz de producir (Alkin, Daillak y White, 1979). Por impacto se entiende la influencia perceptible que la evaluación ejerce sobre las actividades o actitudes de individuos o grupos, lo que amplía considerablemente el concepto más estrecho de utilidad. Una de las principales implicaciones de tal ampliación es que permite distinguir entre los efectos esperados, deseados y voluntarios de una actividad evaluadora de los inesperados, no deseados e involuntarios. En efecto, no sería razonable considerar útil una evaluación solamente si alcanza sus propósitos declarados, dejando de lado otras consecuencias que haya podido producir.

En el apartado anterior se citaron varios ejemplos capaces de ilustrar adecuadamente este tipo de impacto no estrictamente utilitario. Uno de ellos es el efecto que la puesta en marcha de una evaluación puede producir en las actitudes de los participantes en la experiencia sometida a examen. Sabemos bien que la evaluación de un centro o programa puede producir reacciones que oscilan entre la ansiedad y la implicación entre los profesores, alumnos y responsables del mismo; también sabemos que el inicio de ese proceso puede constituir una fuente de motivación para la mejora o generar rechazo y estimular actitudes defensivas. Todos estos efectos no pueden ser ignorados si hablamos de la utilidad de una evaluación institucional. Otro de los ejemplos citados sería el efecto simbólico que la evaluación de una realidad puede producir en favor de

su legitimación. Más allá de los resultados finales del estudio, su simple puesta en marcha contribuye a forjar una imagen de racionalidad en la gestión de la institución, influyendo en su percepción social. Un caso no citado antes, pero muy interesante de analizar, es el efecto ejemplar que la aplicación de pruebas para la evaluación de un sistema educativo puede tener sobre la práctica didáctica y evaluadora de los docentes. El hecho de utilizar un determinado procedimiento, un enfoque dado y unas técnicas concretas para realizar una evaluación a la que se concede relevancia local, regional o nacional puede incitar a los profesores a adoptar procedimientos, enfoques y técnicas similares en su esfera de actuación, al margen de que sean o no los más adecuados para ese ámbito más específico. En algunos casos, el efecto es aún más evidente, como cuando se produce el fenómeno muchas veces mencionado de “enseñar para el test”, que llega a desvirtuar el carácter de la propia evaluación.

A la hora de llevar a cabo una evaluación es necesario prever todos estos efectos en la medida de lo posible y analizarlos cuando se produzcan, pues ello ayudará a un mejor desarrollo de la actividad. No obstante, dada su propia naturaleza inesperada y no deseada, es difícil anticiparlos completamente. Aquí no nos vamos a extender mucho más en su análisis, bastando por ahora con alertar acerca de su existencia y reconocer la necesidad de extremar la atención y la sensibilidad del evaluador para controlar el impacto no previsto de su actuación.

En este apartado nos centraremos más bien en un asunto que puede y debe ser previsto, planificado y controlado por el evaluador y que consiste en el uso que las distintas audiencias realizan de la información que se les aporta a lo largo de un estudio concreto. Este propósito se desdobra en dos vertientes: por una parte, se debe estudiar el proceso continuo de información que tiene lugar en una tarea de evaluación y los usos que de la misma se hacen, aspectos sobre los que existe una cierta confusión; por otra, se ha de analizar el uso específico que tiene lugar en la última fase del proceso, caracterizada por la emisión del informe final.

2.1. Un proceso continuo de información

La evaluación, por su propia naturaleza, tiene como finalidad ofrecer una información de carácter valorativo acerca de una determinada realidad. Esta afirmación podría hacer pensar que el evaluador se dedica en una primera fase de su trabajo solamente a recoger información, en una segunda fase a elaborarla y en una tercera a ofrecerla a las audiencias interesadas. Sin embargo, la realidad es más compleja, ya que dichas actividades están imbricadas entre sí.

En efecto, la idea de fases discontinuas falsea notablemente una realidad que se caracteriza más bien por la existencia de un *flujo continuo de información* que va definiendo progresivamente sus perfiles e integrándose en unidades de

significado cada vez más amplias, hasta permitir llegar a interpretaciones globales. La imagen más adecuada es la de un proceso en espiral, en el cual cada nuevo desarrollo permite alcanzar un plano más elevado de análisis e interpretación. De acuerdo con esta representación, la recogida y el suministro de información se integrarían en un proceso continuo, en el que las nuevas aportaciones de datos vendrían a completar, refutar o matizar las anteriores, hasta alcanzar una síntesis razonable y suficientemente comprensiva.

La existencia real de este proceso continuo es perceptible tanto desde el punto de vista del evaluador como de sus audiencias. Por una parte, el evaluador aumenta y enriquece progresivamente su conocimiento de la realidad estudiada, sin que se pueda afirmar que en determinado momento esté en situación de ignorancia absoluta, que en otro recoja la información necesaria y en un tercero la procese hasta emitir su informe final. Por otra parte, las audiencias de una evaluación determinada reciben de manera progresiva una información sistemática y valorativa, que viene a superponerse sobre la que ya poseían, modificándola parcialmente en síntesis sucesivas.

Fijémonos con algo más de detalle en el proceso que experimenta el evaluador. Cuando inicia una evaluación, suele poseer alguna información preliminar acerca de la realidad que ha de abordar, merced a un conocimiento superficial o profundo de su objeto de estudio. En alguna ocasión, no obstante, la primera información, generalmente de carácter no sistemático, es recibida al mismo tiempo que el encargo de la evaluación. En uno y otro caso, la información poseída o recibida no siempre está exenta de prejuicios o presupuestos. Haciendo un símil con otros ámbitos de la investigación, como puede ser la histórica, el evaluador posee lo que Topolsky (1985) llama un *conocimiento no basado en fuentes*. Ello quiere decir que, al igual que el investigador o el historiador, recurre a unos conocimientos iniciales para el planteamiento de su trabajo, que consisten básicamente en su experiencia general, los resultados de la investigación anterior y ciertos conocimientos transferidos de otras disciplinas.

Al iniciar su trabajo, el evaluador debe prever cuidadosamente la información que necesitará y organizar el plan general de recogida de la misma. Pero hay que tener en cuenta que sus ideas y percepciones iniciales pueden determinar, e incluso sesgar, las decisiones de evaluación que tome y el plan de recogida de datos que adopte. Por ese motivo, es necesaria una actitud de apertura hacia la recepción de información no prevista y un contraste del proyecto de trabajo con sus principales audiencias, con vistas a evitar los riesgos señalados.

Según avanza el proceso, debe ir sistematizando la información recogida y realizando interpretaciones y valoraciones parciales y preliminares. La confrontación de las mismas en el interior del equipo de evaluación y con los afectados por la misma permitirá su depuración y su correcta interpretación. Normalmente, este tipo de interpretaciones parciales se van avanzando bajo la

forma de informes de progreso, aunque también se llevan a cabo a través de comunicaciones e intercambios informales.

Por último, al final del proceso, el evaluador estará en condiciones de emitir un informe final, en el que queden recogidas las principales conclusiones y valoraciones alcanzadas. Dicho informe ha sido muchas veces considerado el elemento clave y el resultado más característico de la evaluación pero, como hemos visto, dista de ser su único resultado.

Desde el punto de vista de los interesados en la evaluación, también existe un flujo continuo de información, que se desarrolla paralelamente al que acaba de exponerse. Así, previamente al inicio del proceso, los implicados y participantes en la realidad evaluada suelen poseer una información abundante acerca de la misma, mayor sin duda que la que tienen los evaluadores, pero generalmente poco sistematizada e incluso con lagunas de cierta importancia. Además, está condicionada por un amplio rango de situaciones personales que determinan su percepción. En consecuencia, existe un riesgo real de prejuicios y tomas previas de posición.

En el momento en que se inicia el proceso de evaluación, los destinatarios de la misma reciben un proyecto de trabajo que determina el tipo y la cantidad de datos que se considera necesario recoger, establece ciertas prioridades entre los mismos e incluye el plan que ha de seguirse para su recogida. Dicho proyecto inicial suele traslucir la visión preliminar que los evaluadores poseen del objeto que abordan, permitiendo así un primer contraste con la que poseen otros grupos implicados. Aunque pueda parecer que esta primera información, aún poco completa y sistematizada, es irrelevante, puede sin embargo resultar muy valiosa para las diversas audiencias de una evaluación, ya que constituye un elemento de contraste con su propia percepción.

Según se va desarrollando el proceso de evaluación, sus destinatarios pueden ir recibiendo información de manera asistemática, además de informes de progreso, en los que aparecen versiones cada vez más elaboradas de las interpretaciones y valoraciones que los evaluadores construyen a partir de la información que van recogiendo. Dichos informes parciales cumplen al menos una doble función. En primer lugar, permiten conocer la selección de información que los evaluadores han realizado, así como las interpretaciones y las valoraciones que establecen tentativamente a partir de la misma. En segundo lugar, permiten contrastar los puntos de vista de unos y otros, enriqueciendo progresivamente el análisis. En consecuencia, dichos informes de progreso representan un elemento importante de información durante el desarrollo de la evaluación.

Al final del proceso, los destinatarios reciben un informe final, que sintetiza el trabajo realizado y constituye su resultado más característico. Aunque se hará una referencia más amplia a sus características más adelante, puede señalarse

aquí que se trata de un elemento fundamental de información, pero de ningún modo el único, como se puede ir apreciando.

A lo largo de este proceso continuo, es mucha la información que se recoge, se trata y se entrega mutuamente por parte de unos y otros. No todos estos trasiegos de información tienen un carácter sistemático, pero ello no quiere decir que no sean valiosos y relevantes. Una parte de dicha información es transmitida de modo informal, a través de los intercambios personales que abundan en toda evaluación. Pero otra parte, la más interesante de analizar puesto que tiene carácter intencional y sistemático, es recogida y transmitida formalmente. Los evaluadores recogen la información formal que necesitan por medio de diversos instrumentos de evaluación (entrevistas, cuestionarios, pruebas, observación,...) y transmiten la que han ido elaborando a través de tres medios principales: el proyecto de evaluación, los informes de progreso y el informe final.

