Evaluación, ética y poder

Por

Ernest R. HOUSE

Traducción de

Pablo Manzano

Colección: PEDAGOGÍA
Manuales

Scanned by CamScanner

Ernest R. HOUSE

Evaluación, ética y poder

Segunda edición

PROYECTO DE
FORMACION INICIAL
UNIVERSIDAD DE PLAYA ANCHA
DE CIENCIAS DE LA EDUCACION
SISTEMA DE BIBLIOTECAS
BIBLIOTECA CENTRAL



EDICIONES MORATA, S. L.
Fundada por Javier Morata, Editor, en 1920
C/ Mejía Lequerica, 12
28004 - MADRID

Scanned by CamScanner

Titulo original de la obra:
EVALUATING WITH VALIDITY

© Ernest R. House.
Publicada originalmente por Sage Publications Inc., 1980.

Primera edición: 1994
Reimpresión: 1997

361.6
H 842
c.1

No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros métodos, sin el permiso previo y por escrito de los titulares del Copyright.

27/10/99
fm

65984

© EDICIONES MORATA, S. L. (1997)
Mejía Lequerica, 12. 28004 - Madrid

Derechos reservados
Depósito Legal: M-16.716-1997
ISBN: 84-7112-389-4

Compuesto por: Ángel Gallardo
Printed in Spain - Impreso en España
Imprime : LAVEL. Humanes (Madrid)
Ilustración de la cubierta: Postal publicitaria de Fósforo Ferrero. Reproducida con permiso de Syntex Latino.

CONTENIDO

Págs.

SOBRE EL AUTOR 10

AGRADECIMIENTOS 12

PREFACIO 13

PRIMERA PARTE: Los enfoques de la evaluación 15

CAPÍTULO PRIMERO: **El evaluador en la sociedad** 17
Evaluación privada y evaluación pública, 18.—*El proceso de evaluación,* 20.

CAPÍTULO II: **Los enfoques principales** 23
El enfoque de análisis de sistemas, 25.—*El enfoque de objetivos conductuales (o basado en metas),* 28.—*El enfoque de decisión,* 29.—*El enfoque que prescinde de los objetivos,* 31.—*El enfoque del estilo de la crítica de arte,* 33.—*El enfoque de revisión profesional (acreditación),* 35.—*El enfoque cuasijudicial (contradictorio),* 38.—*El estudio de casos (o negociación),* 40.—*Otros enfoques,* 43.

CAPÍTULO III: **Premisas en las que se basan los enfoques** 45
Liberalismo, 45.—*Ética subjetivista,* 48.—*La epistemología objetivista liberal,* 49.—*La epistemología subjetivista,* 55.—*Premisas políticas de los modelos: los utilitaristas,* 58.—*Premisas políticas de los intuicionistas-pluralistas,* 59.

SEGUNDA PARTE: Los criterios fundamentales de la evaluación: veracidad, belleza y justicia 63

CAPÍTULO IV: **La lógica de la argumentación evaluativa** 65
El gran terremoto que se va a producir en California, 65.—*Equivocidad*

© Ediciones Morata, S. L.

Scanned by CamScanner

de las pruebas: certidumbre frente a credibilidad, 66.—*La evaluación como persuasión,* 69.— *Los destinatarios de la evaluación,* 72.—*Las premisas del acuerdo,* 75.—*La argumentación cuantitativa,* 77.—*La argumentación cualitativa,* 80.—*Revisión de la objetividad, validez e imparcialidad,* 83.—*El discurso evaluativo: la buena vida (junto a la falla de San Andrés),* 90.

CAPÍTULO V: **Coherencia y credibilidad. La estética** 93
El conductor ebrio, 93.—*La trama,* 97.—*La forma,* 101.—*"Outward Bound" (Salida),* 103.—*Autenticidad,* 106.—*La belleza del conductor ebrio,* 109.

CAPÍTULO VI: **Justicia** 112
Utilitarismo, 114.—*Pluralismo/Intuicionismo,* 118.—*Justicia como equidad,* 120.—*La utilización de las teorías de la justicia,* 126.

TERCERA PARTE: Los principios de la evaluación 129

CAPÍTULO VII: **La democratización de la evaluación** 131
La política de la elección, 131.—*Evaluación democrática liberal,* 133.—*Intereses,* 134.—*El interés público en la evaluación,* 137.—*Concentración de poder frente a difusión de poder,* 140.—*Principios distributivos,* 142.—*La evaluación como procedimiento de decisión moral,* 145.

CAPÍTULO VIII: **Un contrato de evaluación equitativo** 148
La caracterización de la equidad, 149.—*Algunas condiciones para un contrato equitativo,* 150.— *Reflexiones sobre las condiciones de equidad,* 158.—*¿Hasta qué punto es vinculante el contrato?,* 160.

CAPÍTULO IX: **Poder y deliberación** 162
Las debilidades del liberalismo, 162.—*Poder e intereses,* 164.—*Poder y autoridad,* 167.—*Deliberación,* 169.—*Determinación colectiva de las opciones,* 173.—*El fundamento moral de la evaluación,* 175.—*La concepción subyacente de justicia,* 176.

CUARTA PARTE: Metaevaluación 183

CAPÍTULO X: **La objetividad, equidad y justicia de la política federal de evaluación** 185
La política federal de evaluación, 185.—*Los comienzos del "Folow Through",* 188.—*La fase Egbert/Stanford Research Institute: 1968-1971,* 189.—*La fase McDaniels/Huron Institute: 1972-1973,* 192.—*La fase Evans/Abt Associates: 1974-1977,* 194.—*La lógica de la evaluación,* 197.—*Objetividad,* 200.—*Equidad,* 202.—*Justicia,* 204.—*Una revisión de la política federal de evaluación,* 206.

CAPÍTULO XI: **Crítica de los enfoques** 209
El Análisis de sistemas, 210.—*Objetivos conductuales,* 211.—*Decisión,* 213.—*Independencia de los objetivos del programa,* 215.—*Crítica*

© Ediciones Morata, S. L.

de arte, 218.—*Revisión profesional,* 221.—*Cuasijudicial,* 224.—*Estudio de casos,* 227.

CAPÍTULO XII: **La realización de evaluaciones válidas** 231

APÉNDICE A: **Un análisis de la lógica de una evaluación** 239
Formas de razonamiento, 239.—*Análisis de la "Educational Product Evaluation" de Glass,* 244.—*Análisis de la respuesta de Scriven a la evaluación de Glass,* 250.

APÉNDICE B: **La evaluación naturalista** 255

APÉNDICE C: **Un contrato de evaluación** 259

BIBLIOGRAFÍA 262

OTRAS OBRAS DE EDICIONES MORATA DE INTERÉS 270

© Ediciones Morata, S. L.

Scanned by CamScanner

Sobre el autor

Ernest R. HOUSE es profesor de *Administration, Higher and Continuing Education del Center for Instructional Research and Curriculum Evaluation* (CIRCE) de la *University of Illinois* en Urbana. Se graduó con el de *Bachelor* en la *Washington University* en 1959 (graduación *Phi Beta Kappa*), ejerció la docencia en educación secundaria, graduándose como *Master* en la *Southern Illinois University* en 1964 y obteniendo su doctorado en la *University of Ilinois* en 1968.

Su primer proyecto consistió en una evaluación a gran escala del *Illinois Gifted Program*. Desde entonces, se ha dedicado a evaluar diversos programas educativos y sociales. En los últimos años ha trabajado en la evaluación de evaluaciones. Con anterioridad ha publicado los siguientes libros: *School Evaluation: The Politics and Process* (1973), *The Politics of Educational Innovation* (1974) y *Survival in the Classroom* (con S. LAPAN, 1978). Ha sido consultor de muchas organizaciones de los Estados Unidos y de otros países. En 1976, fue presidente del congreso anual de la *American Educational Research Association* y en la actualidad pertenece al consejo de redacción de *Educational Evaluation and Policy Analysis*.

© Ediciones Morata, S. L.

El individuo instruido aunque imprudente, que camina en línea recta desde las verdades generales a la particulares, se pierde en medio de las tortuosas sendas de la vida. Pero el sabio que, en medio de las vueltas e incertidumbres de la acción y de los acontecimientos humanos, mantiene su mirada fija en la verdad eterna, sigue una vía indirecta, cuando no puede proseguir por la línea recta y, en el terreno de la acción, toma decisiones que, con el tiempo, se muestran tan provechosas como lo permite la naturaleza de las cosas.

Giambattista VICO.
On the Study Methods of Our Time, 1709, pág. 35.

Scanned by CamScanner

AGRADECIMIENTOS

Este libro constituye la culminación de mis cinco últimos años de trabajo. Diversas personas han aportado su contribución significativa al mismo. Gene GLASS y Rochelle MAYER lo han revisado y han influido ampliamente en las ideas que se exponen. Steve LAPAN, Don HOGBEN y Barry MACDONALD han desarrollado conmigo muchos de estos pensamientos hasta extremos, a veces, inconcebibles. Mis colaboradores del CIRCE (Gordon HOKE, Tom HASTINGS, Bob STAKE y Harry BROUDY) me han proporcionado un ambiente excepcionalmente estimulante en el que trabajar.

Mike ATKIN, decano de Illinois durante estos años, ha alentado de forma especial mis esfuerzos. Tom SERGIOVANNI, director de mi departamento, se ha mostrado muy generoso a la hora de repartir las tareas. Muchos alumnos de CIRCE han hecho lo posible para evitar que me desviase en diversas cuestiones y les agradezco su ayuda en determinados capítulos. Otra tarea ingrata, desarrollada con todo entusiasmo por Jeri WILLEN y Donna KOENIG, ha sido la de trabajar con mis manuscritos.

© Ediciones Morata, S. L.

PREFACIO

Este libro se dirige a los evaluadores y a quienes deseen adquirir un conocimiento más profundo de la evaluación moderna, qué es y adónde se dirige. Su objetivo consiste en lograr que la práctica de la evaluación sea más reflexiva. Para ello, el libro presenta una conceptualización global de la evaluación moderna, que va desde los enfoques principales hasta las normas y principios que deben presidir esa práctica, y la crítica de aquellos enfoques sobre la base de tales normas.

El panorama actual de la evaluación se caracteriza por la vitalidad y el desorden. La escala a la que se llevan a cabo las actividades de evaluación, su omnipresencia y diversidad hacen difícil su comprensión, incluso a quienes se mueven en este campo. Más alarmante aún es el hecho de que una mala evaluación puede deteriorar un programa social y ocasionar perjuicios a toda una clase social. La importancia social de la evaluación es enorme; su interpretación, relativamente insignificante. Esperamos que este libro contribuya a un orden conceptual más completo y evidente, y a un sentido más decidido de responsabilidad moral.

La *Primera parte* del libro se ocupa de la caracterización de las modernas tentativas de evaluación. En el Capítulo Primero se describen las formas actuales de evaluar, mostrando cómo, en un contexto moderno, la evaluación no sólo difiere de los intentos precedentes en cuanto al ámbito que abarca, sino también en sus características esenciales. En el Capítulo II, afirmamos que hay relativamente pocos enfoques de la evaluación, a pesar del esfuerzo masivo que se lleva a cabo, describiéndolos con detalle. Establecemos una taxonomía de los enfoques fundamentales de la evaluación. En el Capítulo III, relacionamos entre sí de forma sistemática los distintos enfoques fundamentales a partir de sus supuestos básicos subyacentes de carácter epistemológico, ético y político, explicando dichos supuestos. Al caracterizar los enfoques preponderantes en la actualidad, damos por supuesto que los evaluadores actúan de manera racional, o sea, que tienen buenas razones

© Ediciones Morata, S. L.

Scanned by CamScanner

para hacer lo que hacen, aunque ello les lleve a adoptar enfoques diferentes.

La *Segunda parte* del libro consiste en una reflexión sobre los niveles de verdad, belleza y justicia necesarios para orientar y juzgar la calidad de una evaluación. La veracidad en la evaluación tiene más implicaciones que en el caso de la verdad en la ciencia, al menos eso se dice. La belleza, que se manifiesta en la coherencia y la imaginación en la evaluación, tiene una influencia sustancial en la credibilidad que otorgue un público concreto a la evaluación. Por último, el sentido de la justicia con el que actuemos tiene importantes consecuencias para el tipo de evaluación que llevemos a cabo y para los resultados de la misma. Revisaremos las concepciones de la justicia y sus consecuencias sobre la evaluación.

La *Tercera parte* del libro trata de extraer los principios políticos y morales sobre los que debe basarse la evaluación. Se la concibe como un aspecto de los procedimientos más generales de decisión social. Se limita a la persuasión racional. Estudiamos qué significa el que una evaluación sea aceptable desde los puntos de vista democrático y moral. En sentido más práctico, afirmamos que, para considerar equitativa una evaluación, hace falta que se cumplan determinadas condiciones. Por último, indicamos que el fundamento subyacente de la evaluación consiste en valores morales considerados en el marco de una concepción pluralista de la justicia.

La *Cuarta parte* del libro constituye una crítica de la política federal de evaluación y de sus principales enfoques. Cada uno tiene sus virtudes y sus defectos, tanto desde el punto de vista teórico como del práctico, pero ninguno tiene valor universal. El libro acaba con una exposición del carácter de la validez en la evaluación sosteniendo que, para que ésta se considere válida, ha de ser veraz, creíble y correcta, desde el punto de vista normativo. El objetivo del libro consiste en conseguir que la práctica sea más autorreflexiva.

© Ediciones Morata, S. L.

PRIMERA PARTE

Los enfoques de la evaluación

> *El primer paso para que los hombres se comprendan consiste en hacer que reconozcan el modelo o modelos que dominan y penetran su pensamiento y acción. Como todos los intentos de hacer conscientes a los hombres de las categorías con las que piensan, se trata de una actividad difícil y, a veces, dolorosa, que produce con facilidad resultados profundamente inquietantes. La segunda tarea consiste en analizar el modelo en sí, lo que compromete al analista respecto en su aceptación, modificación o rechazo y, en este último caso, en la elaboración de otro más adecuado que lo sustituya.*
>
> Isaiah BERLIN.
> *Philosophy, Politics and Society,* 1962, pág. VII.

Scanned by CamScanner

CAPÍTULO PRIMERO

El evaluador en la sociedad

Separar lo malo de lo bueno es, desde muy antiguo, una actividad humana, esencial para la misma sociedad. Aunque a menudo se ha concedido autoridad a las instituciones para hacer juicios de valor, sólo en los últimos tiempos se ha formalizado la evaluación de los programas públicos de manera que constituya una actividad independiente como tal e incluso una nueva disciplina.

En los últimos veinte años, la evaluación se ha convertido en una actividad importante al surgir los programas de bienestar a gran escala. Cada año, se llevan a cabo decenas de miles de evaluaciones de programas públicos, sólo en los Estados Unidos. Miles de personas están empleadas en estas evaluaciones y cientos de universidades y empresas compiten para conseguir contratos de evaluación. La mayor parte de las evaluaciones están patrocinadas por el gobierno federal; otras lo están por los gobiernos de los estados y los ayuntamientos. La evaluación de los programas públicos se ha convertido en una actividad exigida por la legislación y cada año se invierten en ella cientos de millones de dólares.

Tanto por su omnipresencia, como por el papel que desempeña, la evaluación tiene gran relieve social, y transforma, justifica o desacredita los programas públicos. La enorme cantidad de las actividades de evaluación hacen difícil la supervisión. No siempre las evaluaciones de los programas públicos son de buena calidad. Con excesiva frecuencia, los evaluadores se limitan a hacer lo que quieren sus patrocinadores. Demasiado a menudo confunden el carácter de su trabajo y no hacen justicia a los programas sociales que evalúan. No se trata de que los evaluadores sean más incompetentes o avariciosos que las personas que se dedican a otras ocupaciones, sino de que las posibilidades de causar daños suelen ser más amplias, menos evidentes y más perdurables.

Este libro pretende fomentar la reflexión en la evaluación. Parece necesario basar la evaluación en alguna forma de responsabilidad moral, de

© Ediciones Morata, S. L.

Scanned by CamScanner

manera que las reflexiones sobre su justicia, veracidad e, incluso, belleza configuren su práctica. Deben existir consideraciones morales que trasciendan el marco de los cientos de organizaciones y miles de evaluadores que tratan de atraer la atención de unos pocos patrocinadores. Sin embargo, estas consideraciones no pueden derivarse o imponerse de forma arbitraria, sino desarrollarse a partir de un análisis racional de lo que es la evaluación y de lo que parece razonable a los que la realizan como fundamento de una práctica consciente.

Evaluación privada y evaluación pública

La evaluación moderna es descendiente directa del modernismo. La modernización era la liberación de la tradición, el paso de una realidad incontrovertida ofrecida por la tradición, a un contexto social en el que todo es controvertible y mutable; era el paso de "lo dado" a "la opción" (BERGER, 1974). En el mundo moderno, podían escogerse más cosas. Se suponía que esto traería una vida material mejor y un desarrollo personal más profundo. En realidad, la modernización ha llegado a tal extremo que la misma opción se ha convertido en un problema. La evaluación aparece como elemento de ayuda para poder elegir.

En época más reciente, la realización de evaluaciones formales a escala masiva, el patrocinio brindado por los gobiernos a las evaluaciones y la proliferación de organismos que las efectúan han planteado problemas nunca vistos antes. Pensemos en una mujer que elige por su cuenta y riesgo. Acude a una tienda de automóviles para comprar un coche nuevo. Visita los concesionarios cuyos vehículos puede adquirir y mira los coches disponibles. Tras un regateo con el vendedor respecto del precio, escoge el coche pequeño más barato. Le gustan los coches pequeños y no quiere tener que esperar un par de meses hasta que se lo entreguen. Para ella son importantes el precio de compra, el tamaño y la posibilidad de entrega inmediata. No le preocupan demasiado ni el color ni los accesorios. Lleva a cabo su propia evaluación para su satisfacción particular.

Pensemos ahora en una situación en la que el exceso de opciones posibles abruma a la persona en cuestión. Hay tal cantidad de vehículos disponibles que no sabe ni por dónde empezar y, además, el comprador no es experto en mecánica. Una organización, como la *Consumers' Union**, le ofrece sus servicios: evalúa en su nombre los coches y le dice cuál le conviene más. El interesado puede conseguir sus informes de evaluación de automóviles. Ahora bien, en este caso, el evaluador es distinto de la persona que decide. Ambas situaciones difieren en varios aspectos.

¿Cómo debe proceder el evaluador? Para que la evaluación sea útil, ha

* El equivalente en España sería la Organización de Consumidores y Usuarios (OCU). *(N. del E.)*

© Ediciones Morata, S. L.

de seleccionar normas de valoración adaptadas a los gustos y necesidades de la interesada. Si no la conoce, es difícil que adivine sus preferencias personales. Pasará por alto la forma externa; la interesada puede hacer tal elección sin ayuda. La entrega inmediata es un problema circunstancial. El evaluador también lo dejará de lado, salvo como factor limitador. Por supuesto, a todo el mundo le interesa el precio de compra. Además, el evaluador tendrá en cuenta en su valoración la economía de consumo de combustible y otros aspectos por el estilo, dejando de lado la velocidad máxima como criterio.