Cada uno de estos tres instrumentos tiene su importancia particular. En las páginas que siguen se hará referencia más extensa a los dos últimos. Pero aquí debe hacerse una mención especial al primero, cuya importancia no se subrayará nunca demasiado. El proyecto de una evaluación constituye una concreción de los propósitos que la guían, los supuestos que la inspiran, los principios metodológicos que la rigen y los valores que la orientan. Por lo tanto, la discusión de dicho proyecto con los destinatarios de la evaluación y los afectados por la misma es un requisito indispensable para asegurar el carácter democrático de la operación. Obviamente, confrontar y contrastar el proyecto de evaluación con sus correspondientes audiencias no tiene el mismo significado en diversas iniciativas ni en ámbitos diferentes. En el caso de la evaluación de un centro escolar o un programa educativo, la discusión del proyecto habrá de realizarse directamente con los implicados en la misma, mientras que en el caso de una evaluación del sistema educativo los cauces de discusión serán más formales y representativos (órganos responsables del sistema educativo, comisiones parlamentarias, órganos diversos de participación social en educación).

Además de lo que ya se ha indicado en los párrafos anteriores, es importante destacar que el propio proyecto debe recoger la preocupación manifestada por los evaluadores acerca de cómo utilizar la información suministrada por la evaluación, sentando explícitamente las bases de su correcto tratamiento y uso. Ello quiere decir que los evaluadores no deben esperar simplemente a que se produzca de manera espontánea algún tipo de uso de la información obtenida, sino que deben formularse las cuestiones adecuadas y prever las mejores estrategias para conseguirlo.

Las consideraciones anteriores acerca de la necesidad de prever momentos e instrumentos concretos (como los informes mencionados) que permitan poner en común, discutir y contrastar la información disponible se insertan en una perspectiva más amplia, que enfatiza el papel central que la *negociación* ocupa

en todo proceso de evaluación. Aunque no nos extendamos mucho más aquí en este asunto, merece la pena recoger una frases de Joan Mestres que sitúan esta cuestión en su justo lugar:

"La negociación es una dimensión inherente a todo proceso de evaluación de cualquier tipo y modelo en sus distintos momentos: concepción previa, establecimiento de una política de evaluación, recogida de datos, tratamiento e interpretación de los mismos, elaboración de informes, divulgación y atención a las repercusiones y efectos del proceso evaluador [...] ¿Qué se puede negociar en evaluación educativa? ¿Se puede negociar todo?, ¿existen algunos límites en la negociación evaluativa? Excepto los principios éticos, probablemente todo lo demás pueda y deba ser negociado: objetivos, procesos, instrumentos, estrategias, tiempos,... [...] la meta de la negociación en evaluación estriba en conseguir el punto de equilibrio a través del diálogo, normalmente sin intermediarios, sino directamente entre las partes, mostrando sus intereses y valoraciones personales, es decir, que predomina el estilo de negociación orientada a la mutua satisfacción" (Mestres, 1995:37-40).

2.2. El informe de evaluación

Aun aceptando que la información proporcionada por la evaluación, a través de canales tanto formales como informales, puede y debe utilizarse a lo largo de todo el proceso de trabajo, merece la pena prestar una atención especial a las características y los problemas que implica la elaboración de los informes de resultados, ya que constituyen un elemento fundamental de dicho proceso. En efecto, la elaboración del informe representa la concreción de la tarea central de la evaluación, consistente en transformar los datos en evidencias (Sáez Brezmes, 1995). Incluso si algunos autores llegan a destacar que los informes de progreso suelen ser más útiles que el final (Escudero, 1995), no es posible ignorar la importancia real y simbólica de este último, lo que aconseja analizarlo con cierto detenimiento.

Una primera observación que conviene subrayar es que la elaboración del informe de evaluación tiene un carácter dinámico. Como indica Escudero, se trata de "un elemento resultante y dependiente de otras series de procesos y decisiones más trascendentales" (Escudero, 1995:96). En efecto, lejos de constituir un producto exclusivo de la última fase de trabajo, el informe comienza a perfilarse desde las etapas iniciales de la evaluación y se va construyendo y enriqueciendo progresivamente. Ya desde el momento de la elaboración de su proyecto de estudio, el evaluador debe comenzar a pergeñar el modelo de informe a presentar, haciéndose una representación mental de sus principales características. La consulta de informes de calidad contrastada, procedentes de otros estudios de evaluación, constituye una práctica aconsejable a la hora de realizar dicha tarea. Hay quien incluso defiende la necesidad de elaborar un

informe simulado, en el que queden explícitos el conjunto de información a incluir, su estilo de presentación, el tipo de análisis a realizar y el tono de las conclusiones y eventuales recomendaciones. La elaboración de este modelo permite concretar los propósitos que persigue la evaluación y contrastar con los interesados la utilidad de la información que se pretende ofrecer. Este modo de actuación resulta muy recomendable para los evaluadores que se encuentran en el comienzo de su carrera profesional, ya que facilita su trabajo posterior y evita peligrosos malentendidos iniciales.

A partir de esas primeras previsiones sobre el informe final, e incluso después de haber diseñado un modelo para su elaboración, el evaluador se enfrenta a la tarea de presentar diversos informes de progreso. Como se indicó anteriormente, la función de éstos es muy importante, puesto que permiten avanzar hacia la fase final por aproximaciones sucesivas y, al mismo tiempo, contrastar con los destinatarios y los usuarios de la evaluación la información reunida y las interpretaciones realizadas. Esta construcción progresiva del informe a través de avances sucesivos es especialmente importante cuando se utilizan determinados métodos cualitativos, en los que dicha elaboración no es resultado sólo del trabajo de los especialistas, sino fruto de un proceso compartido de construcción. Como puede lógicamente inferirse, la metodología empleada determina en buena medida las características y el proceso de elaboración del informe. Nuevamente conviene insistir en que la negociación, de la que antes se hablaba y que constituye la esencia del proceso de evaluación, se manifiesta con especial nitidez al tratar este aspecto del proceso.

Pese a la importancia que los evaluadores conceden a los informes de resultados, todavía son muchas las voces que se elevan para denunciar lo difícilmente comprensibles que a veces pueden resultar para los no especialistas y el riesgo en que llegan a caer de permitir malas interpretaciones. Sin duda este tipo de deficiencias, que muchos suscribirían, dificulta la extensión de la práctica evaluadora y su utilización real y fructífera por parte de sus destinatarios.

Estos problemas están en buena medida provocados por no diferenciar adecuadamente las características de un informe de evaluación de otro de investigación. En efecto, en muchas ocasiones el informe de evaluación ofrece un aspecto, unos contenidos y un formato similares a una memoria de investigación, que dificultan su uso por parte de aquellos destinatarios que carezcan de formación específica para su interpretación y manejo. No se trata tanto en estos casos de que el informe sea incorrecto, como de que resulta inadecuado. Aun cuando sus contenidos sean irreprochables desde el punto de vista técnico, muchas veces no se adaptan a las características, formación y necesidades de sus destinatarios. Así, por ejemplo, no es extraño que un informe de evaluación adopte el esquema incluido en el Cuadro 1, en el que la preocupación por el rigor técnico predomina sobre las necesidades de la comunicación.

CUADRO 1. *Esquema de un informe técnico de evaluación* (Glasman y Nevo, 1988:53)

1. Resumen
2. El contexto de la evaluación
 - a. Descripción de la actividad a evaluar y las razones de la evaluación
 - b. El papel de la evaluación y sus clientes
3. El marco conceptual de la evaluación
4. Las preguntas de la evaluación
 - a. Preguntas principales y secundarias
 - b. Justificación de las preguntas
5. Métodos de investigación
 - a. Instrumentos de medida y procedimientos de recogida de datos
 - b. Diseño de investigación
 - c. Procedimientos de muestreo
 - d. Procedimientos de análisis de datos
 - e. Fases de trabajo
 - f. Proceso de información sobre los resultados alcanzados
 - g. Limitaciones metodológicas
6. Resultados
 - a. Resultados acerca de cada una de las preguntas de la evaluación
 - b. Resultados adicionales
7. Resumen y conclusiones
 - a. Resumen de respuestas a las preguntas principales de la evaluación
 - b. Discusión de la significación de los resultados y sus limitaciones
 - c. Conclusiones y recomendaciones
8. Bibliografía
9. Apéndices

En el origen de este fenómeno de mimetismo hacia los informes de investigación se encuentra sin duda la formación recibida por muchos evaluadores y sus tradiciones de trabajo. Siendo académica y universitaria la procedencia de la mayor parte de ellos, no es extraño que recurran frecuentemente al modelo de una memoria de investigación a la hora de elaborar un informe de evaluación. Razones de formación, de hábito de trabajo, pero también de seguridad en el desarrollo de su actividad, pesan poderosamente en esta práctica. Y sin embargo, en ella se encuentra la causa de los problemas mencionados.

Para evitar este riesgo, algunos autores han defendido la aplicación de un modelo diferente, en el que las necesidades de una comunicación eficaz primen sobre la demostración del dominio técnico. El Cuadro 2 recoge un esquema diferente de informe al que antes se presentaba, con una organización más original y mejor adaptada a las necesidades de quienes utilizarán los resultados de la evaluación para tomar decisiones, del tipo que éstas sean. Obviamente, cada uno de ambos modelos tiene sus ventajas e inconvenientes, que el evaluador debe sopesar cuidadosamente. Mientras que el primero es un buen modelo de informe técnico, dirigido a especialistas, el segundo resulta ser más útil para audiencias no especializadas.

CUADRO 2. *Esquema de un informe de evaluación orientado a la toma de decisiones* (Glasman y Nevo, 1988:54)

1. Recomendaciones principales
2. Principales resultados
3. Resultados detallados
 - a. Resultados acerca de cada una de las preguntas de la evaluación
 - b. Resultados adicionales
4. Recomendaciones detalladas
 - a. Recomendaciones para la toma de decisiones específicas
 - b. Recomendaciones específicas acerca de las diversas partes de la actividad evaluada
5. Diseño de la evaluación y desarrollo del trabajo
 - a. Marco conceptual de la evaluación
 - b. Preguntas de la evaluación y su justificación
 - c. Instrumentos de medida y procedimientos de recogida de datos
 - d. Diseño de investigación y procedimientos de muestreo
 - e. Procedimientos de análisis de datos
 - f. Fases de trabajo
 - g. Limitaciones metodológicas
6. Bibliografía
7. Apéndices

Una vez expresadas estas cautelas acerca del esquema de informe que resulta más conveniente en cada circunstancia concreta, vale la pena detenerse a analizar cuáles serían sus características más deseables. Hay que comenzar diciendo que, en éste como en otros asuntos, no existe un acuerdo absoluto acerca de cuáles sean tales características. Mientras que el consenso es

generalizado en torno a ciertos extremos, también existen divergencias acerca de otros.

Una primera característica del informe en que todos coinciden es en la necesidad de incluir en él *toda la información que precisen sus destinatarios potenciales o reales*. Ello implica realizar una tarea previa de identificación de las audiencias, que se lleva a cabo en las primeras fases de la evaluación. En el caso de la educación, es habitual encontrar diversos grupos interesados en una evaluación. Aunque pueden variar, dependiendo de cada caso concreto, las familias de los alumnos, los profesores, los equipos directivos de los centros escolares y las autoridades y administradores de la educación suelen ser destinatarios privilegiados de los informes producidos. Si bien todos ellos pueden estar implicados e interesados en la evaluación, suelen diferir sus expectativas acerca de la información que esperan recibir.

Así, por ejemplo, los padres demandan una información clara y rigurosa, pero presentada en términos comprensibles y generalmente individualizada, demostrando interés en conocer con cierto detalle cuál es la situación de sus hijos, aunque sea en un contexto más amplio. Los profesores, por su parte, esperan recibir una información que les permita reflexionar acerca de su actividad profesional (sin que se perciba como una amenaza personal) y que a ser posible pueda orientarles en su trabajo cotidiano, ofreciendo suficientes garantías de rigor, con un cierto grado de elaboración técnica, pero sin sofisticación innecesaria ni abuso de la jerga profesional. Los equipos directivos de los centros docentes requieren una información que cumpla requisitos similares a los que piden los profesores, pero referida a las características de conjunto y los rasgos diferenciales de su institución, sin excluir posibles comparaciones internas y externas. En lo que respecta a las autoridades y administradores, demandan análisis de un mayor nivel de generalidad, pero capaz de iluminar los efectos de las políticas y las decisiones adoptadas. En resumen, como puede verse, las características y contenidos de la información demandada por las distintas audiencias coinciden sólo parcialmente. Ello ha llevado a algunos evaluadores a defender la necesidad de redactar distintos informes para sus diferentes destinatarios, cuestión debatida y sobre la que se volverá más adelante.

Una segunda característica deseable estrechamente asociada a la anterior es la *pertinencia y relevancia* del informe. Si una evaluación se pone en marcha buscando unas respuestas lo más concretas posibles a una serie de cuestiones más o menos explícitas, es responsabilidad del evaluador abordar dichos asuntos en su informe, sin derivar hacia aspectos irrelevantes.

También existe un acuerdo bastante generalizado acerca de algunas características formales que debe cumplir el informe. Así, todos los autores que se han ocupado de este asunto insisten en su *brevedad y concisión*. Sus destinatarios deben recibir tanta información como pueda obtenerse, pero con un

grado de detalle que no sea superior al que realmente puedan utilizar. Ello implica realizar una selección razonable de datos relevantes, incluyendo en anexos aquellos otros que puedan ser de interés pero no resulten fundamentales para el análisis o que constituyan la base de tratamientos posteriores. También coinciden en la utilización de un *lenguaje accesible*, sin tecnicismos innecesarios ni jerga profesional pero preciso y riguroso, y de un *estilo atrayente*, capaz de mantener la atención y el interés de sus lectores y destinatarios. En este sentido, el formato narrativo suele resultar más motivador que otro secamente analítico. No son pocos quienes consideran que ambos estilos han de estar sabiamente mezclados. El informe debe así mismo incluir las *referencias metodológicas* necesarias para su correcta valoración e interpretación, aunque sin convertirlo en un documento únicamente para uso de especialistas. Por último, conviene utilizar una *diversidad de recursos de comunicación*. Frente al informe exclusivamente escrito, con texto corrido, la introducción de gráficos, tablas, esquemas o ilustraciones favorece su asimilación. No se trata, obviamente, de “adornarlo” innecesariamente sino de utilizar distintos instrumentos de comunicación con el fin de facilitar a sus lectores potenciales su comprensión y asimilación.

No existe, sin embargo, tanto acuerdo en torno a otras características del informe de evaluación. Uno de los principales asuntos sometidos a debate es la conveniencia o no de incluir *recomendaciones*. Mientras que algunos evaluadores defienden la conveniencia de utilizar un esquema como el incluido en el Cuadro 2, que incluye las recomendaciones como un elemento fundamental, otros creen que esa es una tarea que corresponde a los destinatarios del informe. Para estos últimos, lo más que puede hacer el evaluador es presentar sus conclusiones, dejando a los receptores del informe la decisión acerca de cómo actuar a continuación. En esta concepción subyace una clara distinción entre conclusiones y recomendaciones. Aunque las segundas deriven de las primeras, suponen una primera toma de decisiones (o una sugerencia explícita para la misma) que desbordaría la función del evaluador.

Dada la vertiente práctica de toda evaluación, que no pretende sólo conocer sino también orientar la acción, la posición más extendida consiste en aceptar la inclusión de recomendaciones o, al menos, defender un modelo de informe orientado hacia su extracción. La dificultad mayor que esto plantea es la de conseguir plantear recomendaciones razonables, apoyadas en los análisis realizados y, sobre todo, viables en la situación concreta en que han de aplicarse. Sin duda, la madurez profesional de un evaluador se demuestra en buena medida en cómo sea capaz de afrontar esta tarea.

Otra cuestión sobre la que no existe acuerdo unánime, y que está en parte relacionada con la que se acaba de tratar, se refiere al *carácter expositivo o interpretativo* del informe. Mientras que autores como Scriven son claramente partidarios de adoptar el segundo tipo de enfoque, incluyendo explícitamente

una interpretación final, otros prefieren el primer enfoque, argumentando que violenta menos la realidad y no invade el campo de decisión propio de los destinatarios. Al igual que ocurría con las recomendaciones, la posición mayoritaria es favorable a la interpretación, aunque adoptando todas las cautelas que resulten necesarias.

Aun subrayando el papel central que el informe final desempeña en un proceso de evaluación, como se hacía al comienzo de este apartado, no pueden dejar de señalarse algunas de sus limitaciones. En primer lugar, hay que recordar algo que se mencionaba en páginas anteriores, en lo que nunca se insistirá bastante. Frente a la creencia bastante extendida de que un informe de evaluación debe traducirse inmediatamente en medidas concretas, sabemos que los elementos de información utilizados en un proceso de toma de decisiones son diversos y variables. Ello no implica una debilidad de la evaluación, sino tan sólo situarla en su justo lugar, sin fomentar espejismos acerca de su impacto e influencia. Como sabemos, es extremadamente raro que un informe de evaluación produzca por sí solo decisiones inmediatas y racionales; si lo consigue, será más bien a través de diversos canales de intermediación.

Otra limitación importante de los informes de evaluación consiste en la dificultad de recoger rigurosamente todos los perfiles de una determinada realidad, como pretende hacer. Hasta aquí se ha venido hablando de atender a una diversidad de audiencias con expectativas y necesidades diferentes, de interpretar y valorar la realidad evaluada sin distorsionarla, de ofrecer síntesis ajustadas pero asequibles. No es fácil para un equipo de evaluación atender a todas estas exigencias, corriendo siempre un riesgo de escorarse hacia algún lado determinado. Por ese motivo, hay quien ha defendido la elaboración de informes “adversarios” por equipos diferentes y a partir de los mismos datos (Popham, 1993). Esta técnica consiste en invitar a dos equipos o personas a interpretar de manera opuesta los mismos datos, permitiendo así a los destinatarios de la evaluación alcanzar sus propias conclusiones con el mínimo condicionamiento posible. Sería algo así como trasponer la práctica eclesial del “abogado del diablo” al ámbito de la evaluación.

La última dificultad que merece la pena destacar consiste en la dificultad de responder adecuadamente a las necesidades a veces tan diferentes de las diversas audiencias. No se trata solamente de que demanden o precisen información de diverso tipo, sino también que sus expectativas y propósitos no son necesariamente idénticos. En consecuencia, son muchos los evaluadores que siguen o defienden la costumbre de elaborar informes distintos para sus diversos destinatarios.

Esta es una cuestión polémica, ya que tiene diversas implicaciones. Por una parte, la existencia de informes diferenciados puede jugar en contra de la credibilidad de la evaluación, al extender la impresión de que no se dice lo mismo a todos los grupos afectados o interesados. Incluso podría argumentarse

que una práctica democrática de la evaluación exige dar a todos toda la información disponible. Por otra parte, la diversidad de informes puede facilitar una mejor adaptación a las diversas audiencias y una respuesta más adecuada a sus necesidades. Desde este punto de vista, no se trataría de esconder información a nadie sino de seleccionar la más relevante en cada caso. Incluso puede llegar a argumentarse que la difusión de informes técnicamente sofisticados entre grupos sin formación suficiente puede ser un modo de esconder los resultados, al hacerlos inasequibles.