El evaluador tiene que tomar otras decisiones importantes también. ¿Qué coches ha de evaluar? Cuando la mujer efectuaba la evaluación por su cuenta, definía el conjunto de vehículos que le serviría de base para elegir. Para el evaluador no es una cuestión sin importancia, porque los resultados variarán de modo espectacular según los coches que incluya en su comparación. Tiene que seleccionar un conjunto que evaluar. Aplicará las normas por él elegidas al conjunto de coches. En la medida en que se ajusten mejor o peor a las normas, los calificará como buenos o malos (lo que supone una comparación con todos o con el "coche medio") o clasificará los automóviles del mejor al peor, dependiendo de cómo haga la comparación en el conjunto seleccionado. Por último, elaborará su informe.

En ese momento, la consumidora tiene que optar de nuevo: debe decidir si acepta o no la evaluación. Ha de cumplir determinados requisitos: ¿tomaría ella como punto de partida los vehículos preseleccionados? ¿Las normas utilizadas son pertinentes respecto a su decisión? Al emitir sus juicios finales, ¿el evaluador sopesa las normas empleadas como lo hubiera hecho ella misma? Y todavía más importante: ¿es veraz la evaluación?, ¿puede fiarse de ella?, ¿es fiable el evaluador?, ¿le paga alguna de las empresas automovilísticas?, ¿tiene credibilidad la evaluación? La consumidora sigue teniendo libertad para aceptar o rechazar los resultados y consejos de la evaluación. Aunque la evaluación se haya convertido en un proceso social que establece un vínculo entre el evaluador y la consumidora, la elección final le compete a ésta.

Pensemos ahora en una tercera situación. El gobierno financia a un evaluador para que valore un programa de acción social. En este caso, ¿para quién se realiza la evaluación?: ¿para el gobierno?, ¿para el público?, ¿para los beneficiarios del programa?, ¿para los administradores del mismo? La respuesta no es sencilla porque los resultados afectan a todos estos grupos. Sin embargo, aquí no existe un individuo concreto, como en el caso de la compradora de coches, a quien pueda dirigirse la evaluación. Hay muchas personas interesadas, porque las decisiones sobre un programa basadas en la evaluación afectarán a todos. La evaluación forma parte, de manera fundamental e inextricable, de una situación *pública:* una decisión colectiva.

A diferencia del contexto privado, el consumidor insatisfecho no puede conformarse con dejar de lado la evaluación después de encontrarla inaceptable porque ésta le afectará de alguna manera. El carácter público de

© Ediciones Morata, S. L.

Scanned by CamScanner

manera que las reflexiones sobre su justicia, veracidad e, incluso, belleza configuren su práctica. Deben existir consideraciones morales que trasciendan el marco de los cientos de organizaciones y miles de evaluadores que tratan de atraer la atención de unos pocos patrocinadores. Sin embargo, estas consideraciones no pueden derivarse o imponerse de forma arbitraria, sino desarrollarse a partir de un análisis racional de lo que es la evaluación y de lo que parece razonable a los que la realizan como fundamento de una práctica consciente.

Evaluación privada y evaluación pública

La evaluación moderna es descendiente directa del modernismo. La modernización era la liberación de la tradición, el paso de una realidad incontrovertida ofrecida por la tradición, a un contexto social en el que todo es controvertible y mutable; era el paso de "lo dado" a "la opción" (BERGER, 1974). En el mundo moderno, podían escogerse más cosas. Se suponía que esto traería una vida material mejor y un desarrollo personal más profundo. En realidad, la modernización ha llegado a tal extremo que la misma opción se ha convertido en un problema. La evaluación aparece como elemento de ayuda para poder elegir.

En época más reciente, la realización de evaluaciones formales a escala masiva, el patrocinio brindado por los gobiernos a las evaluaciones y la proliferación de organismos que las efectúan han planteado problemas nunca vistos antes. Pensemos en una mujer que elige por su cuenta y riesgo. Acude a una tienda de automóviles para comprar un coche nuevo. Visita los concesionarios cuyos vehículos puede adquirir y mira los coches disponibles. Tras un regateo con el vendedor respecto del precio, escoge el coche pequeño más barato. Le gustan los coches pequeños y no quiere tener que esperar un par de meses hasta que se lo entreguen. Para ella son importantes el precio de compra, el tamaño y la posibilidad de entrega inmediata. No le preocupan demasiado ni el color ni los accesorios. Lleva a cabo su propia evaluación para su satisfacción particular.

Pensemos ahora en una situación en la que el exceso de opciones posibles abruma a la persona en cuestión. Hay tal cantidad de vehículos disponibles que no sabe ni por dónde empezar y, además, el comprador no es experto en mecánica. Una organización, como la *Consumers' Union**, le ofrece sus servicios: evalúa en su nombre los coches y le dice cuál le conviene más. El interesado puede conseguir sus informes de evaluación de automóviles. Ahora bien, en este caso, el evaluador es distinto de la persona que decide. Ambas situaciones difieren en varios aspectos.

¿Cómo debe proceder el evaluador? Para que la evaluación sea útil, ha

* El equivalente en España sería la Organización de Consumidores y Usuarios (OCU). *(N. del E.)*

© Ediciones Morata, S. L.

de seleccionar normas de valoración adaptadas a los gustos y necesidades de la interesada. Si no la conoce, es difícil que adivine sus preferencias personales. Pasará por alto la forma externa; la interesada puede hacer tal elección sin ayuda. La entrega inmediata es un problema circunstancial. El evaluador también lo dejará de lado, salvo como factor limitador. Por supuesto, a todo el mundo le interesa el precio de compra. Además, el evaluador tendrá en cuenta en su valoración la economía de consumo de combustible y otros aspectos por el estilo, dejando de lado la velocidad máxima como criterio.

El evaluador tiene que tomar otras decisiones importantes también. ¿Qué coches ha de evaluar? Cuando la mujer efectuaba la evaluación por su cuenta, definía el conjunto de vehículos que le serviría de base para elegir. Para el evaluador no es una cuestión sin importancia, porque los resultados variarán de modo espectacular según los coches que incluya en su comparación. Tiene que seleccionar un conjunto que evaluar. Aplicará las normas por él elegidas al conjunto de coches. En la medida en que se ajusten mejor o peor a las normas, los calificará como buenos o malos (lo que supone una comparación con todos o con el "coche medio") o clasificará los automóviles del mejor al peor, dependiendo de cómo haga la comparación en el conjunto seleccionado. Por último, elaborará su informe.

En ese momento, la consumidora tiene que optar de nuevo: debe decidir si acepta o no la evaluación. Ha de cumplir determinados requisitos: ¿tomaría ella como punto de partida los vehículos preseleccionados? ¿Las normas utilizadas son pertinentes respecto a su decisión? Al emitir sus juicios finales, ¿el evaluador sopesa las normas empleadas como lo hubiera hecho ella misma? Y todavía más importante: ¿es veraz la evaluación?, ¿puede fiarse de ella?, ¿es fiable el evaluador?, ¿le paga alguna de las empresas automovilísticas?, ¿tiene credibilidad la evaluación? La consumidora sigue teniendo libertad para aceptar o rechazar los resultados y consejos de la evaluación. Aunque la evaluación se haya convertido en un proceso social que establece un vínculo entre el evaluador y la consumidora, la elección final le compete a ésta.

Pensemos ahora en una tercera situación. El gobierno financia a un evaluador para que valore un programa de acción social. En este caso, ¿para quién se realiza la evaluación?: ¿para el gobierno?, ¿para el público?, ¿para los beneficiarios del programa?, ¿para los administradores del mismo? La respuesta no es sencilla porque los resultados afectan a todos estos grupos. Sin embargo, aquí no existe un individuo concreto, como en el caso de la compradora de coches, a quien pueda dirigirse la evaluación. Hay muchas personas interesadas, porque las decisiones sobre un programa basadas en la evaluación afectarán a todos. La evaluación forma parte, de manera fundamental e inextricable, de una situación *pública:* una decisión colectiva.

A diferencia del contexto privado, el consumidor insatisfecho no puede conformarse con dejar de lado la evaluación después de encontrarla inaceptable porque ésta le afectará de alguna manera. El carácter público de

© Ediciones Morata, S. L.

Scanned by CamScanner

la situación impone ciertas limitaciones a la evaluación. Al cabo del tiempo, quedan magnificados todos los problemas relacionados con la selección de normas apropiadas y la elección de comparaciones adecuadas. Sin embargo, los evaluadores tienen que escoger y han descubierto distintos modos de habérselas con este problema. Los enfoques adoptados por ellos —de ahí el moderno predicamento del evaluador— constituyen el tema inicial de este libro. La mayor o menor adecuación de estos enfoques es su final. Entretanto, trataré de establecer las normas de evaluación que conduzcan a estos juicios de adecuación.

El proceso de evaluación

En su sentido más sencillo, la evaluación conduce a una opinión fundada de que algo es de un cierto modo. No tiene por qué llevar a una decisión respecto a una determinada forma de actuar, aunque hoy día a menudo pretende tal cosa. Podemos valorar a César como un gran general sin que de ello se derive ninguna decisión. La evaluación aboca a un juicio acerca del valor de algo.

Con frecuencia, se llega a tal juicio mediante la calificación o clasificación de algo según cumpla mejor o peor un conjunto de normas o criterios. La evaluación es comparativa por naturaleza, y suele presentarse explícitamente como tal. Esto significa que ha de existir un conjunto de normas y una clase con la que comparar el objeto. Si éste se califica como "bueno" o "malo", la clase de comparación está constituida por toda la clase de objetos o por el objeto promedio de la clase. Si el objeto recibe la clasificación de "mejor" o "peor", se le compara directamente con un subconjunto concreto de objetos similares; por ejemplo, un coche se compara con otros cuatro en cuanto a su precio. Los coches se clasifican por orden, indicando el nivel relativo de cada uno, en relación con esta norma y en comparación con los otros cuatro, aunque sin necesidad de asignar una valoración al grupo de coches como tal. El mero hecho de seleccionar una clase de comparación puede provocar una diferencia espectacular en la evaluación, aunque las normas utilizadas sean las mismas (TAYLOR, 1961).

Por tanto, la evaluación supone, por naturaleza, adoptar un conjunto de normas, definirlas, especificar la clase de comparación y deducir en qué grado el objeto satisface las normas. Cumplimentadas estas etapas, el evaluador debe poder llegar a establecer un juicio sobre el valor del objeto evaluado (TAYLOR, 1961). Muchos enfoques, como los de los estudios de casos, presentan las normas, comparaciones y juicios de forma más implícita e intuitiva. Algunos tienden más a la comprensión que al juicio.

Es frecuente que el juicio de valor se exprese como "bueno" o "malo", o mediante otras fórmulas valorativas, pero no hace falta que así sea. En determinados contextos, pueden utilizarse expresiones descriptivas ordinarias para expresar juicios de valor. Los ejemplos y anécdotas pueden transmitir

© Ediciones Morata, S. L.

mensajes de evaluación. El juicio de valor estriba en la forma de utilizar las expresiones más que en estas mismas.

Los juicios de valor que se desprenden de las evaluaciones no equivalen a directrices. Sólo podemos orientar a alguien de manera razonable para que haga algo si está en sus manos hacerlo. Podemos hacer juicios de valor sobre cosas cuya modificación se nos escapa. Sin embargo, en la práctica moderna de la evaluación, se pide a menudo al evaluador que ayude al responsable de las decisiones cuando trata de decidir entre distintos cursos generales de acción. Éstos se transforman en los objetos evaluados. Los juicios de valor pueden convertirse en recomendaciones, pero, aun en ese caso, el responsable de la decisión goza de libertad para dejar de lado el consejo del evaluador.

En el caso más sencillo de evaluación, en el que evaluador y responsable de la decisión es la misma persona, es más fácil seleccionar la clase de comparación y adoptar un conjunto de normas. Sin embargo, el proceso real de evaluación es más complejo de lo que parece. Las personas no sólo evalúan mediante la aplicación de normas, sino comprobando también sus juicios en situaciones particulares cuyos resultados creen conocer. El razonamiento real de evaluación se desarrolla como un proceso dialéctico entre principios abstractos y casos concretos más que como una deducción directa a partir de las normas (BARRY, 1965).

Es más, en una evaluación real, las normas pueden ser contradictorias. Queremos un coche amplio y también un automóvil que consuma poca gasolina. No podemos conseguir el máximo respecto de ambas características de manera simultánea. En estos casos, nos inclinamos al equilibrio entre ellas, planteándonos cuánta amplitud queremos en relación con el consumo que deseamos. En vez de dar una preferencia absoluta a una u otra norma, optamos por determinar la inclinación por normas contradictorias mediante consideraciones intuitivas de las proporciones de ambas. ¿Qué amplitud estamos dispuestos a sacrificar en beneficio de cuánta economía de consumo? Estas proporciones relativas varían de una persona a otra. La evaluación es un proceso complejo aun cuando se concibe como una operación personal.

Cuando evaluador y responsable de las decisiones son personas distintas, se añade otra dimensión. Está en juego la credibilidad del evaluador y, además, en ese caso ha de comunicarse de alguna manera la evaluación al responsable de la decisión. Existe una mayor incertidumbre respecto a la clase de comparación y las normas que emplear y a las proporciones relativas que asignar a las normas. El responsable de la decisión no sólo ha de preocuparse de que los datos sean correctos, sino también de si el evaluador los presenta de forma veraz. En las sociedades liberales modernas, esta relación se concibe aún como un acto privado, aunque interpersonal.

Por último, en la tercera situación, el objeto sometido a evaluación es un programa público. El evaluador debe preocuparse de que la evaluación satisfaga las normas de los procedimientos de decisión previstas en aras

© Ediciones Morata, S. L.

Scanned by CamScanner

del interés público. La evaluación no sólo es ya un asunto interpersonal, sino colectivo. Debe tenerse en cuenta una comunidad más amplia de intereses pertinentes. Éste es el problema fundamental del evaluador moderno y es lo que echa sobre sus espaldas la carga más pesada de todas. Su evaluación no sólo debe ser veraz y creíble, también debe ser justa.

Los evaluadores han enfocado la solución de su moderno problema de diversos modos. Sus soluciones han evolucionado a partir de sus convicciones filosóficas, profesionales y personales. Su preparación y práctica habitual constituyen influencias significativas. En el próximo capítulo presentamos los principales enfoques de la evaluación y, en el siguiente, los ponemos en relación con los sistemas de creencias.

© Ediciones Morata, S. L.

CAPÍTULO II

Los enfoques principales

Se defienden docenas de enfoques de la evaluación. No obstante, la mayoría de ellos puede agruparse en unos pocos tipos básicos. Hay quienes denominan "modelos" a estos tipos básicos, para poner de manifiesto que se trata de diseños o tipos estructurales, defendidos por importantes teóricos como enfoques dignos de imitación. Son paradigmas. Podemos llamarlos "enfoques", sin más.

Todos estos modelos básicos tienen importantes partidarios, presentan unos fundamentos racionales explícitos, aparecen en las bibliografías, hay un grupo de profesionales que los ponen en práctica y —con no menos relieve— existen evaluaciones concretas llevadas a cabo según sus correspondientes diseños básicos. En sí, los modelos constituyen elaboraciones idealizadas de los enfoques de evaluación. Son muchas las contingencias distintas que configuran una evaluación real; por tanto, ésta puede adoptar muchas formas aunque comience, desde el punto de vista conceptual, como un tipo concreto. En otras palabras, un modelo es un tipo ideal.

Conviene señalar también que muchos teóricos eminentes de la evaluación (CRONBACH, CAMPBELL y GLASS, por ejemplo) no se comprometen con ningún enfoque concreto. Y otros, por ej., GUBA, han publicado trabajos que defienden dos enfoques diferentes. Yo he llevado a cabo evaluaciones utilizando todos los enfoques principales. En consecuencia, no podemos identificarlos como propios de ninguna persona determinada ni considerarlos típicos de nadie.

La Figura 1 muestra una taxonomía de los principales modelos de evaluación. Al definir éstos, he utilizado con profusión otras clasificaciones anteriores, sobre todo la de STAKE (1976), así como las de POPHAM (1975) y WORTHEN y SANDERS (1973). Sin duda, habré omitido sin querer algunos tipos. He excluido algunos posibles modelos porque aún no se han utilizado lo suficiente para que constituyan una escuela de práctica profesional. Los

© Ediciones Morata, S. L.

Scanned by CamScanner

Figura 1: *Taxonomía de los principales enfoques de la evaluación*

Modelo	*Principales destinatarios o grupos de referencia*	*Concuerda en*	*Metodología*	*Producto*	*Cuestiones típicas*
Análisis de sistemas	Economistas, directivos	Objetivos, causa y efecto conocidos, variables cuantificadas	PPBS, programación lineal, variación planificada, análisis coste/producto	Eficiencia	¿Se han alcanzado los efectos previstos? ¿Pueden lograrse los mismos efectos de manera más económica? ¿Cuáles son los programas más eficientes?
Objetivos conductuales	Directivos, psicólogos	Objetivos especificados de antemano, variables de producto cuantificadas	Objetivos conductuales, tests de resultados finales	Productividad, responsabilidad	¿El programa alcanza los objetivos? ¿Es productivo el programa?
Decisión	Ejecutivos, en especial, administradores	Objetivos generales, criterios	Encuestas, cuestionarios, entrevistas, variación natural	Eficacia, control de calidad	¿Es eficaz el programa? ¿Qué partes son eficaces?
Sin objetivos definidos	Clientes	Consecuencias, criterios	Control de sesgos, análisis lógico, *modus operandi*	Elección del cliente, utilidad social	¿Cuáles son *todos* los efectos?
Crítica de arte	Entendidos, clientes	Crítica, normas, niveles	Revisión crítica	Normas mejores, mayor consciencia	¿Resiste el programa la crítica? ¿Aumenta el aprecio del público?
Revisión profesional	Profesionales y público	Criterios, tribunal, procedimientos	Revisión a cargo del tribunal, estudio	Aceptación profesional	¿Cómo clasificarían los profesionales este programa?
Cuasi-jurídico	Jurado	Procedimientos y jueces	Procedimientos cuasijurídicos	Resolución	¿Qué argumentos hay en pro y en contra del programa?
Estudio de casos	Clientes, profesionales	Negociaciones, actividades	Estudios de casos, entrevistas, observaciones	Comprensión de la diversidad	¿Qué les parece el programa a distintas personas?

© Ediciones Morata, S. L.

modelos de la Figura 1 constituyen los principales enfoques que se utilizan en la actualidad.

A continuación, expongo un breve esquema de cada uno. En el Capítulo XI, haré una crítica de cada enfoque, desde un punto de vista tanto teórico como práctico.