Ante la complejidad de esta cuestión, merece la pena destacar realizar varias observaciones. Desde el punto de vista de la lógica democrática de la evaluación, no hay por qué aceptar que los informes deben ser únicos, salvo que se entienda este término de manera generosa. El hecho de que existan diversos informes adaptados a las distintas audiencias no supone *a priori* un déficit democrático. El asunto central radica más bien en la accesibilidad de los diversos informes producidos. O sea, que si un ciudadano puede consultar un informe, aunque no sea el que se preparó específicamente para el colectivo al que pertenece, no debe hablarse de falta de transparencia. Ello supondría hacer públicos varios informes, con distinto grado de extensión, profundidad y complejidad, aunque su distribución pueda realizarse de manera diferenciada.

Obviamente, el hecho de seleccionar la información incluida en cada uno de ellos no implica aceptar su distorsión. Por ese motivo, es necesario que el equipo evaluador realice los diversos informes, de modo que se eviten sesgos interesados o no. La responsabilidad única de los evaluadores en la elaboración de los distintos informes asegurará el respeto a criterios comunes. Ello no impide que en dicha tarea puedan participar otras personas (periodistas, expertos en comunicación, etc.), con el fin de adaptar el mensaje a sus destinatarios. Pero se trata de una tarea sumamente importante para dejarla en manos de otros.

Por último, hay que señalar que la experiencia de muchos países avala este tipo de aproximación diversificada a la elaboración de informes. Ya se mencionó anteriormente que los operativos nacionales de evaluación de la República Argentina han producido informes de carácter general, destinados a los responsables ministeriales y a un público más general, junto a otros específicamente orientados a los profesores de las materias y los niveles evaluados. Este tipo de informe resulta de indudable interés para una audiencia muy señalada, proporcionándole elementos de reflexión para la mejora de su tarea docente. Otros países han adoptado estrategias similares, como es el caso del Sistema de Medición de la Calidad de la Educación (SIMCE) de la República de Chile, cuyos resultados se dan a conocer por diversas vías.

Otro tanto podría decirse de los informes de evaluación elaborados por diversos organismos internacionales. Por ejemplo, la International Association for the Evaluation of Educational Achievement (IEA) suele elaborar en sus estudios más recientes diversos tipos de documentos: informes globales de resultados,

informes detallados acerca de algunos aspectos concretos (como diferencias de género o factores explicativos de los resultados, etc.), informes técnicos y folletos de síntesis. Mientras que los primeros se destinan a las autoridades nacionales y al gran público, ofreciendo una información relevante pero sintética, el segundo y tercero van más orientados a los investigadores y especialistas, ofreciendo datos técnicamente más elaborados. El cuarto tipo de informe constituye una fuente de información amplia a públicos diversos, poco especializados y que no requieren una información exhaustiva.

En resumen, podría afirmarse que la elaboración de informes diferenciados para las diversas audiencias de la evaluación presenta más ventajas que inconvenientes. Sin embargo, es necesario adoptar las cautelas necesarias para prevenir los riesgos que dicha decisión implica, sobre todo el posible déficit democrático y las eventuales manipulaciones o distorsiones que podrían producirse.

3. La comparación en evaluación

Uno de los métodos utilizados más habitualmente en los informes de evaluación para tratar y presentar la información obtenida es el análisis comparativo. En algunos ámbitos concretos, como es el caso de la evaluación de los resultados académicos alcanzados por los alumnos o de los logros de las instituciones educativas, esta práctica se ha convertido en la predominante, bien sea empleándola de manera exclusiva o en combinación con otras. Para comprobar lo ajustado de esa observación, basta la simple consulta de algunos de los informes elaborados por los organismos e instituciones citados en el epígrafe anterior.

El recurso a la comparación como uno de los procedimientos habituales de análisis y tratamiento de datos en los estudios de evaluación no es fruto de una decisión caprichosa ni casual, sino que tiene que ver con la propia esencia de la tarea evaluadora. En efecto, al inicio del documento se afirmaba que evaluar consiste en emitir juicios a partir de una información recogida y tratada de forma sistemática. Pero esa emisión de juicios implica realizar una comparación con algún elemento de referencia, bien sea interno o externo. En rigor, no puede formularse una conclusión acerca del valor o mérito de una determinada realidad si no se compara con otra o con un modelo ideal previamente construido. Por eso puede afirmarse que, en última instancia, la evaluación implica la comparación.

3.1. El tratamiento comparativo de la información

Entre las operaciones comparativas más frecuentemente utilizadas en los informes de evaluación destacan tres que conviene examinar con cierto detenimiento. La primera de ellas consiste en comparar los datos obtenidos con un *criterio o norma* que sirva de referencia para la emisión de un juicio de valor.

Es el enfoque que se utiliza, por ejemplo, para realizar lo que se conoce como evaluación criterial del aprendizaje, consistente en el establecimiento de unos criterios precisos para determinar los niveles de rendimiento y en el estudio subsiguiente de su cumplimiento o no por parte de cada alumno. También se aplica un procedimiento semejante para llevar a cabo la acreditación de centros escolares, a partir del contraste de las características de la institución evaluada con un cierto número de requisitos previamente acordados. En uno y otro caso, el mérito de aquello que se valora se determina de acuerdo con el grado de cumplimiento de los criterios preestablecidos y no mediante la posición relativa ocupada por el alumno o el centro en el seno del grupo al que se encuentre adscrito.

La segunda de las operaciones mencionadas consiste en realizar comparaciones entre varias entidades semejantes, de manera *sincrónica o transversal*. En este caso, las diversas realidades objeto de evaluación se comparan entre sí, de acuerdo con la situación que presentan en un momento determinado. Es el tipo de procedimiento utilizado para llevar a cabo la evaluación normativa del aprendizaje, basada en otorgar calificaciones mediante la comparación del alumno con sus compañeros y la determinación del lugar que ocupa en su grupo de referencia. Es también el tratamiento que se realiza, por ejemplo, en los estudios internacionales de rendimiento educativo, que tanto auge han experimentado en los últimos años, o el que se emplea para construir tablas de clasificación de centros escolares, como las que algunos países han publicado recientemente.

La tercera operación consiste en comparar la información obtenida a partir de una misma realidad, en dos o más momentos sucesivos, de manera *diacrónica o longitudinal*. En esta aproximación, la realidad evaluada no se compara con otras semejantes ni con un modelo ideal, sino consigo misma, pero introduciendo en el análisis la dimensión temporal. Es el tipo de tratamiento que se realiza, por ejemplo, en los estudios de seguimiento de cohortes de estudiantes con objeto de conocer su progreso académico a lo largo del tiempo, o cuando se pretende evaluar el impacto de determinadas medidas de política educativa adoptadas en un momento dado.

Cada una de estas tres operaciones ofrece diversas posibilidades, al tiempo que plantea algunos problemas. El examen de sus respectivas ventajas e inconvenientes puede ayudar a entender mejor su contribución en cada situación concreta, así como las cautelas con que deben utilizarse.

El primero de los procedimientos apuntados, la comparación con una norma o criterio prefijado, puede parecer el más deseable desde el punto de vista de la evaluación, ya que supone la valoración de una realidad en y por sí misma. De hecho, es un enfoque muy utilizado en la actualidad y que ha inspirado un buen número de iniciativas. Entre todas ellas merece la pena destacar especialmente los intentos emprendidos en diferentes países para establecer *criterios o*

estándares (standards) de rendimiento, entendidos como una formulación explícita de los niveles que los alumnos deberían alcanzar en las diversas etapas de su proceso de aprendizaje y que sirven de referencia a la hora de la evaluación (OCDE, 1995). Uno de los ejemplos mejor conocidos y más asentados de este tipo de aproximación es la iniciativa desarrollada desde comienzos de la década de los noventa en Inglaterra y Gales, consistente en el establecimiento de unos *objetivos de logro (attainment targets)* para cada una de las áreas del currículo escolar, que son objeto de evaluación individualizada en unos momentos clave del proceso de escolarización (a los siete, once, catorce y dieciséis años de edad). Otros ensayos de este tipo se llevan a cabo en la actualidad en los Estados Unidos.

Como se indicaba anteriormente, también los procedimientos de acreditación de centros se han regido por una lógica similar. El mecanismo utilizado para realizar dicho proceso no es otro que determinar los requisitos que deberían cumplir las instituciones candidatas, examinar el grado de cumplimiento de los mismos en cada caso concreto y realizar una valoración global en términos de aceptación o denegación.

Este primer enfoque resulta indudablemente atractivo, ya que constituye el modo más transparente y menos controvertido de valorar el mérito de una realidad sometida a evaluación. En efecto, la existencia de una referencia externa y explícita evita la subjetividad en la valoración y confiere mayor credibilidad y confiabilidad al juicio emitido. La tarea de acordar los criterios en los que se debe basar el juicio de valor está obviamente lejos de ser sencilla, pero ofrece la innegable ventaja de prestar un apoyo sólido a la actividad evaluadora.

Sin embargo, a pesar de su atractivo, el principal problema que plantea este procedimiento consiste precisamente en la dificultad de determinar los criterios que deben servir de referencia para la comparación. En esta operación intervienen consideraciones de muy diverso tipo y no solamente técnicas. De hecho, en la mayor parte de los casos, la determinación de dichos criterios es fruto de un doble proceso, de carácter respectivamente político y científico-técnico. Así, por ejemplo, para fijar los objetivos de aprendizaje a alcanzar por los alumnos de un grado escolar determinado hay que comenzar teniendo en cuenta las aportaciones de la psicología y la pedagogía acerca de cómo se desarrollan las capacidades individuales asociadas a dichos objetivos. Pero a continuación hay que determinar cuáles son los niveles académicos realmente alcanzables o deseables en ese grado concreto, lo que supone tener en cuenta otras variables de carácter contextual, tales como el nivel educativo general y las demandas de la población, la formación de los profesores, la disponibilidad de recursos o los logros previamente conseguidos. Por último, la revisión del nivel de cumplimiento de los mencionados objetivos, a partir del análisis de los resultados efectivamente alcanzados por los alumnos, aconsejará su mantenimiento o su modificación. Como puede fácilmente apreciarse, para tomar decisiones razonables en este campo se han de combinar adecuadamente el

bagaje científico y técnico disponible con el conocimiento de la realidad educativa, la consideración de lo que es socialmente deseable y, por qué no decirlo, una cierta prudencia y sentido de la oportunidad. Al desbordar el ámbito estrictamente científico y técnico e invadir el propiamente político, aumentan las dificultades para fijar este tipo de criterios y corre el riesgo de disminuir el grado de acuerdo en su elaboración.

Por otra parte, la determinación de este tipo de criterios plantea el interesante problema conceptual y operativo de su aplicabilidad general o contextualizada. Dicho de otro modo, hay que responder a la pregunta de si pueden definirse unos criterios de aplicación general o si la definición debe depender de los contextos particulares. Si bien la primera opción resulta más atractiva desde el punto de vista político y cuando se trata de proporcionar una información amplia, dada su simplicidad, la segunda resulta más rigurosa desde el punto de vista científico-técnico, pero es indudablemente más compleja. La cuestión continúa estando abierta en buena medida, requiriendo todavía la realización y el análisis de más experiencias de esta índole.

Cuando no se cuenta con unos criterios ampliamente aceptados con los que contrastar las características del objeto a evaluar, por los motivos que sea, un procedimiento habitual consiste en compararlo con otras realidades semejantes. Esta segunda aproximación resulta especialmente atractiva para los políticos y para el gran público, dada su simplicidad y su inmediata interpretación. Sin embargo, es la más criticada por los especialistas, por los riesgos que encierra y los posibles abusos que implica. Por citar un solo ejemplo, Miguel Angel Santos incluye entre los usos abusivos de la evaluación el de “aprovechar[la] para hacer falsas comparaciones entre lo que es, en realidad, incomparable” (Santos, 1993:72).

Este modo de tratamiento y presentación de la información es frecuentemente utilizado en los estudios de evaluación del rendimiento educativo. El ejemplo más característico lo constituyen las tablas comparativas de resultados (*league tables*), semejantes a clasificaciones (de centros, países o regiones) elaboradas a partir de la aplicación de pruebas estandarizadas, que son una práctica muy extendida en el mundo educativo. En el ámbito internacional, este procedimiento es sin duda el predominante. Así, la I.E.A. (a la que ya se ha hecho mención) viene publicando este tipo de tablas desde hace más de treinta años, como resultado de los diversos estudios que ha emprendido en su dilatada existencia (Degenhart, 1990). También se basan en esta aproximación la mayor parte de los indicadores internacionales de la educación incluidos en los volúmenes publicados desde 1992 por la OCDE con el título *Education at a Glance/Regards sur l'éducation*. Y encontramos ejemplos recientes de su uso en un contexto nacional en la publicación (por el propio Ministerio de Educación y Empleo) de tablas clasificatorias de los centros escolares ingleses, o en las que elabora la prensa francesa o la chilena a partir de los resultados obtenidos por los alumnos en exámenes oficiales o en pruebas estandarizadas de rendimiento.

A pesar de su atractivo, simplicidad y tradición, el uso de este procedimiento plantea diversos inconvenientes y dificultades. Un primer problema es de orden conceptual y tiene que ver con la selección de los indicadores utilizados para llevar a cabo la comparación. En efecto, ante la dificultad o imposibilidad de medir directamente una realidad compleja y no completamente aprehensible, como es la educativa, la práctica común consiste en seleccionar algunos indicadores capaces de representarla adecuadamente y centrarse en el análisis de los mismos (Tiana, 1997).

Este problema se plantea con especial agudeza cuando se trata de evaluar el rendimiento educativo, lo que constituye un caso muy frecuente. Así, la mayor parte de las tablas clasificatorias pretenden, en última instancia, valorar la *calidad de la educación* en los centros o países objeto de comparación. Para ello se utiliza un reducido grupo de indicadores, que suelen consistir en los resultados alcanzados por los alumnos (o por una muestra de ellos) en diversas áreas escolares. Pese a las cautelas generalmente expresadas en los informes de evaluación que adoptan este tratamiento, sus lectores e intérpretes suelen efectuar una inferencia mediante la cual tales indicadores parciales pasan a representar el grado absoluto de calidad educativa. En tales circunstancias, la búsqueda de unos indicadores suficientemente comprensivos, relevantes y significativos se convierte en una prioridad conceptual, no siempre bien resuelta.

Un segundo problema (o mejor dicho, grupo de problemas) es de orden metodológico y técnico. Por una parte, incluso aunque la selección efectuada de los indicadores más adecuados para valorar una determinada realidad sea inobjetable, no siempre resulta fácil reunir la información necesaria para proceder a su comparación. Dicha dificultad puede ser de carácter técnico (falta de instrumentos adecuados, válidos y/o fiables de evaluación), económico (coste excesivo de la recogida de datos) u operativo (excesiva complejidad de los procedimientos a utilizar). En cualquier caso, la consecuencia es que los indicadores previamente seleccionados se ven muchas veces reducidos en número o en amplitud, produciéndose una simplificación abusiva y no siempre deseada. Por otra parte, cuando la comparación se basa en la aplicación de cuestionarios estandarizados, no se puede desdeñar el efecto producido por la heterogeneidad de condiciones de su aplicación, que a veces puede sumarse a la existencia de sesgos culturales. Aunque ésta no sea la objeción principal que pueda hacerse a este procedimiento, no deja de constituir un inconveniente serio en algunos estudios internacionales de rendimiento, como algunas voces han puesto de relieve (Theisen, Achola y Boakari, 1990).

Desde el punto de vista metodológico, más serias resultan las objeciones expresadas por algunos autores prestigiosos y que han dado pie a una interesante polémica internacional (Goldstein, 1993; Goldstein y Spiegelhalter, 1996). Según dichos análisis, el tomar la puntuación media obtenida por un país o un centro como expresión precisa de su rendimiento, olvidando sus respectivos intervalos de confianza, es la causa de muchas interpretaciones

incorrectas de los datos. En efecto, si se tienen en cuenta los intervalos de confianza de las medidas efectuadas, que suelen superponerse en mayor o menor proporción, la conclusión es que no puede descartarse el efecto del azar en el orden en que aparecen colocadas la mayor parte de las unidades comparadas. Dicho de otro modo, no puede excluirse que el orden pudiera ser muy diferente, simplemente por efecto del azar. En opinión de los autores citados, el análisis de las diferencias entre instituciones suele autorizar solamente a individualizar los casos extremos, sin que se aprecien diferencias estadísticamente significativas entre la mayor parte de los componentes de la lista. Tal objeción no llega a invalidar este tipo de comparación, pero obliga a introducir serias cautelas en el análisis, sobre todo cuando los resultados se comunican al gran público. Goldstein y Spiegelhalter incluso llegan a decir, refiriéndose a este tipo de comparación, que “los gobiernos deben preocuparse de que los usuarios potenciales están debidamente informados de sus limitaciones” (Goldstein y Spiegelhalter, 1996:405).

Un tercer problema consiste en la necesidad de contextualizar la comparación. Esta exigencia es especialmente urgente cuando se habla de evaluar el rendimiento, ya que no es ésta una realidad que se produzca en el vacío, sino que se explica en buena medida por su contexto. De ahí deriva la necesidad de tener en cuenta lo que se ha dado en llamar el *valor añadido* por el centro, efectuando lo que Goldstein prefiere denominar *comparaciones ajustadas* (*adjusted comparisons*), tema éste que será objeto de reflexión más detenida en el apartado siguiente.

Un cuarto problema, que también está lejos de ser irrelevante, tiene que ver con la unidad de comparación elegida. La elección de una perspectiva microscópica, centrada en el centro escolar, o de otra macroscópica, abarcando conglomerados más amplios, determina la interpretación que se haga de los datos comparativos. Cuando una evaluación tiene una finalidad diagnóstica o de información pública, el recurso a conglomerados de centros o alumnos resulta más aconsejable. Pero cuando el propósito es fundamentalmente de mejora, no se debe olvidar que ésta se produce con mayor éxito en el nivel escolar, debiendo orientarse hacia los centros el foco de atención (baste recordar al respecto lo que se decía hace varias páginas acerca de los niveles de aplicación y de análisis de la evaluación). Incluso cuando se trata de realizar análisis desde una perspectiva amplia no se deben olvidar las reservas expresadas por algunos autores: “la mayoría de los sistemas educativos nacionales son conglomerados altamente complejos y heterogéneos de alumnos, escuelas y administradores; en consecuencia, la atribución de características relacionadas con el rendimiento a poblaciones locales sobre la base de resultados de pruebas nacionales está viciada por una falacia ecológica” (Theisen, Achola y Bokari, 1990:39-40).

Como puede apreciarse por esta simple enumeración de problemas, las dificultades que encuentra este tipo de comparación son importantes. No quiere ello decir que no se deba utilizar, sino más bien que hay que hacerlo con mucha

cautela y precaución. Aunque resulta un procedimiento muy atractivo, los riesgos que entraña no son en absoluto desdeñables.

El tercer procedimiento de los aquí presentados complementa los dos anteriores y ayuda a superar algunas de sus dificultades. Por ejemplo, ante la falta de criterios ampliamente acordados acerca de los niveles de rendimiento deseables o ante la inseguridad metodológica de la comparación transversal, la longitudinal permite evaluar el progreso educativo mediante un análisis diacrónico, efectuado a lo largo del tiempo.