El enfoque del análisis de sistemas

En este enfoque, se definen unas pocas medidas de resultados, como las puntuaciones de tests en educación, tratando de relacionar las diferencias halladas entre programas o normativas con las variaciones que se descubran en los indicadores. Los datos son cuantitativos y las medidas de resultados se relacionan con los programas mediante análisis de correlacion u otras técnicas estadísticas. Recientemente, vienen utilizándose más los diseños experimentales, prefiriéndose la "variación planificada" del programa a la "natural" (véase COOLEY y LOHNES, 1976).

Uno de los antecedentes principales de este enfoque fue el análisis de sistemas desarrollado en el *Department of Defense,* siendo secretario MCNAMARA. Desde 1965, más o menos, constituyó la principal perspectiva sobre la evaluación en el *Department of Health, Education and Welfare* (Departamento de Salud, Educación y Bienestar). En ese año, el presidente Johnson amplió el *Planning, Programming and Budgeting System* (PPBS) (Sistema de planes, programas y presupuestos) en el ámbito del gobierno federal. Siendo secretario de *Health, Education and Welfare* John GARDNER, se creó un nuevo organismo denominado *Assistant Secretary for Program Evaluation* (ASPE) (Subsecretaría de Evaluación de Programas). Para hacerse cargo del mismo, fue nombrado un grupo de economistas y auditores, dirigido por William GORHAM, que contaba con gran experiencia en el PPBS del Departamento de Defensa. Los ayudantes de GORHAM eran Robert GROSSE y Alice RIVLIN, quien se convertiría más adelante en subsecretaria y directora de la *Congressional Budget Office* (Oficina de presupuestos del Congreso).

> El modelo de gestión, promovido y perfeccionado por Robert MCNAMARA en el Departamento de Defensa y llevado al de Salud, Educación y Bienestar por William GORHAM, se derivaba del mundo de la teoría microeconómica. Los analistas de la Subsecretaría de Evaluación de Programas trataron de aplicar perspectivas teóricas semejantes a las áreas de servicios humanos. En el diseño de TEMPO *(la primitiva evaluación del Título I)* había diversos supuestos implícitos de partida que configuran una visión económica de cómo funciona el mundo. El presupuesto básico consiste en que los individuos y las organizaciones se comportan de manera que se eleve al máximo algún resultado o conjunto perceptible de resultados. Por tanto, el analista debe ser capaz de representar las opciones organizativas y deducir los objetivos deseados, así como la eficacia relativa de las distintas estrategias para alcanzarlos. Este marco analítico supone la existencia de una función estable de producción, una relación regular y

© Ediciones Morata, S. L.

Scanned by CamScanner

cuantificable entre los factores iniciales de una actividad y los productos de la misma.

(McLAUGHLIN, 1975, pág. 35.)

La historia de la evaluación del Título I de la ESEA (Ley de Educación especial), que financiaba a los niños en situación desventajosa, es la de una tentativa de utilizar el enfoque del análisis de sistemas. La *Office of Education** resolvió que se utilizara este enfoque en los 30.000 proyectos del Título I, que afectaban a cinco millones de niños. Dos funcionarios del Departamento de Salud, Educación y Bienestar explican el enfoque:

> Para cumplir este mandato legislativo, la *U. S. Office of Education* contrató a la *RMC Research Corporation* para que elaborase modelos para evaluar los incrementos de la adquisición de destrezas básicas conseguidos por los proyectos del Título I. Cada modelo de los tres de RMC incluye tests pre y postratamiento, un método de estimación de incrementos de la "línea base" en ausencia de los servicios del Título I y procedimientos de conversión de los resultados de los tests locales en equivalentes de "curva normal" que permitieran la integración entre proyectos.
>
> (BARNES y GINSBURG, 1979, pág. 7.)

Por tanto, estas evaluaciones prescritas utilizan como única medida de éxito las puntuaciones en los tests (aunque también se recogieran la duración semanal de los servicios, las proporciones entre profesores y alumnos, los gastos por cada niño y la cantidad total de participantes), preocupándose más por la máxima consecución de éxitos y por la integración de resultados que por la distribución de los mismos. Todos los resultados debían ser comunicados en un "equivalente de curva normal" elaborado al efecto, e integrados en el nivel estatal y en el nacional. Podemos presumir que una puntuación podría representar el estado o la nación.

En *Systematic Thinking for Social Action,* decía RIVLIN (1971) que el problema fundamental consistía en cómo podría el gobierno tomar decisiones mejores. Esta autora consideraba que las cuestiones clave eran: 1) ¿cómo definimos los problemas sociales y cómo se distribuyen?; 2) ¿a quién y cuánto ayudarían las soluciones propuestas?; 3) ¿qué sería más beneficioso y cómo podrían compararse los beneficios derivados de los distintos enfoques?; 4) ¿cómo pueden proporcionarse los servicios del modo más eficaz? De estas dos últimas preguntas podría ocuparse la evaluación. La característica más importante de este enfoque es el análisis comparativo coste/producto de los programas. El caso paradigmático es la evaluación de seguimiento.

En general, existe un consenso respecto a los objetivos, y el problema

* La *U.S. Office of Education,* integrada en el *Department of Health Education and Welfare,* correspondería a lo que en el organigrama español es una secretaría de estado o subsecretaría ministerial de educación. *(N. del T.)*

© Ediciones Morata, S. L.

estriba en medir los resultados de los programas gubernamentales. Con el fin de establecer las relaciones de causa a efecto entre programas y resultados, hace falta un buen diseño experimental. El diseño preferible incluye un grupo de control escogido al azar, pero no siempre es factible. Una vez medidos los resultados —de modo psicométrico, sociométrico o econométrico—, se comparan los costes de los programas para averiguar qué resultado puede obtenerse por menos dinero. La metodología de las ciencias sociales se basa, en gran parte, en este proceso.

El libro *Evaluation: A Systematic Approach,* de ROSSI, FREEMAN y WRIGHT (1979) hace suyo el enfoque del análisis de sistemas. Para ellos, las evaluaciones se efectúan con fines de gestión, planificación, desarrollo normativo y efectos fiscales. Las cuestiones clave son las siguientes:

1. ¿Alcanza la intervención a la población objetivo?
2. ¿Está implementándose del modo especificado?
3. ¿Es eficaz?
4. ¿Cuánto cuesta?
5. ¿Cuál es su razón coste/eficacia? (pág. 20).

La evaluación debe ser "lo más objetiva posible; es decir, para proporcionar una *valoración firme:* una valoración cuyos resultados no varíen si la realiza otro grupo o si la repiten los mismos evaluadores" (pág. 21). Se pretende conseguir "el desarrollo e implementación ordenados de normas". Las evaluaciones sistemáticas se basan en las técnicas de las ciencias sociales. ROSSI y cols. desestiman las evaluaciones de sentido común porque conducen a conclusiones erróneas, y las evaluaciones basadas en reglas y normas profesionales porque no se fundan en pruebas científicas. Tampoco el juicio clínico proporciona una orientación "firme" para la decisión.

Las evaluaciones sistemáticas proporcionan pruebas válidas y fiables que pueden repetir otros observadores, que no podrían haberse producido sin la interacción y que incluyen información respecto al empleo eficaz de los fondos. Una evaluación global da respuesta a cuestiones relativas a la planificación y supervisión de programas, valoración de su impacto y eficiencia económica. La información sobre la planificación ayuda a los planificadores a poner a punto intervenciones adecuadas. La información de supervisión nos dice si un programa "concuerda con su diseño". La información sobre el impacto nos muestra si el programa provoca cambios en el sentido pretendido. La información económica pone de manifiesto si el programa es eficiente. ROSSI, FREEMAN y WRIGHT advierten que no es posible llevar a cabo la evaluación de un modo competente si el proceso de desarrollo normativo no se efectúa de forma clara y explícita.

En todo caso, el enfoque del análisis de sistemas adopta, sin lugar a duda, la metodología de la ciencia social positivista y, por regla general, excluye otras metodologías. Este enfoque cuenta con el respaldo de muchos altos funcionarios del gobierno federal, sobre todo economistas. Además, reclama para sí la categoría de ciencia social.

© Ediciones Morata, S. L.

Scanned by CamScanner

Como señalamos en el capítulo anterior, el problema que se plantea al evaluador moderno consiste en que debe valorar un programa público para unos destinatarios externos. El enfoque del análisis de sistemas resuelve este problema adoptando la perspectiva de los legisladores federales, dando por supuesto que el programa que se somete a investigación constituye una parte funcional de la estructura social y gubernamental. Si asignamos ese papel al programa, éste debe observar determinados criterios y funciones. Por tanto, el evaluador ha de responder a ciertas cuestiones, aceptando que el programa desempeña ese papel. Se trata de un enfoque funcionalista. Dada la función de los programas, todos ellos deben cumplir ciertos requisitos.

El enfoque de objetivos conductuales (o basado en metas)

El enfoque por objetivos resuelve de otra manera el dilema del evaluador moderno. Toma las metas del programa tal como están formuladas y recaba pruebas para comprobar si se han alcanzado. Los objetivos son la única fuente de normas y criterios. El evaluador valora lo que dicen que tratan de conseguir quienes han elaborado el programa. La discrepancia entre los objetivos formulados y los resultados obtenidos constituye la medida del éxito del programa. Los objetivos formulados descargan al evaluador de la tarea de suponer determinadas funciones de los programas, por ejemplo, la eficiencia.

En educación, el modelo basado en objetivos fue promovido por TYLER (1950), que defendía la definición de los objetivos educativos en términos de la conducta de los estudiantes. La evaluación de un programa debería definir sus resultados y sus objetivos como conductas individuales específicas. El cometido del evaluador consistiría en determinar si los estudiantes mostraban estas conductas concretas tras participar en el programa. Este procedimiento acabó conociéndose como enfoque por "objetivos conductuales" o modelo tyleriano de elaboración y evaluación de programas.

MAGER (1962) perfeccionó más tarde la técnica, insistiendo en que no sólo había que especificar de antemano los objetivos en términos conductuales, sino que los mismos objetivos encerraban los criterios y niveles de consecución de las metas previstas por quienes elaboraran el programa. Los medios empleados para medir la conducta de los alumnos consistían en variables de logros cuantificadas, sobre todo los tests de rendimiento en educación. BLOOM (1956) y otros elaboraron taxonomías de objetivos educativos apropiados. BLOOM y cols. (1971) y POPHAM (1975) aplicaron este enfoque a la evaluación de distintas asignaturas.

En los últimos años, el centro de atención de la técnica se ha desplazado desde la adecuada formulación de los objetivos a la preocupación por el modo de medirlos. Los tests normalizados de rendimiento tradicionales consisten en ítems de prueba que no se basan en objetivos especificados de

© Ediciones Morata, S. L.

antemano. Ha surgido un marcado interés por la elaboración de tests "referidos a un criterio". Invariablemente, se refieren a objetivos. O sea, los creadores de los tests establecen un conjunto de objetivos, basándose en ellos para elaborar los ítems del test.

El *National Assessment of Educational Progress,* cuyo principal defensor es TYLER, es de este tipo. Los tests de competencia constituyen un desarrollo más reciente de esto. A menudo, se parte de la base de que los objetivos que sirven de fundamento al test son objetivos mínimos correspondientes a un nivel determinado. En otras palabras, todos los niños tienen que llegar a dominarlos. Para que se le declare mínimamente competente, el alumno tiene que responder bien todos estos ítems. Casi todos los estados cuentan con un programa de tests de competencia de este tipo. En algunos, como en Florida, el paso al curso superior depende de la puntuación que el alumno consiga en el test. El enfoque por objetivos goza de mayor popularidad en el nivel estatal.

El enfoque por objetivos conductuales también ha adquirido relieve en otros campos. Por ejemplo, la gestión por objetivos consiste esencialmente en el enfoque por objetivos aplicado a los negocios y a los organismos gubernamentales. Se pide a organismos e individuos que definan sus objetivos y se les juzga con arreglo a la medida en que los cumplen. Nada tiene de particular si tenemos en cuenta que el movimiento a favor de los objetivos surge del análisis de tareas y de la psicología industrial de principios del siglo XX.

El primer libro de evaluación de SUCHMAN (1967), sobre la evaluación de la salud pública se basaba primordialmente en el enfoque por objetivos. El proceso de evaluación se concebía como la identificación de una actividad meta, como la puesta en marcha de esa actividad, la valoración del efecto de la actuación meta, la formación de valores, la formulación de objetivos y la medida de los mismos (pág. 34). En palabras de SUCHMAN: "La característica más peculiar de la investigación evaluativa es la presencia de alguna meta u objetivo, cuya medida de consecución constituye el principal centro de atención del problema que se va a investigar" (pág. 37). Un objetivo claro de un programa equivale a la hipótesis en un trabajo de investigación.

Podemos contrastar este enfoque con el antes descrito del análisis de sistemas. La evaluación de seguimiento, según el enfoque del análisis de sistemas, se basaba sólo en cuatro baterías de tests para evaluar trece programas relativos a la primera infancia. Es de suponer que lo que median estos cuatro tests era fundamental. En realidad, cada programa de primera infancia tenía docenas de metas y objetivos explícitos. Una evaluación por objetivos conductuales tendría que haber determinado si se cumplía cada uno de los objetivos o bien debería reducir éstos a algún conjunto mensurable.

El enfoque de decisión

Todos los enfoques modernos de la evaluación tienen presente la conexión entre ésta y la toma de decisiones, aunque varíen los responsables de

© Ediciones Morata, S. L.

Scanned by CamScanner

las decisiones y el modo en que éstas se llevarán a efecto. El enfoque de decisión sostiene que la evaluación ha de estructurarse a partir de las decisiones reales que haya que tomar, lo que alude, por regla general, a las que corresponden al responsable máximo, director, o administrador.

En el campo de la educación, el principal exponente de este enfoque ha sido STUFFLEBEAM (1973). "La evaluación es el proceso de delimitar, obtener y proporcionar información útil para juzgar posibles decisiones alternativas" (pág. 129). STUFFLEBEAM define tres ámbitos de decisión (homeostasia, incrementalismo y neomovilismo), cuatro tipos (planificación, estructuración, implementación y reciclado), tres fases en el proceso de evaluación (delimitación, obtención y comunicación) y cuatro tipos de evaluación [de contexto, de entrada *(input)*, de proceso y de producto].

Con independencia del tipo de evaluación de que se trate, su diseño se establece mediante la identificación del nivel de decisión al que se dirige, proyectando la situación de decisión, definiendo los criterios de cada situación y las normas para el evaluador. Tras ello, se recoge, organiza, analiza y transmite la información requerida. Cada fase se divide en diversas tareas subordinadas del evaluador (STUFFLEBEAM, 1969).

GUTTENTAG (1973, 1974) trató de realizar un análisis cuantitativo de la estructura preferida por el responsable de la decisión. Su idea consiste en identificar las acciones o programas que eleven al máximo relieve los valores de los responsables de las decisiones. Esto se llevó a cabo mediante un método de escalado de utilidad de atributos múltiples, elaborado por EDWARDS. La "utilidad" es el resultado de un curso de acción relativo a los valores de los responsables de las decisiones. Su objeto consiste en identificar los cursos de acción que favorecen al máximo la utilidad global.

El enfoque exige que se definan las dimensiones pertinentes de valor y que el responsable de la decisión clasifique las entidades que han de evaluarse en relación con estas escalas de valor. Todo ello se basa en los juicios subjetivos de los responsables de las decisiones. La utilidad de las alternativas de acción se resume en relación con las diversas dimensiones de valor para determinar el curso de acción de máxima utilidad.

PATTON (1978), que hace hincapié en el factor personal en la utilización de los hallazgos de la evaluación, presenta la tercera perspectiva respecto al enfoque de decisión de la evaluación. "El primer paso en la perspectiva que se centra en la utilización de la evaluación consiste en la identificación y organización de los responsables pertinentes de la decisión y de los usuarios de la información que aporte la citada evaluación" (pág. 61). PATTON sostiene que los hallazgos de la evaluación se utilizan cuando algún individuo se responsabiliza de forma directa y personal de conseguir la información para las personas adecuadas. La responsabilidad de determinar qué información necesitan corresponde a los encargados de tomar las decisiones.

El segundo paso de una evaluación consiste en identificar y centrar las cuestiones pertinentes. La cuestión correcta tiene varias características. Es posible aportar datos que se refieran a ella y existe más de una posible

© Ediciones Morata, S. L.

respuesta a la misma. Los responsables de las decisiones identificados quieren y creen necesitar información que les ayude a responder la pregunta. Los encargados de las decisiones quieren contestarse a sí mismos y les preocupan las respuestas. Por último, los responsables de las decisiones pueden especificar cómo utilizarán la respuesta a la cuestión.

En resumen, el enfoque de decisión de la evaluación resuelve el problema del evaluador tomando como destinatario (a quien se dirige la evaluación) al responsable de las decisiones, adoptando como preocupaciones y criterios significativos los propios de aquél. Varía la manera de identificar a los responsables de las decisiones y cómo se manifiestan éstas. Se supone que este enfoque incrementará la utilización de los hallazgos de la evaluación.

La metodología descansa, en gran medida, en la metodología de encuesta, como cuestionarios o entrevistas, y el evaluador trabaja más sobre las variaciones naturales en el medio en donde se desarrolla el programa que tratando de montar experimentos. Las cuestiones que se resuelven corresponden a los responsables de las decisiones, aunque, por regla general, incluyen la eficacia del programa en relación con alguna dimensión y, en especial, qué partes del programa funcionan.

El enfoque que prescinde de los objetivos

El enfoque que prescinde de los objetivos constituye una reacción directa frente a la omnipresencia de la evaluación determinada por los objetivos. SCRIVEN (1973) sostiene que el evaluador no sólo no tiene que basar su evaluación en los objetivos del programa, sino que debe procurar evitar deliberadamente informarse sobre ellos, con el fin de que no le lleven a sesgos tendenciosos. La evaluación sin objetivos debe contemplarse en el contexto de la preocupación más general de SCRIVEN por la reducción de los efectos de los sesgos en la evaluación. Los objetivos producen sesgos en el evaluador. Como dice SCRIVEN, "por desgracia, yo no puedo evaluar sus logros (del personal del proyecto) con una objetividad razonable y, al mismo tiempo, tener una larga sesión de adoctrinamiento con ellos" (página 320).

El modelo sin objetivos reduce el sesgo de buscar sólo los intentos del responsable del programa, especificados de antemano, ya que éstos no se comunican al evaluador. Por tanto, el evaluador debe investigar todos los resultados. Muchos de ellos no son sino efectos colaterales imprevistos, que pueden ser positivos o negativos. Como dice SCRIVEN:

> A menudo, los llamados "efectos colaterales", sean buenos o malos, determinan completamente el resultado de la evaluación. Es obvio que, para el evaluador, carece de relieve si se trata de efectos "colaterales" o "directos". Esa forma de hablar se refiere a las intenciones del promotor y el evaluador no valora intenciones, sino resultados. En realidad, es peligroso incluso prestar oídos a descripciones generales de las intenciones porque esto aparta la atención

© Ediciones Morata, S. L.