Sin embargo, también este procedimiento plantea algunos problemas. En primer lugar, volvemos a encontrar la dificultad conceptual y metodológica antes mencionada para definir los indicadores más convenientes y recoger la información necesaria para su cálculo. El hecho de que este tipo de comparación permita contrastar una realidad consigo misma no evita este inconveniente. En segundo lugar, la comparación longitudinal exige tiempo suficiente, ya que la presentación de resultados se demora notablemente. Ello le resta atractivo desde el punto de vista político, debido al gran desfase que existe entre el momento de inicio del estudio y el del análisis de sus conclusiones. Por último, exige instrumentos y criterios estables a lo largo del tiempo, lo que a menudo entra en contradicción con el deseo legítimo de mejorar los proyectos de trabajo tan pronto como se detecta alguna deficiencia. Por todos estos problemas, no es extraño que haya sido el procedimiento menos utilizado de los tres analizados, pese al interés que sin duda ofrece.

Como puede apreciarse tras la lectura de las líneas anteriores, la comparación puede considerarse uno de los modos más habitualmente utilizados de tratamiento y presentación de la información en la tarea evaluadora. No obstante, los conceptos, los métodos y las técnicas de que actualmente disponemos para realizarla distan de ser los mejores. Ante la falta de otros procedimientos es legítimo y razonable recurrir a los que se acaban de presentar. Pero ello no excusa de la necesidad de avanzar más decididamente en la exploración de nuevas posibilidades que reduzcan los riesgos denunciados, por una parte, y de utilizar mientras tanto los existentes con suma cautela y precaución, por otra.

3.2. Las condiciones de una comparación justa

De acuerdo con una sugerente formulación de E.R. House, incluida en uno de sus textos más difundidos e influyentes, cualquier empresa evaluadora debe respetar tres principios clave: veracidad, coherencia y justicia (House, 1994). Si la aplicación efectiva de los tres constituye una exigencia ineludible, el tercero cobra especial importancia cuando se realizan análisis comparativos, dadas las consecuencias fuertemente dañinas que la comparación injusta puede acarrear para las entidades evaluadas.

El respeto al principio de justicia enunciado por House tiene mucho que ver con la adecuada contextualización de la información recogida y presentada a los destinatarios de la evaluación. En el caso de estudios de rendimiento de alumnos, centros o sistemas educativos, este requisito cobra una importancia capital. En efecto, como se indicaba en el apartado anterior, la evaluación del rendimiento debe ser debidamente contextualizada, teniendo en cuenta una serie de factores que inciden en los resultados conseguidos por los alumnos y los centros y que no tienen su origen en la simple diferencia de capacidades individuales. Variables tales como la extracción social de los alumnos, la actitud de las familias hacia la educación, el *clima* o ambiente escolar, la organización del centro, el estilo pedagógico de los profesores o el funcionamiento de la dirección pueden contribuir a explicar buena parte de las diferencias registradas entre los centros escolares.

De todos los factores señalados (y algunos más que podrían apuntarse), unos son fruto de la actuación educativa de las familias, los profesores y los centros escolares, mientras que otros vienen determinados de antemano y no puede influirse en ellos fácilmente. Así, un centro puede mejorar su organización, reforzar la formación de sus profesores, crear un ambiente favorable al estudio o poner en marcha programas valiosos de intervención psicopedagógica, pero no puede modificar la extracción social o cultural de sus alumnos, salvo que los sustituya por otros. Por ese motivo, la diferencia existente entre las diversas condiciones de partida debe ser tomada en cuenta a la hora de evaluar rendimientos. No sería justo ensalzar o denigrar a un centro simplemente por el efecto de las condiciones iniciales de su alumnado. De este modo, llevar a cabo una comparación justa implicaría neutralizar el efecto de tales condiciones sobre el rendimiento, valorando los resultados logrados a partir de un nivel teóricamente homogéneo.

Sin embargo, sabemos que dicha situación ideal está lejos de darse en la realidad. La diversidad de condiciones de partida de los alumnos y de los centros es muy considerable, explicando por sí sola parte de las diferencias observadas en los resultados. La necesidad de tener en cuenta esa diversidad a la hora de comparar rendimientos, sentando así las bases para una comparación justa, es lo que ha llevado a introducir la noción de *valor añadido*. La idea intuitiva que subyace bajo ese término consiste en reconocer que los centros pueden producir efectos diversos dependiendo de sus condiciones particulares, pero que la valoración de sus logros debe hacerse en función de la progresión conseguida (en términos de diferencia entre la situación final y la inicial) y no solamente de los resultados brutos obtenidos. De acuerdo con esta noción, el rendimiento de un centro A podría ser superior al de otro B aunque el resultado bruto del segundo fuese superior al primero; la clave estaría en la magnitud del avance registrado.

Esta idea de tomar en consideración la situación de partida de los centros y de sus alumnos ha inspirado un buen número de iniciativas recientes. El propio país

que más decididamente ha impulsado en los últimos años la publicación de tablas clasificatorias, Inglaterra, se ha visto obligado a introducir este tipo de matizaciones debido a la presión de los centros más desfavorecidos, aunque por el momento dichos intentos no han pasado de ser exploratorios (Department for Education, 1995). La consecuencia de los avances registrados es que en la actualidad el término *valor añadido* se utiliza extensamente en el lenguaje de la evaluación educativa, si bien algunos autores ponen de manifiesto la poca precisión del concepto y prefieren utilizar otros como *comparaciones ajustadas* (Goldstein y Spiegelhalter, 1996).

Uno de los análisis más interesantes y prácticos realizados acerca de los diversos modelos teóricos y metodológicos existentes para hacer operativo el concepto de *valor añadido* se debe a Sally Thomas y Peter Mortimore. A partir de los resultados obtenidos en el GSCE de 1993 por más de once mil alumnos de 87 escuelas de Lancashire, los tests (*NFER Cognitive Abilities Test*) aplicados a la entrada en la etapa secundaria, los datos del Censo de 1991 y los registros escolares de los alumnos, elaboraron y contrastaron empíricamente cinco modelos de *valor añadido*, progresivamente más sofisticados. A continuación, compararon los resultados obtenidos por medio de la aplicación de los cinco modelos y llegaron a interesantes conclusiones (Thomas y Mortimore, 1995).

De acuerdo con su estudio, el factor con mayor importancia a la hora de estudiar el *valor añadido* por un centro consiste en el *rendimiento inicial (prior attainment)* de sus alumnos. La inclusión en el análisis de datos acerca de dicho rendimiento inicial, junto a algunas características del alumnado, tales como su sexo, edad, extracción familiar y contexto socioeconómico, permite controlar la mayor parte de la variación debida a las condiciones de partida y estimar así con justicia y validez el *valor añadido* por el centro. En ausencia de dichos datos, las características contextuales del centro pueden proporcionar una aproximación, de mucha menor significación, pero digna de ser tenida en cuenta. Según este análisis, la estimación precisa del *valor añadido* requeriría disponer de datos relativos al rendimiento inicial de los alumnos y a sus variables contextuales individuales. Este requisito debe ser tenido en cuenta a la hora de emprender estudios de evaluación del rendimiento que quieran superar el simple estadio de la medición de los resultados brutos y efectuar una comparación justa.

Además de este primer resultado, que refuerza la necesidad de obtener y manejar datos relativos al rendimiento inicial de los alumnos y de sus características básicas, el estudio de Thomas y Mortimore proporcionó otras conclusiones de gran interés para el diseño y la puesta en marcha de estudios de evaluación del rendimiento:

- 1) En primer lugar, los análisis realizados demostraron la existencia indudable de un denominado *efecto escuela* (entendido como la influencia ejercida específicamente por el centro sobre el rendimiento de sus alumnos), ya que se

apreciaron diferencias de rendimiento entre las instituciones participantes a pesar de controlar el efecto de las mencionadas variables individuales.

2) En segundo lugar, el cálculo del *valor añadido* proporcionó una imagen más realista y fiel de los logros de un centro que los simples resultados brutos obtenidos por sus alumnos en una prueba o examen. No obstante, los análisis realizados venían condicionados por la calidad, fiabilidad y validez de los datos utilizados, insistiendo así el estudio en la necesidad de cuidar los procedimientos utilizados para obtenerlos.

3) No obstante lo anterior, incluso teniendo en cuenta el cálculo del *valor añadido*, hay que señalar que los intervalos de confianza de los resultados logrados por los centros presentaron un grado considerablemente de solapamiento, lo que obligó a valorar con precaución las diferencias obtenidas. En realidad, sólo podría hablarse de diferencia real de eficacia entre dos centros cuando sus intervalos de confianza no se solapasen en absoluto, lo que constituye un caso no excesivamente habitual.

4) Otro resultado realmente interesante consistió en que el *efecto escuela* difería al considerar diversas áreas (Matemáticas y Lengua, por ejemplo), lo que induce a pensar que el *valor añadido* por los diferentes departamentos de un mismo centro no es idéntico. En cierto sentido, el centro no es completamente homogéneo a la hora de influir sobre sus alumnos.

5) Igualmente, el *efecto escuela* no se manifestaba con la misma intensidad en relación a distintos grupos de alumnos. Los análisis realizados pusieron de manifiesto que es frecuente que cada escuela beneficie especialmente a algún tipo determinado de alumno (unas a los de mayor rendimiento, otras a los de menos, o a los de determinado estilo de aprendizaje, etc.). Este resultado abre la puerta a análisis posteriores acerca de la posible concordancia entre ciertas características del centro y el tipo de alumno al que más beneficia.

6) De acuerdo con las observaciones anteriores, un planteamiento riguroso acerca del *valor añadido* obligaría a efectuar su cálculo para las puntuaciones totales promedio, para las de cada materia y para las obtenidas por grupos de alumnos de diferentes características, ya que el efecto no es idéntico en todos los casos.

Como puede fácilmente apreciarse, la investigación acerca del *valor añadido* se encuentra todavía en sus fases iniciales. No obstante, las posibilidades que encierra son tan ricas y prometedoras que animan a los evaluadores a continuar avanzando en esa dirección. El logro de una comparación efectivamente justa tiene mucho que ver con el éxito del empeño por conceptualizar y operativizar esa noción.