Scanned by CamScanner

de los "efectos colaterales" y tiende a hacer que los pasemos por alto o los infravaloremos.

(Pág. 321.)

De todos los enfoques de la evaluación aquí descritos, el que prescinde de los objetivos ha sido el menos utilizado, hasta el punto de que algunos cuestionan que se le incluya entre los modelos principales. En el campo de los servicios sociales, es frecuente que, tanto a los evaluadores como a los responsables de los programas, les resulte difícil imaginar incluso dónde puede el evaluador encontrar criterios de evaluación fuera de los objetivos del responsable del programa. Dan por supuesto que el evaluador se limita a reemplazar sus propios criterios por los correspondientes a los responsables de los programas.

Sin embargo, la *Consumers' Union* evalúa productos de consumo con regularidad sin preguntar a los productores cuáles son sus metas. Por ejemplo, este organismo evalúa automóviles sin utilizar los objetivos de la *General Motors* como base de la evaluación. La *Consumers' Union* emplea las normas y criterios que cree benefician más al consumidor que al productor. Del mismo modo, la evaluación sin objetivos toma como destinatarios o grupo de referencia a los beneficiarios de los programas sociales, en vez de a los directivos. En gran parte, el pensamiento de SCRIVEN parece tener presente como modelo el enfoque de la *Consumers' Union*.

Por supuesto, los programas sociales no son automóviles y la obtención de un conjunto de normas para su evaluación no resulta tan sencilla como en el caso de los productos de consumo. Con este objeto, SCRIVEN ha tratado de desarrollar el concepto de "necesidad" como fundamento de la evaluación. Frente a los simples gustos o deseos, las necesidades se descubren mediante su evaluación. El establecimiento de la validez objetiva de tales necesidades proporcionaría al evaluador una fuente autorizada de normas. Por tanto, la evaluación independiente de los objetivos se basa en el análisis de las necesidades del usuario y no de los objetivos de los productores.

La metodología de la evaluación sin objetivos no es tan clara como en otros enfoques. En general, SCRIVEN es partidario de un enfoque "sin sesgos" para garantizar la objetividad. Se considera al evaluador como experto en la tarea de eliminar e impedir que los sesgos derivados de sí mismo y de los demás intervengan en la evaluación. Entre las técnicas utilizadas están los experimentos de tipo "doble ciego", en los que ni el sujeto ni el experimentador saben cuál es el tratamiento ni cuál el placebo. SCRIVEN ve la evaluación sin objetivos como una operación a "triple ciego", en la que nadie conoce las intenciones. Otras técnicas abarcan la mezcla antecedente y posterior de los elementos que se juzgan, la calibración de los jueces por medio de casos normalizados, la separación entre el individuo y el personal del programa, códigos deontológicos, réplica concurrente de las experiencias, análisis secundario a cargo de agentes independientes y otras por el estilo (SCRIVEN, 1976).

© Ediciones Morata, S. L.

SCRIVEN (1975) insiste mucho en la independencia entre el evaluador y el personal del programa. Su principio de retroalimentación (*feedback*) independiente dice que ninguna unidad dependerá por completo de una determinada subunidad en relación con la retroalimentación evaluativa sobre esa subunidad. Su principio de la inestabilidad de la independencia da a entender que, en el transcurso del tiempo, los evaluadores se ven continuamente solicitados, debiendo cuidarse de mantener una renovada vigilancia sobre la evaluación. Se refiere de forma reiterada al evaluador como a un cazador o detective que ha de descubrir claves y extraer la información. En un caso, compara el trabajo del evaluador con el de un juez instructor que trata de determinar conexiones causales en condiciones no experimentales (SCRIVEN, 1976). Por medio del método del *modus operandi*, el investigador establece una cadena causal entre causa y efecto. El evaluador debe operar de modo semejante.

Las técnicas utilizadas en las pocas evaluaciones llevadas a cabo prescindiendo de los objetivos son muy diversas. En una evaluación de ciertos materiales curriculares nuevos, HOUSE y HOGBEN (1974) examinaron los resultados de un pretest, leyeron revisiones de los materiales efectuadas por expertos, visitaron aulas, entrevistaron a profesores y alumnos, y examinaron los materiales concretos. En otra evaluación de materiales curriculares, WELCH (1978) utilizó un tribunal de jueces para clasificar los materiales en una larga lista de comprobación. HARRINGTON y SANDERS (1979) tratan de elaborar algunas orientaciones para la realización de la evaluación, prescindiendo de los objetivos.

El enfoque del estilo de la crítica de arte

Los roles que desempeñan la crítica de arte, la crítica literaria, la crítica teatral y la crítica de cine son bien conocidos. Constituyen formas valiosas de juzgar la calidad de las obras de arte. EISNER (1979) y otros han estudiado el modelo análogo de crítica educativa o curricular para juzgar los programas educativos. El mismo enfoque puede utilizarse en otras áreas del ámbito social.

En la concepción de EISNER, la crítica es cualitativa por naturaleza, como las mismas obras de los artistas. "El crítico o crítica de arte se enfrenta con la difícil tarea de traducir las cualidades esencialmente inefables que constituyen las obras de arte a un lenguaje que ayude a los demás a percibir la obra con mayor profundidad" (pág. 191). EISNER sostiene que la crítica es una operación empírica. Las cualidades que describe el crítico pueden situarse en la obra misma. Es más, todo puede ser objeto de crítica. Ésta no es la valoración negativa de algo, sino la iluminación de sus cualidades, de modo que pueda apreciarse su valor. La prueba de la crítica está en los efectos instrumentales que produce en su destinatario.

EISNER distingue entre la actuación del experto y la crítica. La característica del experto es el arte de la apreciación, mientras que la crítica es el arte de

© Ediciones Morata, S. L.

Scanned by CamScanner

la revelación (pág. 193). El arte del experto consiste en reconocer y apreciar las cualidades de la obra en cuestión, pero no exige un juicio o descripción públicos. Es el preludio necesario de la crítica.

El carácter de experto requiere del perceptor que tenga gran experiencia para ser capaz de distinguir lo que de significativo tiene un vino, un juego, un aula educativa. Sin embargo, existen diferencias entre reconocimiento y percepción. El simple reconocimiento y la clasificación basados en la experiencia no bastan. Hace falta ser capaz de percibir las sutilezas y comprender cómo contribuyen a un patrón general, como en la vida del aula.

Para desarrollar este carácter de experto es preciso tener la oportunidad de asistir a los acontecimientos y compararlos. Las percepciones han de perfeccionarse, integrarse y valorarse. Hay que ser sensible a las cualidades que surgen del aula o del programa que se va a evaluar. Esto requiere también un conjunto de ideas (teorías, modelos o lo que sea) que permitan al sujeto diferenciar lo significativo de lo trivial. Para acceder a estas discriminaciones hace falta una preparación adecuada y también experiencia.

Por otra parte, "la crítica es el arte de revelar las cualidades de los hechos u objetos que percibe el arte del experto" (pág. 197). La crítica trata de "traducir" una situación de tal manera que se pongan en evidencia los aspectos significativos, de la situación, objeto o programa. EISNER plantea aquí la distinción entre las formas discursiva y no discursiva de presentación. La crítica trata de captar la esencia de lo concreto. Su lenguaje es no discursivo en la medida en que utiliza formas para presentar sensaciones más que para representar hechos.

La crítica realizada de manera adecuada aumenta el conocimiento y la apreciación. "Apreciación" no significa que guste algo, sino un mayor conocimiento. La función de la crítica consiste en aplicar criterios, de manera que los juicios sobre los hechos puedan fundarse en una consideración de lo importante. Esa investigación cualitativa permite al lector participar de manera vicaria en las cualidades descritas.

En cuanto a la validez y fiabilidad, EISNER habla de la "corroboración estructural" y de la "adecuación referencial". La evidencia se corrobora cuando los elementos de la misma encajan entre sí. La adecuación referencial alude a la posibilidad de encontrar en la obra o programa criticado las características señaladas por el crítico. Se hace gran hincapié en la necesidad de un buen conocimiento de la situación. No obstante, no hay que esperar la unanimidad entre los críticos. Según EISNER, el objetivo consiste en ampliar la percepción y no en llegar a una crítica definitiva. La consecuencia de la crítica se basa en el desarrollo del carácter de expertos de los demás. Hasta ahora, el enfoque crítico de la evaluación ha sido divulgado sobre todo por EISNER y sus discípulos (p. ej.: MCCUTCHEON, 1978, y VALLANCE, 1978), limitándose a la educación.

En realidad, hay diferentes enfoques respecto a la crítica misma. KELLY (1978) establece una analogía entre la evaluación y la crítica literaria, más que con la crítica de arte. De la literaria extrae conceptos establecidos como: metáfora, voz, argumento y tema, utilizándolos para ilustrar la evaluación. Su análisis muestra cómo los evaluadores emplean conceptos similares al considerar los programas y redactar sus informes. En su análisis, se mezclan descripción y evaluación. Asimismo, la credibilidad de la crítica constituye un problema para él más que para EISNER. Como señala KELLY, aceptamos la crítica con la que estamos de acuerdo. Por tanto, la naturaleza de la crítica puede variar de una materia a otra.

© Ediciones Morata, S. L.

La crítica puede cambiar también dentro de una misma área. JENKINS y O'TOOLE (1978) señalan las diferencias entre críticos literarios. Existen los críticos académicos de las universidades, que juzgan todo según los niveles de las mejores obras. Hay críticos escritores que perciben la crítica como una consecuencia de su propia obra y tratan de determinar las normas de gusto mediante las que podrían ser evaluados. En particular, la crítica de estos últimos asume algunas características de la escritura creativa. Por último, se encuentran los críticos libres o periodísticos, que se dirigen a un público amplio.

Ante todo, el crítico da por supuesto que escribe para un grupo de usuarios o para expertos. El público descubrirá si el programa es bueno, pero también tendrá una sensibilidad acrecentada. Esto supone que el público acepta la crítica y, por regla general, sus normas implícitas. No existe una metodología normalizada, salvo la revisión crítica, que puede llevarse a cabo de diversos modos. La inmersión y la familiaridad con el objeto o programa suelen considerarse fundamentales. Para el crítico, son esenciales una experiencia y preparación adecuadas. En definitiva, la crítica lleva a una consciencia reforzada, una mejora de los niveles y una actuación perfeccionada. El enfoque crítico de la evaluación resuelve el dilema de ésta mediante la designación, como evaluador, de una persona experimentada y con la adecuada preparación, en definitiva, un experto en la cuestión de que se trate. Con su sensibilidad, la persona experimentada comunica sus juicios de un modo que informa y educa a los menos sensibles.

El enfoque de revisión profesional (acreditación)

Durante más de cincuenta años, las asociaciones profesionales han llevado a cabo evaluaciones de la preparación profesional. Ya se trate de médicos, abogados, trabajadores sociales o maestros, la revisión profesional de los centros de formación, tal como la realiza la *North Central Association* en las escuelas secundarias, ha constituido un medio fundamental de evaluación. Concebida de manera algo más amplia, la revisión profesional supone que los cirujanos, profesores o abogados juzguen el trabajo de sus colegas. Por regla general, la participación en la revisión se circunscribe a los compañeros de profesión.

Muchas de ellas son evaluaciones que sirven para dar fe de la calidad de los centros de preparación profesional. Por ejemplo, a partir de 1885, ciertos grupos de escuelas secundarias organizaron la normalización de los requisitos de ingreso en los centros de enseñanza superior. De ello

© Ediciones Morata, S. L.

Scanned by CamScanner

surgió el *National Study of School Evaluation* (1978), que publica *Evaluative Criteria*. En este trabajo aparecen criterios mediante los cuales se evalúan las escuelas secundarias.

Los criterios están organizados por materias específicas, como agricultura, arte, formación empresarial, etcétera. Algunos criterios se refieren a los servicios escolares, como medios audiovisuales de aprendizaje, servicios a los alumnos o instalaciones de la escuela. El programa educativo se divide en dieciocho áreas temáticas. Además, existen nueve áreas de servicios. Hay diversas listas de comprobación para evaluar cada materia.

Por ejemplo, en el área de inglés, aparecen ítems como: "Se exige a todos los alumnos que cursen la asignatura de inglés" (pág. 103); "Algunas obras contemporáneas y clásicas, seleccionadas por su adecuación a las capacidades, intereses y madurez de los alumnos, se estudian de manera intensiva" (pág. 195); "Se presta atención al perfeccionamiento de las técnicas de estudio" (pág. 111). El evaluador puntúa cada ítem en una escala de cinco opciones, desde "nada en absoluto" hasta "excelente".

Este enfoque está pensado para involucrar al profesorado en la autoevaluación. Antes de que una comisión externa efectúe una visita, el profesorado del centro se divide en varias subcomisiones que se encargan de puntuar los ítems de la lista de comprobación. Los resultados se comunican a todo el profesorado. El organismo externo encargado de la certificación (p. ej.: la *North Central Association*) nombra una comisión externa encargada de validar la autoevaluación. Los centros tienen derecho a vetar a los miembros de dicha comisión, cuyo número puede oscilar entre diez y treinta.

La comisión visitadora se divide en subcomisiones por materias, cada una de las cuales comprueba las puntuaciones de la autoevaluación, modificándolas si no están de acuerdo con ellas. Tanto las subcomisiones como el pleno de la comisión visitadora elaboran informes escritos. Antes de abandonar el centro, en un breve informe oral al profesorado, la comisión indica los principales aspectos positivos y negativos y hace las correspondientes recomendaciones. Dependiendo del informe, se concede o no la certificación a la escuela. Tras la visita, se prevé que ésta adopte las medidas pertinentes para corregir los aspectos negativos señalados. Aunque existen ciertas variaciones de criterios y procedimientos, diversos organismos con capacidad para otorgar certificaciones, como asociaciones profesionales y gobiernos estatales, efectúan visitas similares a casi todas las escuelas, desde las más elementales hasta las de carácter más profesional.

Una variante interesante del enfoque de acreditación de la evaluación, que registra un rápido crecimiento, consiste en la evaluación de departamentos o unidades de una universidad, patrocinada, por regla general, por la administración universitaria. Una de las primeras evaluaciones de este tipo fue la *Council on Program Evaluation* (COPE), de la Universidad de Illinois. Desde 1972 hasta 1979, fueron evaluadas todas las unidades académicas del *campus* de Urbana.

Para llevar a cabo este proceso, se instituyó un consejo de nueve profesores, presidido por un administrador del rectorado. Más tarde, entraron

© Ediciones Morata, S. L.

a formar parte del consejo dos alumnos graduados y dos no graduados. El consejo tenía autoridad para evaluar los departamentos de acuerdo con seis criterios:

1. calidad del programa docente;
2. calidad de la investigación, de la actividad creativa o del trabajo académico;
3. calidad de servicio a la universidad y a la profesión;
4. contribución prestada por la unidad a otras del *campus* o importancia de aquélla en el mismo;
5. valor del programa para la sociedad o singularidad del mismo en el estado, y
6. posibilidades y previsiones de futuro.

Al principio, el consejo nombraba grupos de trabajo de cinco a diez profesores para que estudiasen un departamento concreto. Dichos grupos celebraban amplias entrevistas con los miembros del departamento en cuestión, redactando más tarde un informe confidencial de evaluación del mismo que elevaban al consejo. Este organismo revisaba el informe y, basándose en él, elaboraba su propia evaluación abreviada, normalmente en forma de carta, que enviaba al vicerrector y al departamento. Este último podía responder a la carta y a las recomendaciones que contuviese.

Después de emplear este método durante unos años, este proceso empezó a exigir tal cantidad de tiempo que se decidió que no hubiese un grupo de trabajo por cada departamento. Se elaboró un amplio conjunto de formularios para que cada uno de ellos los cumplimentase. Éstos recogían cuestionarios sobre la mayoría de los aspectos de las operaciones de los departamentos. El consejo revisaba los formularios cumplimentados, emitiendo un juicio global y una serie de recomendaciones fundadas en criterios formales. En la práctica, la calidad de la investigación pesaba mucho en el juicio final. Cuando parecía que el departamento en cuestión presentaba demasiados aspectos negativos o necesitaba algún tipo de atención especial, se nombraba un grupo de trabajo para que llevase a cabo una evaluación presencial más exhaustiva (PETRIE, 1979).

Por supuesto, no todas las revisiones profesionales se desarrollan del mismo modo. Por ejemplo, los tribunales de evaluación de las especialidades médicas están regidos por un consejo de profesionales, pero el médico se somete a un examen escrito y otro oral a cargo de especialistas. En todas estas evaluaciones, la fuente de normas y criterios son los profesionales mismos. Se da por supuesto que están mejor cualificados para juzgar el mérito profesional. Los procedimientos varían, pero la evaluación culmina en una valoración global de un programa profesional a cargo de otros expertos.

© Ediciones Morata, S. L.

Scanned by CamScanner

*El enfoque cuasijudicial (de contrapruebas)**

Los procedimientos cuasijudiciales se han utilizado desde hace mucho tiempo con fines de evaluación y de elaboración de normativas. Recientemente, en el campo de la evaluación, se viene prestando atención a la simulación de juicios u otros procedimientos de contrapruebas, sobre todo ante la agresiva acción de los tribunales de justicia en la determinación de problemas de los programas sociales. Sin embargo, no todos los procedimientos cuasijudiciales son por naturaleza de contrapruebas. De hecho, si contamos entre los procedimientos de tipo judicial los jurados de premios y las comisiones nombradas con fines parecidos, la mayoría no reviste caracteres de contrapruebas. La legislación [p. ej.: la *Civil Rights Act* (Ley de Derechos Civiles)] establece procedimientos de revisión administrativa que se atienen a los principios procesales, aunque no son de contrapruebas.

Por ejemplo, la *Warren Commission*, que investigó el asesinato del presidente Kennedy, fue un grupo nombrado por el presidente que tomó declaraciones a los testigos, llevó a cabo sus propias investigaciones y llegó a una conclusión sobre los hechos probables. Del mismo modo, la *Kerner Commission on Civil Disorders* llegó a una conclusión sobre causas probables. Con frecuencia, se emplea este tipo de tribunales para tratar cuestiones controvertidas y resolver las dudas que pesen sobre ellas. La comisión de WIRTZ (1977) constituye un ejemplo reciente del descenso de las puntuaciones en los tests normalizados registrado en todo el ámbito nacional. El tribunal correspondiente dedujo las probables causas de tal descenso.

En Inglaterra, se hace frecuente uso de este tipo de comisiones y consejos para extraer orientaciones para la política pública (KOGAN y PACKWOOD, 1974). El *Plowden Committee* de 1967, sobre la educación primaria, y el *Taylor Report*, más reciente, sobre el gobierno local de las escuelas constituyen buenos ejemplos. Para presidir estos tribunales, se nombra a ciudadanos relevantes, contando con la ayuda gubernamental.