4. La difusión de la información

Antes de finalizar este documento, aún conviene detenerse a reflexionar algo más acerca de los modos y los canales a través de los cuales se difunde la información proporcionada por la evaluación. Aunque a lo largo de los apartados anteriores se hayan realizado observaciones que, de manera más o menos implícita, sostienen la importancia de dicha tarea de difusión, vale la pena subrayarla explícitamente, al tiempo que se analizan algunas de sus implicaciones. Es éste uno de esos asuntos sobre los que probablemente nunca se insistirá bastante.

En efecto, frente a la idea extendida de que el evaluador termina su tarea con la emisión del informe final, hoy en día tiende cada vez más a considerarse que ese momento constituye un hito importante pero no clausura el proceso de evaluación. Por una parte, ha de asegurarse que los principales resultados de la evaluación son devueltos a las audiencias implicadas en la misma; por otra, ha de sacarse el máximo provecho de la información obtenida, con vistas a la mejora de actividad educativa. Y para ambos propósitos, la difusión constituye una pieza clave.

4.1. Importancia de la difusión en la evaluación

Ese cambio de perspectiva se ha dejado sentir en los países anglosajones en la propia denominación de la tarea a la que se refiere. Así, el término *difusión* ha ido dejando paso al de *diseminación*, queriendo significar precisamente que no se trataba sólo de organizar la publicación y distribución de los informes de evaluación, sino que la etapa comprendía otro conjunto de tareas de más largo alcance. El término *diseminación*, con esa connotación de *sembrar o esparcir* que recoge el *Diccionario de la Lengua Española* y que también posee en otras lenguas, transmitiría el mensaje de que la evaluación debe depositar su germen de mejora en aquellos terrenos mejor dispuestos para hacerlo fructificar. Para ello no basta una mera distribución de materiales escritos, sino que es necesario adoptar una actitud activa y utilizar una multiplicidad de recursos comunicativos. Aun reconociendo que la idea es sugerente, se ha preferido mantener aquí el término de difusión, por considerarlo más arraigado en la lengua castellana. Pero no debe olvidarse que el significado que se le da en los párrafos que siguen desborda ampliamente la acepción estrecha del mismo.

Al hablar de difusión, nuevamente vuelve a aparecer el concepto de *audiencias* de la evaluación y la consideración de sus necesidades. La exigencia de dar respuesta a las demandas explícitas o implícitas de las audiencias es precisamente una de las razones más poderosas en favor de la difusión. En efecto, el suministro de información únicamente a los promotores de la evaluación podría muy bien llevarse a cabo mediante el informe final, unido quizás con otros medios de comunicación personal. Pero si pensamos en audiencias más diversificadas y destinatarios más amplios, esos medios resultan demasiado limitados. Como se indicaba anteriormente, el conjunto de personas o colectivos implicados e interesados en la evaluación suele ser más extenso

que el de sus promotores. Aunque no debemos pensar necesariamente en grupos demasiado amplios, tampoco podemos reducirnos a una evaluación cuyos resultados competen e interesan tan sólo a los responsables de la actividad evaluada o a sus superiores jerárquicos.

El hecho de considerar que una evaluación debe responder a las necesidades de sus audiencias obliga a pensar en una difusión que las tenga a todas ellas en cuenta. Esta reflexión ha llevado a algunos autores a distinguir una difusión *primaria* de otra *secundaria*. Si aquélla consistiría, en última instancia, en elaborar y suministrar los informes correspondientes a los patrocinadores de la evaluación, ésta implicaría comunicar los resultados y recomendaciones principales de modo tal que responda a las necesidades de las diversas audiencias (Rossi y Freeman, 1993:451-452). Siguiendo este razonamiento, podríamos afirmar que una adecuada difusión constituye un requisito indispensable de una evaluación de carácter democrático.

Esta exigencia (de transparencia activa, podríamos decir) se convierte en ineludible cuando hacemos referencia a evaluaciones que interesan a sectores sociales o institucionales muy diversos. El caso más extremo es el de los estudios que tienen como objeto el conjunto del sistema educativo, que pueden llegar a interesar a toda la población de un país. En estas circunstancias, la restricción en el acceso a la información obtenida resulta especialmente reprobable. Pero también pueden aplicarse estas reflexiones a otros estudios de evaluación de menor envergadura, aunque no de menor importancia para sus audiencias, como sería el caso de la evaluación de centros o de programas educativos. No quiere ello decir que la información ofrecida deba ser idéntica para todas las audiencias en cada situación concreta, pero debe insistirse en que se trata de diferencias de cantidad y detalle de datos y análisis más que de calidad y naturaleza de los mismos.

Por otra parte, no se puede olvidar la segunda de las razones mencionadas a favor de una adecuada difusión. No sólo se trata de difundir información en cumplimiento de un deber democrático, sino también de ofrecer elementos de juicio orientados a la mejora. Si la evaluación es rigurosa y creíble, debe permitir a los implicados en la misma revisar sus pautas de actuación, los medios que utilizan y los efectos de su acción, con el propósito explícito de contribuir a mejorar su parcela concreta de la realidad educativa. Y mal puede pensarse en planes o programas de mejora sin un sentimiento compartido acerca de su necesidad y un conocimiento preciso de los principales puntos fuertes y débiles. La difusión juega un papel crucial en la creación de ese sentimiento y en la generación de dicho conocimiento, de donde deriva su importancia.

Esta preocupación por difundir adecuadamente la información no es exclusiva del ámbito de la evaluación. Los investigadores y los promotores de la investigación educativa están también explorando vías semejantes para revitalizar su incidencia en el campo de la práctica y de la política de la

educación (Reimers, McGinn y Wild, 1995). De acuerdo con estas ideas, muchos proyectos recientes de investigación y/o evaluación han dedicado una atención especial a los procedimientos utilizados para llevar a cabo su difusión. Como muestra de esta realidad puede bastar con un ejemplo significativo, tomado del ámbito internacional.

Hace unos años, la OCDE inició un proyecto para estudiar y analizar las experiencias de innovación que se desarrollan en diversos sistemas educativos en el ámbito de las ciencias, las matemáticas y la tecnología. La iniciativa se concretó en la realización de veintitrés estudios de casos en trece países miembros de la organización. Aunque el proyecto no tuviese como finalidad explícita la evaluación de los casos estudiados, no cabe duda de que ésta fue una de las vertientes principales del proyecto. Siguiendo una tendencia que se vislumbra claramente en el panorama actual, la investigación, la evaluación y el análisis cualitativo de experiencias de innovación estaban íntimamente ligadas en una misma actividad. Pues bien, a la finalización del proyecto, el equipo internacional responsable del mismo redactó y publicó un informe, que fue ampliamente anunciado a través de diversos medios (Black y Atkin, 1996). No obstante, la tarea de difusión de los resultados que se emprendió a continuación fue mucho más allá de dicha publicación. En primer lugar, se realizaron algunas publicaciones nacionales, recogiendo sus estudios de casos concretos y en ocasiones algunas reflexiones más generales. En segundo lugar, se discutió el informe final en el Comité Director del Centro de Investigación e Innovación Educativa (CERI) de la OCDE. En tercer lugar, se presentaron diversos trabajos y comunicaciones en diferentes congresos y reuniones científicas, algunos de ellos tan prestigiosos y concurridos como los encuentros anuales de la *American Educational Research Association (AERA)*. Pero además de estas iniciativas, el CERI organizó una serie de reuniones de difusión, celebradas en varios países, en las cuales se han ido presentando y discutiendo los resultados obtenidos con audiencias internacionales muy variadas, con participación de profesores de diversos niveles educativos y responsables políticos y administradores de la educación.

Sin duda, este último elemento ha constituido una novedad que se ha revelado fructífera. El hecho de dar la palabra a personas y colectivos no implicados en la realización del estudio pero afectados por el mismo, ha representado la celebración de debates tan ricos como extensos, logrando la comunicación de los resultados obtenidos a públicos muy diversos. Este tipo de experiencias nos está enseñando mucho acerca de cómo llevar a cabo tareas de difusión, entendiéndolas de un modo más actual y renovado de lo que era tradicional.

Otro tipo de iniciativas son las emprendidas por algunos de los países que han puesto recientemente en marcha programas nacionales de evaluación educativa. Uno de los casos más atractivos y sugerentes es el de Francia. La *Direction de l'Évaluation et la Prospective (DEP)* realiza una difusión extensa e intensa de la información obtenida a través de sus diversos proyectos de trabajo. Por una

parte, publica anualmente el folleto titulado *L'Etat de l'Ecole*, que recoge una serie de treinta indicadores del estado y situación del sistema educativo francés. En vísperas del debate presupuestario, lo entrega a los parlamentarios, además de enviarlo a las autoridades educativas y de ponerlo a disposición del gran público. Por otra parte, publica los informes titulados *Géographie de l'école*, donde se recoge información acerca del estado de la educación en las diversas regiones. Difunde así mismo la revista trimestral *Education & Formations* y la serie titulada *Les Dossiers d'Education et Formations*, destinados a un público más especializado, al que proporcionan una información detallada y elaborada, procedente de una diversidad de fuentes. En plan de divulgación, difunde abundantes *Notes d'information*, en las que recoge información sintética acerca de estudios en marcha o recientemente finalizados, para su conocimiento general. Junto a esa impresionante tarea editorial, la DEP efectúa presentaciones públicas de sus datos en reuniones de profesores, inspectores o padres de familia, sus miembros envían trabajos a congresos científicos y no rehuyen su presencia en los medios de comunicación. En conjunto, como puede apreciarse, desarrollan una ingente tarea de difusión de información, que poco a poco va calando en la opinión pública.

Con unos planteamientos más modestos, merece la pena sin embargo destacar las iniciativas que se han puesto en marcha en los últimos años en algunos países iberoamericanos para difundir los resultados de la evaluación. Entre ellos, pueden recordarse la publicación de los *Manuales de orientaciones para la interpretación de resultados* del Sistema de Medición de Calidad de la Educación (SIMCE) de Chile o de las *Recomendaciones Metodológicas para la Enseñanza* tras los Operativos Nacionales de Evaluación de la República Argentina, anteriormente mencionados. Sin duda, la orientación que sigue todo este conjunto de iniciativas apunta nuevas tendencias para el futuro en lo que se refiere a la difusión de la información.

4.2. Estrategias de difusión

Los ejemplos anteriores nos permiten acercarnos a las principales reglas estratégicas que deben regir las actividades de difusión. Aunque no sea posible (ni es la intención de este apartado) elaborar un catálogo de normas precisas de actuación, pueden sin embargo formularse algunos principios generales aplicables a este campo particular.

En primer lugar, hay que insistir en la idea de que es necesario planificar la difusión desde el momento inicial, cuando se elabora el proyecto de evaluación. La voluntad de comunicar la información obtenida a las audiencias interesadas debe estar presente desde el comienzo de la actividad evaluadora, sin dejarse al azar o regirse por la improvisación. Los modos, momentos y estrategias de difusión han de quedar patentes en el propio proyecto de evaluación, como una tarea bien trabada con las demás.

En segundo lugar y en conexión con lo anterior, la difusión ha de planificarse como una actividad continua, aunque con momentos de mayor énfasis que otros. Si se acepta la tesis que planteábamos unas páginas más arriba de que el proceso de información es continuo, hay que concluir que la difusión habrá de acomodarse en líneas generales a ese flujo ininterrumpido.

En tercer lugar, hay que subrayar la importancia que tiene utilizar una diversidad de canales y procedimientos de difusión. Así, los informes escritos deben alternarse con la información oral, los mensajes dirigidos al gran público con las reuniones especializadas, el tratamiento en que prima la accesibilidad con la presentación técnicamente rigurosa. Por una parte, cada uno de ellos se adapta mejor a unos destinatarios determinados; por otra, los diversos procedimientos complementan mutuamente la información que ofrecen. No se trata de abrumar a nadie con una avalancha de resultados de la evaluación, sino tan sólo de facilitar que lleguen a todos los que tengan interés en ellos y del modo más adecuado a sus características personales o colectivas.

En cuarto lugar, la difusión no puede dissociarse de la necesaria formación a los receptores de la información. Incluso en sociedades educativamente desarrolladas e informadas, son muy frecuentes los malentendidos acerca de cómo deben interpretarse los resultados de la evaluación. Hace unos años se produjo un caso muy curioso en España, a raíz de realizarse un estudio internacional de comprensión lectora (Elley, 1992). Dado que los alumnos de nueve años ocupaban el lugar dieciséis entre los alumnos de esa edad de los países participantes y los de catorce el lugar veinticinco entre los de la suya, un periodista concluía y publicaba muy ufano que “los niños españoles leen cada vez peor durante su progreso escolar”. Sin duda, la comprensión lectora que fallaba era la del propio periodista, que demostraba magníficamente esa necesidad de formación que aquí se reclama.

En quinto lugar, hay que tener en cuenta el lugar que ocupan los intermediarios de la información. El ejemplo anterior, más allá de lo que tiene de anecdótico, ilustra acerca del importante y ambivalente papel que desempeñan profesionales tales como los periodistas en el proceso de difusión (Popham, 1993: 291). Sin duda, los medios de comunicación de masas ejercen un indudable atractivo sobre los evaluadores, deseosos de difundir sus resultados. Pero el tratamiento que muchas veces realizan tales medios de los resultados de la evaluación, cayendo en la simplificación abusiva, el deseo de llamar la atención o la selección interesada de la información, puede resultar contraproducente. De ahí que los evaluadores y los responsables de la evaluación hayan de valorar muy cuidadosamente los canales a través de los cuales difunden su información.

En sexto lugar y último lugar, siguiendo esa línea argumental, hay que combinar diversas habilidades en el equipo responsable de efectuar la difusión. Para que ésta sea eficaz ha de mezclar dosis de rigor, precisión, accesibilidad, empatía, atractivo, por no citar sino algunas características. Pero no es siempre fácil

asegurar la presencia de tales cualidades en un equipo evaluador; en ocasiones será necesario recurrir a profesionales externos que puedan ayudar a planificar y desarrollar un proceso coherente y efectivo de difusión.

El dedicar las últimas páginas del documento al tema de la difusión no es una casualidad. En efecto, la difusión de la información obtenida en un proyecto de evaluación constituye el último impacto del mismo. Si bien la evaluación es un proceso continuo, a lo largo del cual se trata y se utiliza la información de diversos modos, la difusión de sus resultados suele constituir su punto final (a veces, punto y seguido). Y sin embargo, esa etapa última debe estar planificada desde su origen e integrarse en una estrategia general.

Al comienzo del documento se avanzaba la intención de abordar cuatro aspectos centrales relativos al tratamiento y uso de la información proporcionada por la evaluación. En este momento se llega al final del proceso, con el cierre del último apartado. Lo que era un plan general del trabajo se ha ido desarrollando hasta completarse precisamente abordando la difusión. Lo que constituye el final natural de un proceso de evaluación se convierte aquí también en el asunto final del documento, aunque esperando dejar suficientes pistas abiertas para continuar explorando caminos que apenas comienzan a recorrerse en la actualidad.

Referencias bibliográficas

Alkin, M.C., Daillak, R.H. y White, P. (1979) : *Using Evaluation : Does Evaluation Make a Difference*. Beverly Hills : Sage.

Black, P. y Atkin, J.M., eds. (1996) : *Changing the Subject. Innovations in Science, Mathematics and Technology Education*. London & New York: Routledge/OECD.

De Landsheere, G. (1994): *Le pilotage des systèmes d'éducation*. Bruxelles: De Boeck.

Degenhart, R.E., ed. (1990): *Thirty years of international research. An annotated bibliography of IEA publications (1960-1990)*. The Hague: IEA.

Department for Education (1995): *GCSE to GCE A/AS Value Added: Briefing for Schools and Colleges*. London: Department for Education.

Elley, W. (1992): *How in the world do students read?*. The Hague: IEA.

Glasman, N.S. y Nevo, D. (1988) : *Evaluation in Decision-Making. The case of school administration*. Boston, Dordrecht & Lancaster : Kluwer.

Goldstein, H. (1993): *Interpreting International Comparisons of Student Achievement*. A report prepared for UNESCO, policopiado.

Goldstein, H. y Spiegelhalter, D.J. (1996): "League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance", *Journal of the Royal Statistical Society*, A. 159, pp. 385-443.

House, E.R. (1994): *Evaluación, ética y poder*. Madrid: Morata.

Kogan, M. (1996): "Les systèmes éducatifs: Comment suivre leur fonctionnement, les contrôler et les diriger", en *Evaluer et réformer les systèmes éducatifs*. París: OCDE, pp. 27-51.

Joint Committee on Standards for Educational Evaluation (1981): *Standards for Evaluation of Educational Programs, Projects and Materials*. New York: McGraw-Hill.

MacDonald, Barry (1995): "La evaluación como servicio público: Perspectivas de futuro", en Sáez Brezmes, M.J., ed.: *Conceptualizando la evaluación en España*. Alcalá de Henares: ICE de la Universidad de Alcalá, pp. 15-23.

Mestres, J. (1995): "Reflexiones sobre la negociación en evaluación", en Sáez Brezmes, M.J., ed.: *Conceptualizando la evaluación en España*. Alcalá de Henares: ICE de la Universidad de Alcalá, pp. 37-49.

Michel, A. (1996): "La conducción de un sistema complejo: la Educación Nacional", *Revista Iberoamericana de Educación*, nº 10, pp.13-36.

Norris, N. (1990): *Understanding Educational Evaluation*. London: Kogan Page.

Popham, W.J. (1993): *Educational Evaluation*. Boston: Allyn and Bacon.

OCDE (1995): *Performance Standards in Education. In Search of Quality*. París: OCDE.

Reimers, F., McGinn, N. y Wild, K. (1995): *Confronting Future Challenges: Educational Information, Research and Decision-Making*. París: UNESCO, International Bureau of Education.

Rossi, P.H. y Freeman, H.E. (1993): *Evaluation. A Systemic Approach*. Newbury Park, London & New Delhi: Sage, 5ª ed.

Sáez Brezmes, M.J. (1995): "Caracterización y problemática de los informes de evaluación", en Sáez Brezmes, M.J., ed.: *Conceptualizando la evaluación en España*. Alcalá de Henares: ICE de la Universidad de Alcalá, pp. 83-93.

Santos Guerra, M.A. (1993): "Los (ab)usos de la evaluación", *Cuadernos de Pedagogía*, nº 215, pp. 70-73.

Shadish, W.R.; Cook, T.D. y Leviton, L.C. (1995): *Foundations of Program Evaluation. Theories of Practice*. Newbury Park, London & New Delhi: Sage.

Simons, E.H., ed. (1980): *Towards A Science of the Singular: essays about case study in educational research and evaluation*. Norwich: CARE – University of East Anglia.

Stake, R.E. (1995): *The Art of Case Study Research*. Beverly Hills & London: Sage.

Theisen, G.L., Achola, P.P.W. y Boakari, F.M. (1990): "La insuficiencia de los estudios internacionales sobre el rendimiento", en Altbach, P.G. y Kelly, G.P., eds.: *Nuevos enfoques en educación comparada*. Madrid: Mondadori, pp. 37-61.

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

Tiana, A. (1997): "Indicadores educativos. Qué son y qué pretenden", *Cuadernos de Pedagogía*, nº 256, pp. 50-53.

Thomas, S. y Mortimore, P. (1995): *Comparison of Value Added Models for Secondary School Effectiveness*, Paper presented at the annual conference of the British Educational Research Association, Bath, 14-17 September 1995.

Topolsky, J. (1985): *Metodología de la historia*. Madrid: Cátedra.