Todas las áreas sociales han recibido orientaciones de esta clase de tribunales y a menudo se instituye el tribunal para resolver problemas en el seno de una colectividad profesional. Por ejemplo, la *National Education Association* (HOUSE y cols., 1974) financió un tribunal para que examinara el sistema de responsabilidad educativa de Michigan. El mismo grupo patrocinó una investigación similar en Florida (TYLER y cols., 1978). En la *Office of Education*, el *Joint Review Dissemination Panel* escucha los testimonios de prueba y los comunica si los programas desarrollados con financiación federal resultan lo bastante "validados" como para divulgarlos por el resto del país.

* El término original es "Adversary", cuya traducción literal sería "adversario" o "contrario". Dado el contexto —cuasijurídico— en el que aparece, optamos por la acepción "contrapruebas", que es como en nuestro ámbito jurídico se conoce este vocablo. *Contrapruebas*, en Derecho Procesal Civil, son las pruebas encaminadas a destruir los asertos alegados por la parte contraria, forman parte de lo que se denomina *"prueba de lo contrario"* o *"prueba en contrario"*. *(N. del T.)*

© Ediciones Morata, S. L.

Aunque estos tribunales y comisiones no forman parte del sistema judicial, en realidad, su autoridad se deriva, en parte, de sus procedimientos cuasijudiciales. Las pruebas suelen presentarse ante un tribunal. Los "testigos" testifican y presentan "pruebas". Se establecen reglas respecto a quién puede presentarlas y en qué condiciones ha de darse testimonio. Las mismas audiencias adoptan los símbolos y la solemnidad de las salas de justicia; a menudo, las salas de audiencia están dispuestas a la manera de aquéllas. Suele ponerse mucho cuidado en tomar declaraciones exhaustivas. En resumen, el problema del evaluador se resuelve mediante el empleo de procedimientos cuasijudiciales que prestan a la evaluación la autoridad de la justicia.

Desde no hace mucho tiempo, vienen utilizándose procedimientos judiciales de contrapruebas para evaluar programas. La audiencia de contrapruebas se basa en el supuesto de que los hechos de un caso pueden comprobarse mejor si cada parte hace todo lo posible, con plena parcialidad, para presentar las pruebas más favorables en apoyo de sus tesis ante el tribunal (OWENS, 1973). La tendenciosidad de las partes en conflicto pone en evidencia pruebas vitales.

En 1970, tuvo lugar en Hawai la primera audiencia de contrapruebas para juzgar si las escuelas públicas de allí debían adoptar el *curriculum Man: A Course of Study*. Presidían tres jueces, todos ellos educadores. Las dos partes pasaron varias semanas preparando el caso. Otra simulación judicial se celebró en la universidad de Indiana para evaluar un programa de formación de profesorado (WOLF, 1975). Los equipos opuestos entre sí presentaron argumentos en pro y en contra del programa. Durante dos días, declararon treinta y dos testigos. Se elaboraron instrucciones judiciales respecto al examen directo y cruzado. El jurado estuvo constituido por trece expertos en educación.

El responsable del diseño del juicio dijo: "quizá, la razón de mayor peso para utilizar la metodología judicial sea que presenta un sistema útil de reglas y procedimientos de prueba dirigidos a conseguir inferencias alternativas a partir de los datos antes de proceder al juicio... El enfoque judicial facilita la consideración estructurada de argumentos e inferencias alternativos que hagan que la evaluación sea intelectualmente honrada y equitativa" (WOLF, 1975, pág. 185).

WOLF concibió el proceso constituido por cuatro fases: planteamiento del problema, selección del mismo, presentación de argumentos y audiencia. En la primera fase, se llevaba a cabo una serie de entrevistas para descubrir treinta problemas. En la segunda, se utilizaba una encuesta dirigida a profesores, estudiantes y administradores con el fin de reducir el número de problemas. En la tercera fase, se preparaban los alegatos. En la cuarta, tenían lugar las sesiones de presentación anterior a la audiencia y ella misma. La duración total del proceso fue de seis meses.

Ante las críticas dirigidas contra los métodos de contrapruebas, WOLF (1979) trató de quitar importancia al carácter de contrapruebas de su enfoque concreto: "quizá el aspecto más significativo de la metáfora judicial... sea que la forma de investigación de los tribunales es educativa" (WOLF, 1979, pág. 21). El

© Ediciones Morata, S. L.

Scanned by CamScanner

testimonio humano es la "piedra de toque de la evidencia". "La evidencia final que orienta la deliberación y el juicio no sólo abarca los 'hechos', sino una gran variedad de percepciones, opiniones, sesgos y especulaciones, en un contexto de valores y creencias" (pág. 21).

STENZEL (1979) estudió el empleo de las comisiones de investigación del Congreso como modelo de evaluación. La *Illinois Office of Education* instituyó una comisión para examinar los aspectos positivos y negativos de la evaluación de un programa de educación especial a cargo de terceros. La comisión recibió el encargo de estudiar la validez de la evaluación, revisar las recomendaciones contenidas en el informe y elaborar por escrito una respuesta al informe original de evaluación. La comisión estaba compuesta por tres miembros del personal profesional de la *Illinois Office of Education*. Los evaluadores comparecieron ante la comisión para responder a las preguntas suscitadas. STENZEL (1979) destaca los procedimientos de las comisiones del Congreso, que difieren significativamente de los de un juicio.

Hay muchos procedimientos cuasijudiciales que pueden aplicarse a la evaluación, de los que el jurado de certámenes y el juicio sólo son los más conocidos. El objetivo de los procedimientos judiciales consiste en resolver de uno u otro modo los problemas. La mayoría de los procedimientos judiciales se caracterizan por alguna forma de audiencia en la que se presentan los argumentos en pro y en contra. La legitimidad de este enfoque depende de la aceptación que se dispense a los procedimientos empleados en concreto, al instructor y al jurado o tribunal. En el ámbito de la misma administración de justicia, la acción de los tribunales suele seguir los derroteros marcados exclusivamente por los jueces.

El estudio de casos (o negociación)

Este enfoque se centra en los mismos procesos del programa y en la visión que las personas tienen de éste. La cuestión principal que se plantea es: "¿qué les parece el programa a los sujetos que tienen contacto habitual con él?". La metodología usual consiste en llevar a cabo entrevistas con muchas personas, hacer observaciones *in situ* y presentar los hallazgos en forma de estudio de casos. El estudio cualitativo de casos prevalece hasta tal punto en cuanto metodología que, para referirme al enfoque, utilizo esta misma expresión junto a la de "negociación". También podría llevarse a cabo un estudio cuantitativo de casos, pero se apartaría mucho de este enfoque. El de estudio de casos-negociación es casi completamente cualitativo por metodología y presentación. Quizá la denominación de "historia de casos" fuera más exacta.

El objetivo de este enfoque consiste en mejorar la comprensión que de la evaluación tenga el lector o destinatarios, mostrándoles, ante todo, cómo perciben otros el programa sometido a evaluación. Los procedimientos antropológicos e, incluso, las herramientas periodísticas y novelísticas son

© Ediciones Morata, S. L.
DAD DE PLAYA ANCHA

análogos. La metodología etnográfica constituye un enfoque relacionado con el que nos ocupa y es habitual calificar estas metodologías como "naturalistas". Un personaje de cabecera es STAKE (1975, 1978).

Este autor defiende así los estudios de casos:

> Con frecuencia, los estudios de casos serán el método preferible de investigación porque están en armonía epistemológica con la experiencia del lector, por lo que, para esa persona, constituirán una base natural para la generalización... Si los lectores de nuestros informes son las personas que pueblan nuestros hogares, escuelas, gobiernos e industrias, y si tenemos que ayudarles a entender los problemas y programas sociales, hemos de percibirlos y comunicarlos de forma que se adapten a sus conocimientos actuales. En su mayoría, esas personas han adquirido sus conocimientos a través de la experiencia directa y vicaria.
>
> (STAKE, 1978, pág. 5.)

Cuando la meta de la investigación es más la comprensión que la explicación y el conocimiento proposicional, el estudio de casos es a menudo superior a otras formas de investigación. Es legítimo descubrir o validar leyes por medio de muchos estudios académicos, pero el objetivo de las artes prácticas es hacer cosas. Para STAKE, los estudios de casos presentan descripciones complejas, holísticas y que implican gran cantidad de variables interactivas. Los datos se obtienen mediante la observación personal y la redacción es informal, narrativa, empleando a menudo citas, ilustraciones, alusiones y metáforas. Las comparaciones son más implícitas que explícitas.

El enfoque de STAKE se centra en la percepción y en el conocimiento como proceso "de negociación", por lo que a veces se le denomina "modelo de negociación". En la evaluación "responsiva"*, el evaluador suele negociar con el cliente sobre lo que ha de hacerse, y "responde" a lo que quieren saber distintos destinatarios.

> La evaluación educativa es responsiva si se orienta de forma más directa hacia las actividades del programa que a sus intenciones, si responde a las peticiones de información de los destinatarios y si se tienen en cuenta las diferentes perspectivas de valor presentes al informar sobre el éxito o el fracaso del programa. Para hacer una evaluación responsiva, el evaluador concibe un plan de observaciones y negociaciones. Prepara la situación de manera que diversas personas observen el programa y, con su ayuda, elabora breves informes narrativos, descripciones, exposiciones de resultados, gráficos, etc. Pone de manifiesto lo que sus destinatarios consideran valioso y recaba manifestaciones de valor de individuos cuyos puntos de vista difieran. Por supuesto, comprueba

* Aunque la traducción de la palabra inglesa *"responsive"* no es, en sentido estricto, "responsiva", sino "sensible (al contexto)", en este caso, esa "sensibilidad" se manifiesta en la disposición a adaptar la evaluación a lo que desean conocer diversos destinatarios de la misma, por lo que parece adecuado traducir el término como aparece en el texto. *(N. del T.)*

© Ediciones Morata, S. L.

Scanned by CamScanner

la calidad de sus anotaciones; hace que el personal que desarrolla el programa reaccione respecto a la precisión de sus descripciones, y los destinatarios reaccionan según la importancia de sus hallazgos. Lleva a cabo la mayor parte de estas operaciones de manera informal: repitiéndolas y anotando acciones y reacciones. Escoge medios accesibles a sus destinatarios para aumentar la viabilidad y fidelidad de la comunicación. Puede preparar, o no, un informe final por escrito, dependiendo del acuerdo al que llegasen sus clientes y él.

(STAKE, 1975, pág. 14.)

Utilizando una metodología antropológica más tradicional, SMITH y POHLAND (1974) evaluaron un programa de enseñanza asistida por ordenador en los Apalaches, y WOLCOTT (1977) evaluó la instalación de un sistema de planificación de programas y de presupuestos en una escuela local. En ambos casos, los evaluadores se basaron en la observación participante, en especial por medio de notas de campo, entrevistas informales y análisis documental.

En Inglaterra, MACDONALD (1974) defiende también el enfoque de los estudios de casos en lo que ha denominado evaluación "democrática":

La evaluación democrática constituye un servicio de información a la comunidad sobre las características de un programa educativo. Reconoce el pluralismo de valores y trata de representar un conjunto de intereses al formular los problemas. El valor básico consiste en una ciudadanía informada y su evaluador actúa como intermediario en los intercambios de información entre distintos grupos. Sus técnicas de recogida y presentación de datos deben estar al alcance de personas no especializadas. Su principal actividad consiste en recoger definiciones del programa y reacciones frente a él. A los informadores, les garantiza la confidencialidad y les confiere el control sobre el uso que pueda hacer de la información. El informe carece de recomendaciones y el evaluador no es responsable del empleo incorrecto de la información. El evaluador mantiene negociaciones periódicas respecto a sus relaciones con los patrocinadores y los participantes en el programa. El criterio de éxito es el ámbito de destinatarios del proceso.

(MACDONALD, 1948, págs. 226-227.)

También en Inglaterra, PARLETT y HAMILTON (1977) defendían un enfoque que denominaban evaluación "iluminadora". Su objetivo consiste en estudiar el programa: cómo opera, cómo influyen en él las situaciones en las que se aplica; qué ventajas e inconvenientes presenta para quienes se ven directamente afectados por él, y en qué aspectos de las tareas intelectuales y experiencias académicas de los alumnos influye de modo especial. "Aspira a descubrir y documentar qué supone participar en el plan, sea como profesor o como alumno, y discernir y exponer, además, las características más significativas, las consecuencias y los procesos críticos de la innovación" (PARLETT y HAMILTON, 1977, pág. 19).

Poco después, GUBA (1978), adoptando en líneas generales este mismo enfoque, trató de definir las dificultades a las que se enfrenta una evaluación

© Ediciones Morata, S. L.

"naturalista". Tal como lo considera, el principal problema es el de la autenticidad: el establecimiento de las bases que permitan fiarse de los resultados de la evaluación. Otros problemas metodológicos son el de fijar los límites de la investigación y el de circunscribirse a las categorías en las que puedan asimilarse y comprenderse los datos. GUBA considera que la metodología naturalista no sólo describe el enfoque del estudio de casos, sino también los modelos crítico y cuasijudicial.

Podemos comparar el enfoque del estudio de casos con el crítico, al que se parece mucho. El crítico se basa en su propia experiencia y aplica normas que escoge él mismo. Quien emplea el estudio de casos se funda tanto en las percepciones de otros como en las suyas, emitiendo sus juicios de valor sobre el programa. En algunos casos, ambos enfoques pueden parecer muy semejantes.

El enfoque del estudio de casos, o de negociación, de la evaluación va adquiriendo gran popularidad. Se están llevando a cabo muchas evaluaciones mediante estudio de casos (consideradas antes muy subjetivas) y existe una bibliografía muy amplia sobre el tema. El enfoque del estudio de casos resuelve el problema del evaluador procurando representar todas las posturas significativas de valor en el estudio del caso de que se trate, extrayendo de esas posturas sus criterios y normas, y dejando al lector del estudio que sopese y equilibre estos elementos por su propia cuenta.

Otros enfoques

He propuesto determinados enfoques, descritos y puestos en práctica por muchos profesionales, como modelos principales de evaluación. He excluido otros por estar muy relacionados con estos tipos principales o ser combinaciones de los mismos. Los enfoques puros no tienen por qué excluirse mutuamente. Está surgiendo un nuevo enfoque, debido a APPLE (1979), LUNDGREN (1977), KALLOS (1978), PAULSTON (en prensa) y otros más. Aunque estos autores presentan diferencias sustanciales entre sí, su enfoque global consiste en asumir una postura teórica que tiene en cuenta de forma deliberada la sociedad. Consideran que ésta se halla organizada de manera que produzca desigualdades de clase, y creen que la educación proporciona los fundamentos ideológicos de esta reproducción social.

Hasta ahora, su trabajo ha consistido sobre todo en críticas de los enfoques de la evaluación vigentes, más que en formular un nuevo enfoque basado en aquél fundamento teórico. Muy pocas evaluaciones se han llevado a cabo de acuerdo con este enfoque. Un estudio que se ha realizado consiste en una investigación a largo plazo de las convicciones que han ido surgiendo en los estudiantes en el transcurso de su formación profesional en la educación superior sueca (FRANKE-WIKBERG, 1979). Podemos suponer que este enfoque de la evaluación se centrará en las desigualdades en el contexto del programa sometido a valoración. Si llega a desarrollarse este

© Ediciones Morata, S. L.

Scanned by CamScanner

enfoque, es probable que los supuestos a él subyacentes se caractericen por ser no liberales, en claro contraste con los dominantes en la actualidad, que hemos presentado antes.

Respecto a la clasificación de las evaluaciones en tipos idealizados, hemos de advertir, por último, que tal clasificación sólo tiene un carácter de aproximación. Muchos evaluadores insistirán en que su enfoque concreto es único y no puede encuadrarse en ninguna de las categorías que he expuesto. En cierto sentido, tienen razón. Los enfoques mencionados sólo son tipos idealizados, constructos utilizados para aclarar el pensamiento en este campo. En cierto sentido, cada enfoque presenta caracteres exclusivos.

Por otra parte, creo que estos tipos básicos dan cuenta de la inmensa mayoría de las evaluaciones que se realizan en la actualidad, y que puede defenderse con fuerza la clasificación de los enfoques concretos utilizados en la práctica en estas categorías generales. Esto no significa que cada evaluación adopte exactamente la forma propia de uno de estos tipos. Las evaluaciones reales están configuradas por determinadas circunstancias y contingencias. Lo que, en ciertas circunstancias, empieza como una evaluación por objetivos puede diferir mucho de ese tipo de evaluación en situaciones distintas. No obstante, es muy probable que, en la mayoría de los casos, una evaluación por objetivos difiera mucho más de otra que siga el enfoque del estudio de casos. La prueba definitiva del esquema de clasificación estriba tanto en su validez como en su utilidad para comprender la evaluación. ALKIN y ELLETT (1979) y ALKIN (1979) han manifestado un punto de vista similar sobre el papel que desempeñan los sistemas categoriales en la evaluación.

© Ediciones Morata, S. L.

CAPÍTULO III

Premisas en las que se basan los enfoques

Liberalismo

Una manera de comprender la evaluación consiste en comparar los numerosos enfoques o modelos entre sí (utilizo los términos "enfoques" y "modelos" como sinónimos). Hay muchas posibilidades de comparación, pero quizá las más significativas sean las que se efectúan entre las premisas teóricas en las que se fundan los enfoques. De este modo, podemos ver hasta qué nivel son similares, desde el punto de vista lógico, y determinar qué posibilidades lógicas tienen y cuáles no.

La cuestión fundamental es que todos los enfoques de evaluación se basan en variaciones de las premisas del liberalismo o, si se prefiere, en las concepciones de la democracia liberal. Los modelos difieren entre sí cuando cambian las premisas básicas. En este contexto, entiendo el término "premisas" según el significado de sentido común, en cuanto cosas que se dan por supuestas o se toman como verdaderas.

En la primera columna de la taxonomía de la Figura 1 del Capítulo II, se citan los nombres corrientes de los enfoques. En la fila superior, aparecen las dimensiones críticas de la comparación: los destinatarios a quienes se dirige la evaluación, sobre qué aspectos existe acuerdo dentro del enfoque en cuestión, la metodología de recogida de datos, los resultados finales previstos y la cuestión característica de la que se ocupa el enfoque. En cierto nivel de abstracción, éstas son las premisas subyacentes al enfoque.

NOTA DEL AUTOR: Una versión de este capítulo apareció con el título: *"Assumptions Underlying Evaluation Modes"*, en: *Educational Researcher*, vol. 7, n.º 3, marzo 1978, págs. 4-12. Copyright 1978: *American Educational Research Association*, Washington, D.C. Reimpreso con autorización. Robert ENNIS, David HAMILTON y Bruce STEWART han aportado sus valiosos comentarios.

© Ediciones Morata, S. L.

Scanned by CamScanner

En la taxonomía, los modelos se relacionan entre sí de forma sistemática. En general, a medida que se desciende por la columna de destinatarios principales, son más democráticos y menos elitistas. Si descendemos por la columna del consenso, menor es el acuerdo sobre los objetivos y otros elementos. En la columna correspondiente a la metodología, cuanto más abajo vayamos, más subjetivista y menos objetivista será la metodología de investigación. Descendiendo por la columna de resultados, decrece la preocupación general por la eficiencia social y aumenta la búsqueda de la comprensión personal. En realidad, se trata de simplificaciones excesivas, puesto que el orden real es más complejo.

Los elementos principales para comprender los enfoques son su ética, su epistemología y sus derivaciones políticas. Todos los modelos vigentes se derivan de la filosofía del liberalismo y las diferencias existentes entre ellos se deben a sus respectivas desviaciones de la corriente principal. No pueden separarse por completo la ética, la epistemología y la política.

El mismo liberalismo surgió del intento de racionalizar y justificar la sociedad de mercado (MACPHERSON, 1966), organizada sobre la base de la libertad de elección. La elección sigue siendo una idea clave en los enfoques de la evaluación, aunque aparezcan diferencias en cuanto a quién elige, qué elige y sobre qué base elige. El ideal último es la elección del usuario, pero difieren las concepciones sobre dicho usuario.

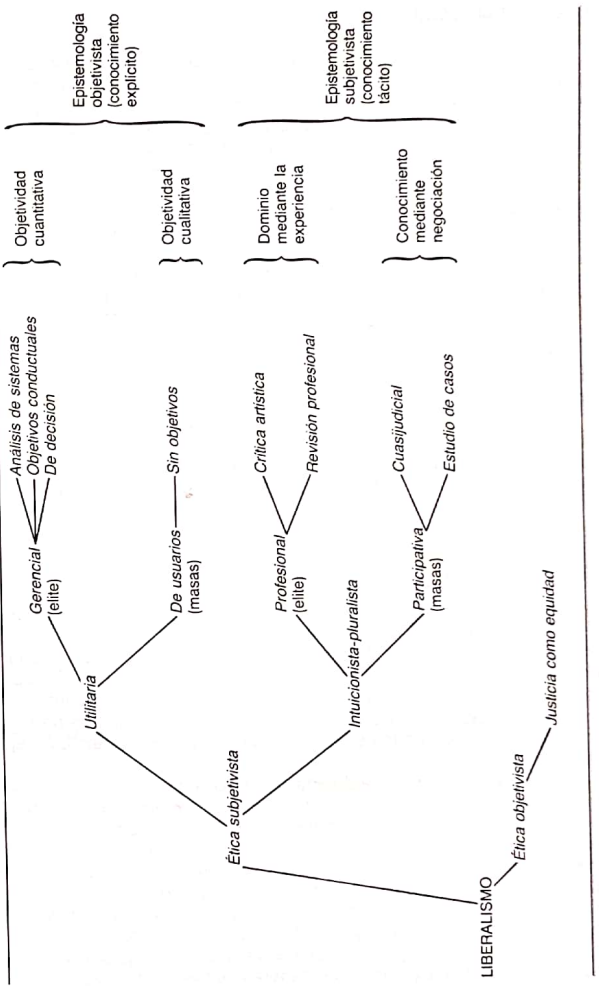
La segunda idea clave del liberalismo consiste en una psicología individualista. Se presume que la existencia de cada mente individual es anterior a la de la sociedad. En principio, no se piensa en el individuo como parte de una colectividad mayor, aunque pueda incluirse más adelante en una de ellas, como en una situación de contrato social. Respecto a sus estructuras intelectuales, el liberalismo es profundamente individualista, desde el punto de vista metodológico.

Otra idea clave consiste en la orientación empirista. Con frecuencia, el liberalismo es radicalmente empirista. Por ejemplo, John Stuart MILL, el apóstol del liberalismo, creía que incluso las matemáticas tenían una base inductiva. Todos los modelos de evaluación muestran ese regusto empirista.

Los enfoques de la evaluación asumen también un mercado libre de ideas en el que los usuarios "comprarán" las mejores. Dan por supuesto que la competencia entre las ideas fortalece la verdad. En último extremo, suponen que el saber hará de alguna manera felices o mejores a las personas. Por tanto, los enfoques de la evaluación comparten las ideas correspondientes a una sociedad mercantil, competitiva e individualista. Pero, la idea más fundamental es la de libertad de elección, porque, si falta, ¿qué utilidad tiene la evaluación? La Figura 2 presenta los enfoques de evaluación relacionados con el liberalismo.

© Ediciones Morata, S. L.

Figura 2: *Esquema de relaciones entre los principales enfoques de evaluación y la filosofía del liberalismo*



© Ediciones Morata, S. L.

Scanned by CamScanner

Ética subjetivista

Todos los modelos principales de evaluación se basan en una ética subjetivista, aunque no sea necesario. Una clase de subjetivismo ético considera que el fin de la conducta ética es la realización de algún tipo de experiencia subjetiva. Se muestra éste en los cuatro enfoques que he denominado "utilitarios". Hablando con propiedad, el utilitarismo se refiere a la idea de elevar al máximo la felicidad en la sociedad. Toda actividad que lo consiga es "correcta" y debe llevarse a cabo (HOUSE, 1976). Por regla general, como indicadores de felicidad se utilizan medidas vicarias, como el producto nacional bruto en economía o las puntuaciones medias de tests en educación.

Junto a los enfoques de análisis de sistemas, objetivos conductuales y de decisión, el modelo de SCRIVEN, que prescinde de los objetivos, se incluye también en esta clase. Todos estos enfoques tratan de llegar a un juicio único de utilidad social general. El enfoque más sencillo es la escala homogénea del enfoque del análisis de sistemas, que trata de reducir todas las variables a un modelo cuantitativo, como el análisis de regresión. Sólo se utilizan una o unas pocas variables de resultados.

El más complejo (el de SCRIVEN) considera muchas variables de resultados, incluyendo efectos secundarios y terciarios. Hay que asignar pesos distintos y combinar todas estas variables de manera que se integren en un juicio sumativo global. Puede establecerse una analogía con un informe de una asociación de consumidores en el que se integran criterios muy diversos para dar como resultado una clasificación global de "mejor" y "óptimo". Como en el caso de la asociación de consumidores, la clasificación final será ordinaria y no de intervalos. Es utilitaria en el sentido de que el resultado final consiste en *una* clasificación final de méritos, encabezada por la mejor opción para el usuario y, en último término, de máxima utilidad social.

He denominado los otros cuatro modelos principales de evaluación "intuicionistas-pluralistas". Los principios éticos no son únicos ni están definidos explícitamente como en la ética utilitaria. Hay varios principios sin reglas para sopesarlos y esos principios se derivan de la intuición y la experiencia. Esto pone de relieve otro sentido del subjetivismo ético: los criterios últimos de lo que es bueno y correcto son sentimientos o apreciaciones individuales.

En el modelo de la crítica de arte, el crítico llega a sus principios invariablemente plurales a través de un largo período de preparación y experiencia. Su gusto se ha perfeccionado tanto que es un juez experto. Gran parte de su maestría se basa en su capacidad para equilibrar de modo intuitivo las premisas de sus juicios. En el enfoque profesional, las premisas pueden expresarse en libros como *Evaluative Criteria (National Study of Secondary School Evaluation,* 1978), pero su origen y aplicación siguen siendo una cuestión que tiene mucho de juicio intuitivo basado en una larga experiencia.

En el enfoque cuasijudicial, como en los juicios, el jurado tiene libertad para utilizar múltiples criterios de manera intuitiva. Los adversarios pueden

© Ediciones Morata, S. L.

introducir lo que gusten en calidad de evidencia que puede estructurarse como pertinente. Por último, en el enfoque del estudio de casos, se introducen criterios plurales solicitando a diversas personas participantes en el programa que emitan sus juicios respectivos. El evaluador y los destinatarios de la evaluación sopesan los datos de manera intuitiva. En general, los utilitaristas suelen adoptar técnicas definidas de forma explícita como base de la ética y el saber, mientras que los intuicionistas-pluralistas asumen como fundamento de aquéllos el saber intuitivo tácito que se deriva de la experiencia profesional y la participación.

Aunque todos los enfoques principales se basan en una ética subjetivista, no tendría por qué ser así. En el marco de la filosofía liberal, podría darse una ética objetivista. Por ejemplo, en su *A Theory of Justice,* John RAWLS (1971) trató de fundamentar su ética social en dos principios de justicia. Esos principios no se justificaban sobre la base de que llevaran al máximo un sentimiento subjetivo, como la felicidad, ni en el hecho de que las personas sostuvieran intuitivamente tales principios. Las premisas se justificaban, en cambio, sobre la base de que el análisis racional, o sea, la teoría de reducir al mínimo las pérdidas, llevaría a *cualquier* persona racional hasta las mismas premisas. Eran objetivistas en el sentido de que se basaban en una lógica pública susceptible de revisión. La teoría de RAWLS seguiría encuadrándose en la tradición liberal de la libre elección y el individualismo metodológico.

El grupo utilitarista se diferencia internamente por su orientación hacia distintos destinatarios. Los enfoques de análisis de sistemas, objetivos conductuales y de decisión se orientan en gran medida a suministrar información a los gestores o dirigentes. El enfoque que prescinde de los objetivos del programa tiende más hacia los usuarios. Esta distinción entre elite y masas aparece también entre los intuicionistas-pluralistas. Los enfoques críticos artístico y profesional se basan, para juzgar el programa que se evalúa, en profesionales expertos, mientras que los enfoques cuasijudicial y de estudio de casos requieren la participación de profesionales y público.

La epistemología objetivista liberal

Los enfoques gerenciales, con su orientación hacia la gestión, buscan la manera de que la dirección pueda supervisar y perfeccionar los programas. Las principales preocupaciones de la dirección son la responsabilidad, la eficiencia y el control de calidad. La evaluación tiene una cierta función de perro guardián. La escuela gerencial suele basarse también en una epistemología común, que he denominado "objetivista". Se considera que la información de la evaluación es "científicamente objetiva". Esta objetividad se consigue utilizando instrumentos "objetivos", como tests o cuestionarios. Se supone que los resultados que producen estos instrumentos son reproducibles. Los datos se analizan por medio de técnicas cuantitativas, que también

© Ediciones Morata, S. L.

Scanned by CamScanner

son "objetivas" en el sentido de que pueden verificarse mediante la revisión lógica de las mismas, con independencia de quien las emplee.

En su forma extrema, el objetivismo epistemológico excluye por completo lo no cuantitativo: lo que existe puede medirse. La reducción a números se equipara a veces a la objetividad. En buena medida, los orígenes de este empirismo y objetivismo extremos pueden encontrarse en la primitiva epistemología de la filosofía liberal.

En parte, podemos situar los orígenes del liberalismo como filosofía en el primer gran empirista británico: Thomas HOBBES. Este autor era un admirador de GALILEO y de su visión mecánica global del universo. HOBBES elaboró una filosofía mecanicista de la mente, basada en la idea de que las sensaciones se producen en la mente a causa de los movimientos de cuerpos externos. Las palabras sólo tienen significado cuando se asocian con estas sensaciones y la verdad consiste en el correcto ordenamiento de nombres y sensaciones. Las definiciones claras son fundamentales. El razonamiento equivale al cálculo y sólo será correcto cuando los signos se vinculen con imágenes coherentes. Causa y efecto constituyen un reflejo directo del mundo. Por tanto, el camino hacia la verdad pasa por la definición clara y el método científico.

LOCKE, BERKELEY y HUME, a su vez, reformularon esta epistemología. Doscientos años después, apareció el último de los grandes empiristas británicos: John Stuart MILL, llamado también el apóstol del liberalismo. Según MILL, la mente es un manojo de impresiones. Las leyes psicológicas de la asociación constituirían el fundamento de la ciencia de la sociedad. Todo proviene de la experiencia y ha de ser comprobado mediante ésta, hasta el punto de que incluso la lógica deductiva (las matemáticas, por ejemplo) tiene una base inductiva. MILL no veía ninguna diferencia fundamental entre las verdades lógicas y las verdades basadas en hechos. Pensaba que todo el conocimiento tenía que fundamentarse en la experiencia y comprobarse mediante ésta. El saber consistía en la suma de las observaciones individuales.

Este individualismo metodológico extremo llegaba hasta lo más profundo de la filosofía de MILL. Según él, la sociedad era una colección de personas individuales unidas por intereses comunes. Su empirismo radical le llevó a formular reglas para el desarrollo de la investigación inductiva en *A System of Logic*. Como la mente consistía en sensaciones, sentimientos y sus asociaciones, para MILL, las categorías del mundo real eran fácilmente accesibles. Pero, ¿en qué se basaban estas categorías? Si fuesen efímeras, nada podríamos conocer. Para solucionar este problema, MILL *dio por supuesto* que el mundo consistía en uniformidades constantes entre clases de sensaciones.

La experimentación era esencial para la comprobación científica de las teorías. Las conexiones causales válidas sólo podían establecerse mediante el meticuloso ordenamiento y la variación independiente de las condiciones antecedentes y las consecuencias. Con frecuencia, éstas quedaban ocultas a la observación casual. Las apariencias "sin depurar" resultaban engañosas.

© Ediciones Morata, S. L.

Sólo las técnicas de la lógica inductiva podían purificarlas. En este contexto, eran útiles las técnicas como el método del acuerdo, el de la diferencia o el de la variación concomitante. Los cánones de la inducción proporcionaban un método para filtrar la evidencia de las correlaciones en las que se basaban las leyes científicas.

Se trataba de llegar a leyes científicas. La ciencia debía ocuparse de las uniformidades causales que afectaran a todos los elementos pertenecientes a una clase dada. La clave de la que dependía todo el sistema lógico de MILL era el supuesto de la uniformidad de la naturaleza. Con palabras de este autor:

> En primer lugar, tenemos que observar que, en el mismo enunciado de lo que es la inducción, hay una premisa básica; un supuesto previo en relación con el decurso de la naturaleza y el orden del universo: en la naturaleza existen casos paralelos, de modo que lo que sucede una vez, si se da un grado suficiente de circunstancias semejantes, ocurrirá de nuevo, y no sólo otra vez, sino tantas veces cuantas vuelvan a concurrir las mismas circunstancias... Y si observamos el curso real de la naturaleza, veremos que la premisa es cierta. El universo, en la medida en que lo conocemos, está constituido de tal manera que lo que es cierto en un caso, lo es en todos los casos de una misma clase; la única dificultad está en descubrir de qué clase se trata.
>
> (MILL, 1893.)

Este empirismo extremo se basaba en el convencimiento de que no era difícil descubrir las categorías, como en el caso de las categorías de las ciencias físicas. Utilizando las técnicas adecuadas, podrían definirse las variables y establecer las relaciones entre ellas. Como la naturaleza sería uniforme, podríamos inducir las leyes de las ciencias sociales, del mismo modo que en las ciencias naturales. Es más, esto podría llevarse a cabo mediante la metodología de encuesta. La evaluación ha heredado ese extremo interés por el empirismo en la exclusión de la teoría a partir de la epistemología liberal: en la naturaleza, podemos descubrir relaciones sin necesidad de constructos teóricos.

HAMILTON (1977) estudia la aparición de algunas de estas ideas en la evaluación. Un contemporáneo de MILL, Francis GALTON, creó técnicas correlacionales para establecer las conexiones entre las categorías que, según MILL, eran fácilmente apreciables. GALTON proporcionó también un importante fundamento para la psicometría en su obra sobre las diferencias individuales. CATTELL, muy influido por GALTON, introdujo los tests mentales en los Estados Unidos.

El uso de los tests en educación con fines de clasificación fue promovido por E. L. THORNDIKE. Este autor tuvo también gran influencia en el movimiento de "gestión científica" de la educación. Dicho movimiento introdujo en las escuelas la ingeniería de la eficiencia. Los tests, las escalas de clasificación y las técnicas cuantitativas estimuladas por los trabajos de THORNDIKE se

© Ediciones Morata, S. L.

Scanned by CamScanner

utilizaron como instrumentos para lograr la eficiencia de las escuelas (CALLAHAN, 1962).

Al hacer hincapié en el análisis de tareas, la especificación previa y muchos otros aspectos, la gestión científica presenta un importante parecido con los enfoques del análisis de sistemas y de los objetivos conductuales de la evaluación (HOUSE, 1978). De hecho, uno de los principales defensores de la gestión científica de la educación fue Franklin BOBBITT. Más tarde, BOBBITT y CHARTERS adoptaron un enfoque de análisis de tarea y especificación de la misma en relación con el desarrollo curricular. Uno de los alumnos posgraduados de CHARTERS fue Ralph TYLER, creador del modelo de evaluación por objetivos conductuales. La orientación gerencial de la evaluación tiene profundas raíces históricas.

Con independencia de sus conexiones históricas, la semejanza entre los modelos gerenciales de evaluación y la epistemología de MILL es grande. Quizá sea Alice M. RIVLIN la principal exponente del enfoque del análisis de sistemas. Pertenecía al grupo que introdujo el análisis coste/producto en la educación y en otros programas sociales en el *Department of Health, Education and Welfare* en los años sesenta. La mayor parte de los más amplios programas federales de evaluación se basaron en la lógica del análisis de sistemas (MCLAUGHLIN, 1975).

Las ideas clave del enfoque de RIVLIN (1971) eran las siguientes (HOUSE, 1978):

- Las decisiones clave deben tomarse en los niveles gubernamentales superiores.
- El fin de la evaluación es la eficiencia en la producción de servicios sociales.
- El único conocimiento verdadero consiste en una función de producción que especifique relaciones estables entre *inputs* y *outputs* educativos.
- Sólo puede conseguirse ese conocimiento mediante los métodos experimentales y las técnicas estadísticas.
- Es posible lograr el acuerdo sobre los objetivos y sobre unas pocas medidas de resultados.
- Existe un paralelismo directo entre la producción de servicios sociales y la de artículos manufacturados. Se aplican las mismas técnicas de análisis.

Como MILL, RIVLIN da por supuesto que el paradigma de las ciencias físicas es adecuado para estos menesteres. Consideraba que las mismas técnicas de investigación servían para los servicios sociales y para la fabricación de productos. El resultado sería unas generalizaciones (funciones de producción, en este caso) válidas en diversas circunstancias, siempre que pudiesen determinarse éstas. Esta concepción mecanicista se remonta hasta HOBBES. Y las premisas de la uniformidad de la naturaleza nos llevan hasta J. S. MILL. Los resultados serían consistentes en cualquier parte.

© Ediciones Morata, S. L.

RIVLIN creía, como también MILL, que hacían falta muestras grandes para evitar los aspectos idiosincrásicos.

MILL pensaba que el principal problema moral era la ilustración y que los hombres resultaban engañados por sus instituciones. RIVLIN creía que el principal problema era la ignorancia respecto a cómo producir servicios sociales de manera más eficaz. Ambos estaban convencidos de que el empleo de expertos eliminaría esos problemas. Asimismo, ambos creían en la utilización de unas cuantas medidas como sustitutos de la felicidad. Esas medidas son necesarias para los cálculos precisos para el funcionamiento de la ética utilitaria.

No obstante, hay un aspecto en el que el liberal clásico J. S. MILL y un liberal moderno se distancian. Para MILL, la libertad era necesaria para que se produjera la ilustración. Fue uno de los grandes defensores del *laissez-faire*. El gobierno debía apartarse de los asuntos de los individuos. Según RIVLIN, una liberal del estado del bienestar, el gobierno tendría que tomar las decisiones y ponerlas en práctica. Con el fin de descubrir lo que funciona, el gobierno necesita mayor control para poder desarrollar experimentos. A excepción de campos especiales, como el de la educación, en el que aceptaba la planificación y la coerción, incluso, MILL no estaba a favor de la planificación social a gran escala. RIVLIN constituye el epítome de dicha planificación.

El objetivismo de los enfoques gerenciales de la evaluación, que tiende a equiparar la objetividad con la cuantificación, se basa en el acuerdo intersubjetivo como indicador exclusivo de objetividad. SCRIVEN (1972) lo llama "sentido cuantitativo de la objetividad". La percepción que una persona tiene de algo se considera subjetiva: es la disposición de un individuo. Lo objetivo está constituido por lo que experimenta un conjunto de observadores. La experiencia común hace pública la observación mediante el acuerdo intersubjetivo.

Ha llegado a equipararse la objetividad con la exteriorización de todas las referencias, de manera que se llegue a un testimonio múltiple, dado que, según la doctrina de la intersubjetividad, no es aceptable lo que no puedan experimentar otros de forma directa. Por tanto, con frecuencia, se equipara la objetividad con la posibilidad de especificar y explicar por completo todos los procedimientos de recogida de datos. La total exteriorización permite la réplica, la piedra de toque de la fiabilidad.

Describiendo los procedimientos de análisis de sistemas, el *Department of Defense* plantea de este modo la cuestión:

> Quizá el más importante de ellos sea el método científico. Esto significa que los datos utilizados en el análisis deben ser susceptibles de verificación, bien mediante observación, bien por deducción a partir de premisas aceptables, y que los procedimientos empleados en el análisis deben ajustarse a las reglas aceptadas de la lógica. El análisis de sistemas se caracteriza, en segundo lugar, por el uso de datos cuantificables y técnicas matemáticas... El común denominador final del análisis de sistemas es la claridad. Las premisas y criterios utilizados en

© Ediciones Morata, S. L.

Scanned by CamScanner

un análisis se especifican y los aspectos inconmensurables y las áreas de incertidumbre se delimitan con todo cuidado.

(SNYDER, 1967, pág. 5.)

De acuerdo con este razonamiento, se asigna la máxima prioridad a los "instrumentos fiables". En realidad, cuanto mayor sea el acuerdo entre observadores (coeficiente de fiabilidad), mejor parecerá el instrumento o procedimiento. Por desgracia, la fiabilidad no garantiza la validez y la falacia intersubjetiva puede llevar a utilizar "instrumentos" con los que se logra un alto grado de acuerdo entre observadores en perjuicio de procedimientos de superior validez. Los instrumentos de elevada fiabilidad constituyen los hitos de este objetivismo epistemológico.

En cierto sentido, el principio de verificación intersubjetiva se traslada a la definición de los objetivos, así como a la medida de los efectos. En el enfoque gerencial de la evaluación, se supone que es posible llegar al consenso sobre los objetivos de un proyecto concreto. Este supuesto es fundamental en los modelos de análisis de sistemas y de objetivos conductuales, porque el consenso define lo que buscará la evaluación. Además, el modelo de análisis de sistemas da también por supuesto el acuerdo sobre las relaciones entre causa y efecto. MILL presuponía que la causa y el efecto son evidentes y discernibles por naturaleza.

Por tanto, en el objetivismo, la noción de objetividad se equipara con los procedimientos para determinar la intersubjetividad. Esto se produce tanto en la ética como en la epistemología. En realidad, el utilitarismo requiere que haya una norma única de utilidad social con la que comparar las cosas.

Aunque he situado la evaluación independiente de los objetivos de SCRIVEN en el grupo utilitarista, ésta es algo más compleja. SCRIVEN emplea la idea *cualitativa* de objetividad. En este caso, ésta se equipara con la ausencia de sesgos y deformaciones. La preocupación suprema de SCRIVEN consiste en el control de los sesgos. Puede utilizar el diseño experimental, como el grupo gerencial, pero ha elaborado todo un arsenal de dispositivos organizativos y sociales para controlar las desviaciones tendenciosas en su sentido más amplio (SCRIVEN, 1975, 1976).

En el marco de SCRIVEN, es posible que un único observador, sin el auxilio de instrumento psicométrico alguno, sea más objetivo que un batallón de observadores cargado de instrumentos fiables, siempre que el observador único busque lo que debe buscar y esté lo bastante protegido contra numerosos sesgos. SCRIVEN no equipara la objetividad al acuerdo intersubjetivo.

Tampoco hace suyo el consenso respecto a metas y objetivos, como pone de manifiesto de manera espectacular el enfoque independiente de objetivos. Es probable que la evaluación independiente de objetivos requiera el acuerdo sobre las consecuencias del programa. (De nuevo, se refiere más a la elección del usuario que a la del directivo.) SCRIVEN combina

© Ediciones Morata, S. L.

múltiples criterios de consecuencias con la escala ordinal y un enfoque comparativo implacable.

La epistemología subjetivista

La evaluación utilitarista se basa, por tanto, en una ética subjetivista, de manera que eleve al máximo un estado mental, como es la satisfacción; pero, para hacerlo, emplea una epistemología objetivista. Cuando se ha determinado lo que debe promoverse al máximo, la metodología puede utilizarse de forma objetivista para asignar recursos que satisfagan diversos deseos.

Los enfoques intuicionistas-pluralistas de la evaluación emplean una ética subjetivista de clase diferente. En general, también su metodología de investigación suele ser subjetivista. Los procedimientos mediante los cuales un crítico educativo o un equipo de acreditación lleva a cabo su investigación son más generales que específicos. La claridad de detalles y la exteriorización de los procedimientos, de manera que otros puedan observar lo mismo, no son denominadores comunes. Tampoco la posibilidad de reproducción es el criterio principal.

Del mismo modo, en los enfoques cuasijudicial y de estudio de casos, no se especifica del todo cómo llega un jurado a una decisión ni por qué un estudioso de casos escoge ciertos materiales y no otros. El modo en que el autor clave desarrolla su inventigación se distingue por una subjetividad importante.

Mientras los objetivistas se fundan en la claridad de detalles para definir técnicas que puedan utilizar otros, los intuicionistas se basan en la preparación y la experiencia para garantizar el descubrimiento de la verdad. En estos enfoques, llevados al extremo, tanto la validez como la utilidad se definen de manera subjetiva. La validez se concibe en relación con los condicionamientos de la mente humana, bien a causa de las limitaciones universales que pesan sobre la forma de pensar de las personas, bien a causa de limitaciones personales. (A menudo, en el marco objetivista, se concibe la validez como la previsión de una categoría observable a partir de otra.)

La validez subjetiva significa que la verdad depende de la naturaleza humana y, quizá incluso, de los humanos concretos. Lo que es válido para uno, puede no serlo para otro. De igual modo, la utilidad subjetiva de algo se basa en el juicio y en los deseos personales. Cada persona es el mejor juez de los hechos para sí misma.

En realidad, la epistemología subjetivista no es tan anárquica como parece. Mientras que la postura objetivista se basa en la exteriorización y plena especificación de las técnicas, la postura subjetivista se funda en la tradición del "caso cerrado". Mediante la preparación, la experiencia y la socialización, el evaluador incorpora precedentes a sus juicios. La justicia constituye el ejemplo más obvio de un conjunto de casos que orientan el juicio.

© Ediciones Morata, S. L.

Scanned by CamScanner

A los subjetivistas les importa menos llegar a una proposición "verdadera" (en el sentido de generalizable) que relacionar la evaluación con la experiencia concreta de los destinatarios. Tratan de conseguir una perspectiva válida en el marco de referencia del grupo para el que trabajan. Se da por supuesta la distancia entre el lenguaje y la experiencia. El evaluador busca más el conocimiento tácito que el explícito. La evaluación está intencionadamente ligada al contexto y sus descubrimientos se interpretan dentro del mismo. Como los destinatarios pueden tener una captación más fundamentada del contexto, basados en su mayor experiencia, la interpretación que ellos hagan de un acontecimiento puede ser superior a la del evaluador. A los evaluadores subjetivistas les preocupan más en su trabajo los enunciados causales específicos (validez interna) que los enunciados causales generales (validez externa) (ENNIS, 1973). Las generalizaciones quedan para los destinatarios.

La metodología subjetivista suele ser naturalista. Aspira a la generalización naturalista (basada en la experiencia de los destinatarios); se dirige más a destinatarios no técnicos, como profesores y público en general; utiliza el lenguaje ordinario y categorías cotidianas de acontecimientos, y se basa más en la lógica informal que en la formal. Los instrumentos preferidos para recoger datos son las entrevistas y observaciones informales. Con frecuencia, sus trabajos se ponen por escrito a modo de estudios de casos.

El evaluador procura recoger múltiples perspectivas, aunque no concuerden entre sí. Asimismo, tienen gran relieve la investigación histórica y el seguimiento de los hechos en el transcurso del tiempo. El modo histórico de investigación es más adecuado que el de las ciencias naturales. Se hace más hincapié en lo cualitativo que en lo cuantitativo.

En la metodología subjetivista, la utilidad se plantea en términos de los intereses del observador. La teoría y la práctica se mezclan. El objetivo consiste en mejorar la comprensión de individuos particulares. En el enfoque objetivista, se da una separación rígida entre el observador y los hechos (una teoría muy abstracta separada de la aplicación). Su objetivo es la previsión.

La técnica del caso cerrado y el uso de los precedentes proporcionan estabilidad a los juicios que se derivan de la epistemología subjetivista. En realidad, en estos enfoques de evaluación, tiene tanto peso la tradición que se pueden criticar por eso mismo. Ocurre esto porque los enfoques subjetivistas se basan en gran medida en el juicio del profesional que, a su vez, se funda en la experiencia de la clase. Los modelos subjetivistas consideran la educación y la enseñanza como un arte, más que como un conjunto de técnicas explícitas y exteriorizadas.

El enfoque de estudio de casos de la evaluación se centra en los hechos que se producen dentro y alrededor del programa concreto en contexto. Se basa en la percepción y en el conocimiento como proceso negociador. En una versión de la psicología transaccional, la percepción humana depende de tres características (ITTELSON y CANTRIL, 1954). *En primer lugar,* los hechos de la percepción se presentan siempre mediante individuos

© Ediciones Morata, S. L.

concretos que se desenvuelven en situaciones concretas. Por tanto, sólo pueden estudiarse las percepciones a través de las transacciones concretas en las que pueden observarse. Ni las percepciones ni el objeto percibido existen con independencia de una situación vital. La situación es una transacción. Todas las partes que intervienen en la situación se incluyen en la transacción como "participantes activos". Las partes deben su existencia a esta participación activa y no aparecen como entidades independientes preexistentes. En consecuencia, la participación activa constituye una idea clave.

En segundo lugar, la percepción siempre es de una persona particular, desde su propia postura, experiencias y necesidades únicas. Hay tantos puntos de vista como participantes. Incluso los científicos observadores entran en la situación como participantes. Ellos influyen en la situación y, a su vez, son influidos por ella. Por tanto, el punto de vista del científico observador sólo es uno entre muchos. Forma parte de la transacción.

Tercero, cada persona crea su propio ambiente psicológico, atribuyendo aspectos de su experiencia al ambiente. Cree que el ambiente existe con independencia de su experiencia. En otras palabras, exterioriza su experiencia. En consecuencia, el mundo, tal como se experimenta, es el *producto* de la percepción y no su causa.

Estas premisas respecto a la percepción difieren por completo de las de J. S. MILL y HOBBES. Desde el punto de vista mecanicista hobbesiano, el ambiente actúa sobre el organismo para producir la percepción. En la perspectiva transaccional, éste también opera en el ambiente de forma activa, por lo que la percepción es producto de ambos. Las categorías no se distinguen con facilidad del ambiente porque están inextricablemente mezcladas con las propias percepciones.

Desde el punto de vista hobbesiano-milliano, el mundo se toma como dado. La uniformidad mecanicista de la naturaleza lo garantiza. Las percepciones pueden aislarse y estudiarse. Un científico puede comunicar enunciados a otro científico, que son reducibles a lo que el otro experimentará si hace determinadas cosas (el punto de vista operacional). En la perspectiva transaccional, la fiabilidad no se consigue con tanta facilidad. Lo que uno observa es función de la posición y experiencia únicas de cada observador e, incluso, de sus fines. Las percepciones sólo pueden estudiarse en situaciones concretas de la vida real.

No todos los defensores de los enfoques de estudio de casos de la evaluación suscribirían todos estos presupuestos, pero esta descripción presenta el punto de vista transaccional como interacción dialéctica entre las personas y su ambiente que se traduce en percepciones. Las percepciones individuales constituyen un centro de atención que estudiar y la participación activa es esencial para el conocimiento.

© Ediciones Morata, S. L.

Scanned by CamScanner

Premisas políticas de los modelos: los utilitaristas

El liberalismo no sólo es una teoría ética y epistemológica, sino también una teoría profundamente política. Como señalamos antes, aunque RIVLIN y los liberales modernos aceptan la ética utilitaria y el empirismo extremo de J. S. MILL, difieren espectacularmente de su postura respecto al papel que haya de desempeñar el gobierno en cuanto a las decisiones y a la iniciativa de acción. El liberalismo de bienestar de RIVLIN sostiene que el gobierno debe tomar las principales decisiones, mientras que MILL era el campeón del *laissez-faire*, de mantener al gobierno lo más apartado posible de las decisiones. En este sentido, los intuicionistas están más cerca del liberalismo clásico que los utilitaristas.

Según MILL, la ilustración era necesaria para la felicidad y la libertad. La ausencia de interferencia del gobierno era necesaria para la ilustración. Según RIVLIN, la ilustración se producirá a través de la investigación patrocinada por el gobierno y, sobre la base de estos conocimientos, el gobierno emprenderá los pasos necesarios para paliar los problemas. En ambos casos, el saber hará felices a los hombres.

Según MILL, hay dos esferas distintas de actividades: interna y externa. La interna abarca la mente, el cuerpo y las acciones privadas del propio individuo. En la esfera interna, el gobierno no tiene absolutamente ningún derecho a interferir en nada. El individuo podría entretenerse o abusar de sí mismo cuando se tratase de cuestiones que sólo le afectaran a él (WOLFF, 1968). El ámbito de actuación del gobierno se limita sólo a lo que concierne a terceros: la esfera externa.

Incluso en la esfera externa en la estructura de MILL, sólo se justifica la interferencia social sobre la base del principio de utilidad. Es decir, la interferencia gubernamental sólo estaría justificada cuando promoviera la mayor felicidad del mayor número de personas, elevando al máximo la satisfacción.

> Las acciones guardan una proporción correcta cuando tienden a promover la felicidad, y errónea cuando tienden a producir lo contrario a ella. Entendemos por felicidad el placer y la ausencia de dolor; por infelicidad, el dolor y la privación del placer.
>
> (MILL, 1861.)

Los argumentos de MILL eran estrictamente utilitarios, o sea, basados en una estimación de las consecuencias futuras de diversas alternativas. Si las estimaciones empíricas de las consecuencias cambian, también lo harán las conclusiones. Los modernos liberales del estado de bienestar han modificado las estimaciones empíricas y, en consecuencia, las conclusiones. Según los liberales modernos, el principio de utilidad requiere que el gobierno emprenda una acción enérgica. El cálculo del principio de utilidad lleva, en último término, al liberalismo del estado de bienestar (WOLFF, 1968).

Según MILL, cualquier persona es el mejor juez de su propio interés.

© Ediciones Morata, S. L.

(Debemos señalar que MILL escribía en un momento en que las masas carecían de voto. Sus teorías sólo se aplicaban a la ciudadanía más "ilustrada".) Para los liberales modernos, como RIVLIN (1971), no es así. A menudo, el gobierno es el mejor juez. En el marco de MILL, la completa libertad de expresión lleva a una competencia de ideas que refuerza la verdad. En un mercado libre de ideas, los consumidores adquirirán lo mejor sin la interferencia gubernamental. Para el liberal moderno, el gobierno debe proporcionar índices de eficacia sobre los que fundar los juicios y, a menudo, elegir en aras del interés público. Los liberales, clásicos y modernos, están de acuerdo en los principios políticos liberales, pero disienten respecto al carácter beneficio o dañino de la interferencia gubernamental.

MILL cree que la interferencia del gobierno se traducirá en el impedimento del desarrollo individual, la centralización y el abuso de poder, la ineficacia, la acción desganada y de baja calidad del gobierno que las partes privadas interesadas podrían realizar mejor y la supresión de la iniciativa individual que sienta las bases para el despotismo. Sin embargo, incluso MILL cree que, si el comprador no fuese el mejor juez de un bien, la competencia del mercado no existiría. En los asuntos culturales y educativos, MILL no cree que los incultos sean los mejores jueces. Por tanto, el estado podría intervenir legítimamente, postura próxima a la de los liberales modernos.

En la esfera privada, que consiste en la libertad de pensamiento, conciencia, opinión, sentimiento, "gustos y pasatiempos" y asociación, MILL pensaba que cualquier persona sería, sin duda, el mejor juez de sus propios intereses por ser la más preocupada por su propio bienestar. Su conocimiento de sus sentimientos y circunstancias sería muy superior al que pudiera tener cualquier otra respecto de ella misma. Por tanto, sería mejor dejar a cada individuo que hiciera sus propios juicios, aunque pareciesen equivocados o dañinos para el interesado.

Pero, en general, el cálculo del principio de utilidad conduce más bien a potenciar una fuerte intervención gubernamental que a evitarla. No nos sorprende, por tanto, que todos los modelos utilitaristas de evaluación (análisis de sistemas, por objetivos conductuales, de decisión e independiente de los objetivos) se basen en una acción de gobierno, normalmente en forma de mandato gubernamental, incluso. La mayor parte de estos modelos tienen como principales destinatarios también a los directivos y administradores de programas gubernamentales. La evaluación independiente de los objetivos del programa, cuyos principales destinatarios son los usuarios, constituye la única excepción.

Premisas políticas de los intuicionistas-pluralistas

¿Cómo puede llevarse a cabo una evaluación si no existe un principio único, como el de utilidad, para usarlo como criterio? En los enfoques intuicionistas-pluralistas hallamos dos tipos de respuestas. Una consiste en la autoridad profesional, como en los enfoques crítico y profesional, y la otra

© Ediciones Morata, S. L.

Scanned by CamScanner

es una combinación de autoridad científica y de participación en la evaluación de quienes desarrollan el programa. El enfoque profesional comenzó como asociación voluntaria en contra de la interferencia del gobierno y se basa en la idea de la maestría por medio de la experiencia.

Los enfoques participativos constituyen intentos de promover una colaboración más directa de las personas más comprometidas en el programa. El enfoque cuasijudicial consigue esto haciendo que las personas se involucren en una operación de decisión semejante a un juicio. El enfoque del estudio de casos implica a las personas a través de la negociación, la entrevista y la respuesta a evaluaciones provisionales, aunque sea el evaluador quien redacte el informe. Para que las personas participen en calidad de destinatarios principales, la evaluación ha de ser inmediata y comprensible. La escuela participativa pretende establecer una democracia más directa entre los implicados en el programa en vez de dejar las decisiones al gobierno, como en el caso de la evaluación utilitaria. Los principios pluralistas se derivan de la participación de diversos individuos y grupos.

Existen dos enfoques distintos del pluralismo político. Uno se deriva directamente de la filosofía liberal clásica de J. S. MILL, sin su epistemología ni su ética utilitarista. El liberalismo clásico considera la sociedad como una asociación de individuos autodeterminados que cooperan con los demás para conseguir sus propios fines. De acuerdo con MILL, es esencial que puedan autorgobernarse, dado que son ellos quienes mejor se conocen a sí mismos y a sus intereses. Esta versión del pluralismo acepta como inevitables las idiosincrasias individuales y los conflictos interpersonales (WOLFF, 1968).

La inviolabilidad del individuo frente a la intrusión de la sociedad es fundamental y la esfera íntima de la persona debe conservarse inviolada. No pueden interferirse sus pensamientos y prácticas. En vez de suprimirse, se estimulan la individualidad y la diversidad. Se concibe la sociedad como un mercado en el que cada individuo trata de conseguir sus propios fines. Cada hombre goza de libertad para procurar alcanzar sus propias metas de manera compatible con la de los demás que pretenden lograr las suyas. En la esfera pública, la sociedad tiene derecho a imponer algunas reglas de equidad. Por lo demás, el mercado regula las cosas de forma automática.

La sociedad ideal que se preconiza es la de una gran ciudad como Londres o Nueva York. El liberalismo clásico considera al hombre como un calculador racional del dolor y el placer. La racionalidad consiste en la prudencia calculada. Cada individuo ve a los demás como instrumentos para conseguir sus propios fines. Los demás son medios para algún fin y no fines en sí mismos. A veces, ésta se denomina teoría "instrumental" del pluralismo (WOLFF, 1968).

La justicia consiste en la protección de la esfera íntima de la persona y en la oportunidad, más o menos igual para todos, de conseguir sus propios fines. En versiones más complejas, se suaviza hasta cierto punto la consecución del máximo placer, haciéndose hincapié en el logro de objetivos más elevados. No obstante, sigue considerándose la sociedad como una

© Ediciones Morata, S. L.

colección de "centros de conciencia" independientes, cada uno de los cuales procura conseguir sus propios fines y se enfrenta a los demás como si de objetos extraños se tratase (WOLFF, 1968).

Asociada con esta versión del pluralismo está la teoría "arbitral" del gobierno. Éste tendrá que establecer las reglas básicas para la competición entre individuos y grupos, aunque sin inmiscuirse de ninguna otra manera. La acción debe ser *directa* de los individuos y grupos interesados y no una acción de gobierno. Cada persona tiene que gozar de libertad para participar de forma directa en los procesos de decisión y este papel no puede delegarse ni en el gobierno ni en nadie más, como tampoco puede ser usurpado por ellos. Esta versión de la democracia directa surge del liberalismo clásico, en el que se promueve al máximo la elección de *cada individuo*.

El enfoque de evaluación que más se corresponde con esta versión del pluralismo liberal es el de la evaluación democrática de MACDONALD (1974). Este autor considera al evaluador como "intermediario en los intercambios de información entre grupos", representando un conjunto de intereses, presentando información accesible a los no especialistas, brindando confidencialidad y proporcionando control sobre los datos a los informadores, sin intención de utilizar mal la información, negociando con los patrocinadores y participantes y absteniéndose de hacer recomendaciones. El evaluador facilita información a los destinatarios y deja que el mercado resuelva. Cada persona hace el uso que le parece de la información, relevando al evaluador de hacer interpretaciones. Éste opera sobre la base de un conjunto de normas de procedimiento que controla el flujo de información.

MACDONALD ha sido también un máximo defensor de la protección de la inviolabilidad de la "esfera íntima", concediendo a los informadores el derecho a vetar lo que pudiera aparecer en el informe. Esto se refleja en su máxima: "las personas son dueñas de los hechos de sus vidas". Se rechaza la interferencia del evaluador tanto en la esfera privada como en la pública, salvo para facilitar el flujo de información que incremente el conocimiento y, quizá, la felicidad, al permitir a los individuos que traten de conseguir con mayor facilidad sus propios fines.

En los Estados Unidos, prevalece una versión diferente del pluralismo. Basada en la idea de que cada persona alcanza su identidad a partir de los grupos primarios, no considera al hombre como un individuo por completo desvinculado, como en el liberalismo clásico. Como los grupos étnicos, religiosos, económicos y culturales determinan lo que es una persona, sus intereses deben expresarse por medio de los grupos más grandes a los que pertenece. Esta versión, denominada "pluralismo democrático" o "liberalismo conservador", acepta la tolerancia para con los distintos grupos establecidos, pero no respecto a los individuos idiosincrásicos. Se considera que, para el individuo, es bueno adaptarse a ciertos grupos, aunque, para éstos, es bueno que difieran entre sí.

En esta visión del pluralismo, en cuanto conflicto de grupos de intereses permanentes, la comunidad tiene un valor mucho mayor que en el liberalismo clásico. El pluralismo se basa en la diversidad de los grupos y no en la de

© Ediciones Morata, S. L.

Scanned by CamScanner

los individuos. La sociedad ideal se parece más a una ciudad pequeña que a una grande. La tolerancia y la aceptación mutua se ejerce entre los grupos.

Asociada con esta visión del pluralismo está la teoría del gobierno de la "suma de vectores" o del "equilibrio del poder". El objetivo consiste en conseguir que el gobierno actúe en una dirección determinada. El individuo debe operar por medio de sus grupos para lograrlo, dado que sólo ellos tienen influencia suficiente para conducir la nave del Estado. Como en el liberalismo clásico, el individuo debe estar cerca de su grupo, ser eficaz y comprometido en el mismo y participar en él. El gobierno (y el evaluador) actúa como mediador, amortiguando las instrucciones y respondiendo a los cambios de presión. En ambos tipos de pluralismo, la capacidad de considerar legítimas las reclamaciones encontradas es esencial para que el sistema funcione.

La evaluación "responsiva" de STAKE se acerca mucho a la versión del pluralismo democrático. El evaluador debe responder a cualesquiera intereses y presiones legítimos en torno al programa, pero no está obligado a representar punto de vista alguno ni a reflejar en la evaluación los puntos de vista, salvo que algún grupo de intereses participante en el programa promueva activamente alguno de ellos. La participación activa es un criterio para reflejar en la evaluación el punto de vista de un grupo.

Los grupos legítimos definen las cuestiones y sólo algunas se estudian. Han de escogerse métodos de investigación que se adapten a las cuestiones planteadas. Los problemas se resuelven mejor cuando se encargan de ello directamente personas próximas a la situación en la que se plantean (STAKE, 1975). Este planteamiento se opone al de la evaluación utilitarista en la que el gobierno evalúa, define el problema y emprende la acción. Las evaluaciones pluralistas tienden a favorecer al máximo la opción individual y local en vez de la utilidad social. Ambas formas de pluralismo suelen identificarse con la interacción libre entre individuos y grupos opuestos, como requiere el funcionamiento del mercado libre.

Aunque todos estos enfoques de la evaluación difieren en cuanto a los aspectos señalados, comparten algunos elementos básicos de la ideología liberal. Entre ellos está el principio de la libertad de elección. Aunque siga en pie el problema de quién elegirá, se favorece al máximo esa elección. Se fundan en una psicología muy individualista. Asimismo, su orientación es rotundamente empirista.

De igual modo, todos los enfoques de la evaluación dan por supuesto un mercado libre de ideas en el que, en un plano ideal, los usuarios seleccionarán lo mejor. Mediante la competición entre las ideas, se fortalece la verdad y se perfeccionan los servicios sociales. Por último, todos los enfoques asumen que un mayor conocimiento hace de alguna manera felices, mejores o satisfechas a las personas, suposición interesante que se remonta hasta la Ilustración.

© Ediciones Morata, S. L.

SEGUNDA PARTE

Los criterios fundamentales de la evaluación: veracidad, belleza y justicia

Cuando se trata de la conducta prudente en la vida, es bueno que tengamos presente que los acontecimientos humanos están dominados por el Azar y la Elección, que están extremadamente sujetos a cambios... En consecuencia, aquellos cuya única preocupación es la verdad abstracta experimentan grandes dificultades para poner los medios y mayores aún para alcanzar sus fines... Por tanto, dado que el curso de acción debe tener en cuenta la importancia de los acontecimientos únicos y sus circunstancias, puede suceder que muchas de estas circunstancias sean ajenas y triviales, algunas malas, otras contrarias, incluso, al propio objetivo. Es, pues, imposible valorar los asuntos humanos con arreglo al criterio inflexible de la justicia abstracta; debemos, en cambio, sopesarlos mediante las... reglas flexibles, que no adaptan los cuerpos a sí mismas, sino que se adaptan a sus siluetas.

Giambattista VICO.
On the Study Methods of Our Time, 1709, pág. 34.

Scanned by CamScanner

CAPÍTULO IV

La lógica de la argumentación evaluativa

Utilizo a propósito la palabra "discurso" porque las demostraciones científicas, incluso las pruebas matemáticas, son fundamentalmente actos de persuasión. Los enunciados científicos nunca pueden ser ciertos; sólo pueden ser más o menos creíbles.
Joseph WEIZENBAUM: *Computer Power and Human Reason*, 1976

Las generalizaciones desaparecen.
Lee J. CRONBACH: *Beyond the Two Disciplines of Scientific Psychology*, 1974

El gran terremoto que se va a producir en California

Vivo en Los Ángeles, pero me pregunto por qué. Ha surgido una repentina elevación de treinta centímetros y medio en una franja de ciento sesenta kilómetros a lo largo de la falla de San Andrés. Basándose en las lecturas de las ondas sísmicas, un científico de California ha previsto un importante terremoto en el área de Los Ángeles durante este año (*Science*, mayo 1976). Fundándose en otras lecturas, un predicador radiofónico advierte

---

El título original del capítulo es: *The logic of evaluative argument.* En inglés, la palabra *"argument"* significa "argumento", "razonamiento", "discusión", "debate", "alegato". En la cita de WEIZENBAUM, la palabra *"argument"* se utiliza en dos sentidos diferentes: "razonamiento" y "discurso persuasivo", derivándose esta última acepción de uno de los significados del verbo inglés *"to argue"*: "persuadir", que concuerda con el sentido de la cita inicial, la cual resultaría confusa traducida de otro modo. *(N. del T.)*

NOTA DEL AUTOR: Este capítulo es una versión reducida de *The Logic of Evaluative Argument*, Monografía n.º 7, Los Ángeles: *Center for the Study of Evaluation*, UCLA, 1977. Agradezco a Lee CRONBACH, Bob ENNIS, Gene GLASS y al personal del CSE sus detallados comentarios.

© Ediciones Morata, S. L.

Scanned by CamScanner

acerca de un importante temblor de tierra. Tanto los científicos como los videntes concuerdan en sus augurios. Sin embargo, ninguno facilita el tipo de información que necesito.

Hablo con los naturales del país acerca de estos signos de mal agüero. Su respuesta corresponde a la necesidad de vivir en esas circunstancias: se encogen de hombros. El Presidente ha sido informado, pero parece que nadie sabe qué hacer con exactitud. Los funcionarios de Washington sugieren que se instale un nuevo dispositivo de instrumentos científicos a lo largo de la falla, aunque no está claro qué resultados pueden proporcionar más medidas.

Mientras tanto, el tiempo es perfecto; el marco de las montañas de Santa Mónica, espléndido; el estilo de vida, sibarítico. Los cálculos de probabilidades de los seísmos a largo plazo no me benefician en absoluto; yo necesito saber cuándo temblará la tierra en relación conmigo mismo. El vocabulario de la acción es complejo. Todo el mundo está de acuerdo en que la información configura de algún modo las decisiones, pero la relación no es directa ni sencilla. A menudo, cuanto más importante es la decisión, más oscura parece la relación. Consideremos la decisión de contraer matrimonio. Para la mayoría, se trata de un proceso largo, arduo, que requiere un período de tiempo para ir tomando forma. Ninguna porción de información es decisiva al respecto. Al contrario, la decisión se fragua despacio, casi imperceptiblemente, hasta que se produce. Se esgrimen y ponen a prueba una razón tras otra. Por último, un conjunto de argumentos diversos se erigen en fundamento racional de la decisión que, con frecuencia, se ha tomado mucho antes de establecer todos los argumentos.

Las decisiones más significativas son las que tienen consecuencias a largo plazo, pero no permiten una extrapolación fácil, las que están tan enredadas con todo lo demás que se oponen a un análisis formal preciso. Ante ellas, nos vemos obligados a aplicar nuestra lógica intuitiva, nuestro sentido común. El hecho de que nuestros conocimientos de estos complejos problemas sea limitado, que no llegue a ser determinado, radica en su mismo carácter. Ante un conocimiento inseguro, la tarea de decidir en medio de esta maraña depende menos de que aparezcan pruebas que nos convenzan de manera absoluta que de que dispongamos de múltiples razones que nos persuadan. El criterio no estriba en lo necesario sino en lo admisible.

Equivocidad de las pruebas: certidumbre frente a credibilidad[1]

Entonces, ¿por qué los funcionarios gubernamentales, el público e, incluso, los miembros de la comunidad de evaluadores reclaman pruebas definitivas

[1]Respecto a esta distinción y a muchas otras ideas expuestas en este capítulo, estoy en deuda con la excelente obra sobre la argumentación de PERELMAN y OLBRECHTS-TYTECA: *The New Rhetoric: A Treatise on Argument.* Scranton (PA): University of Notre Dame Press, 1969.

© Ediciones Morata, S. L.

del éxito de los programas educativos? Hay una tradición que se remonta hasta DESCARTES que sostiene que el *único* conocimiento es el que posee la característica de cierto. El método de análisis de DESCARTES partía del escepticismo total: dudar de todo lo que pudiese dudarse. En su búsqueda de un saber seguro, llegó a lo *de por sí evidente* como última nota de la razón. Para que algo pudiese calificarse de saber, había que partir de ideas claras y distintas e ir progresando a base de pruebas deductivas. Las proposiciones así derivadas tenían que ser necesarias e irresistibles para el intelecto; no podían ser negadas, desde el punto de vista racional.

Este método impedía considerar como saber lo meramente creíble. Según el ideal cartesiano, el único razonamiento verdadero es analítico. La lógica deductiva formal, el método de prueba utilizado en matemáticas, constituye el método por excelencia. El saber puede reducirse a proposiciones de por sí evidentes. No son posibles los desacuerdos respecto al conocimiento cierto. Como decía DESCARTES, si dos hombres no se ponen de acuerdo sobre una materia, uno de ellos estará con seguridad equivocado. Hay una proposición verdadera y otra falsa y la lógica actúa mediante pruebas irrefutables para determinar cuál es cada cual.

Más tarde, quienes prolongaron esta línea de razonamiento tuvieron que enfrentarse con el hecho de que, a menudo, parecía que los hombres racionales pensaban de forma diferente y llegaban a conclusiones contradictorias. Algunas de las proposiciones del mismo DESCARTES parecían sospechosas. PASCAL introdujo la explicación de que estos desacuerdos, así como la negativa a aceptar conclusiones necesarias, era producto de la irracionalidad. Comenzó a verse en el hombre un lado irracional que, con frecuencia, le descarría en su búsqueda del saber. La aparente irracionalidad de quienes no aceptan las conclusiones que otros perciben como necesarias es moneda corriente en la evaluación contemporánea.

Desde la perspectiva cartesiana puede accederse al conocimiento cierto mediante procesos deductivos, y éste debe llevar a una convicción absoluta. Ese razonamiento funciona en geometría, pero lo hace porque excluye la mayor parte del mundo de los sentidos. Como señalaba HUME, nuestras creencias, aun en conceptos tan básicos como la causalidad, no son seguras si se les aplica un escepticismo absoluto. El razonamiento deductivo llega a producir un conocimiento cierto sobre todo mediante la eliminación de la mayor parte del mundo cotidiano.

Desde el punto de vista epistemológico, John Stuart MILL rescató para nuestros usos el mundo sensible. Del mismo modo que los lógicos han estructurado la lógica deductiva formal reflexionando sobre la naturaleza de las pruebas matemáticas, MILL reflexionó sobre la psicología asociacionista de su época y formuló una lógica inductiva que permitió la introducción de la certidumbre en el conocimiento derivado de manera inductiva. Para hacerlo, MILL estableció diversos axiomas que, aún hoy, están presentes en toda la investigación a base de encuestas. Según HAMILTON (1976), éstos son los axiomas:

© Ediciones Morata, S. L.

Scanned by CamScanner

- La naturaleza es uniforme en el espacio y en el tiempo. Esto brinda al razonamiento inductivo la misma certidumbre de procedimiento que a las conclusiones derivadas mediante la lógica silogística.
- Los conceptos pueden definirse por referencia directa a categorías empíricas y las leyes de la naturaleza puede derivarse de un modo inductivo de los datos por lo dicho anteriormente.
- Las muestras grandes pueden eliminar los aspectos idiosincrásicos y poner de relieve "causas generales".
- Las ciencias sociales y naturales tienen el mismo objetivo de descubrir leyes generales (que proporcionan la base de la explicación y de las previsiones).
- Las ciencias sociales y naturales son idénticas desde el punto de vista metodológico.
- Las ciencias sociales sólo son más complejas.

Por tanto, MILL defendía que el conocimiento cierto puede derivarse tanto del razonamiento inductivo, como del deductivo. Podríamos definir las categorías y relacionarlas entre sí mediante técnicas ahora bien conocidas. En realidad, MILL concluía que el método inductivo constituía la *única* forma de descubrir ideas nuevas, dado que la lógica deductiva sólo podría revelar lo que hubiera de antemano. (MILL estaba tan seguro de su método que afirmaba que los principios éticos también podían derivarse mediante el razonamiento inductivo y, por tanto, tenían un fundamento científico.)[2]

El axioma de verdad importante es el primero de MILL. Según sus propias palabras: "El universo, en la medida en que lo conocemos, está estructurado de manera que lo que en un caso es verdadero es cierto en todos los casos de una determinada descripción; la única dificultad estriba en descubrir de qué descripción se trata" (MILL, 1893). ¡Qué familiar resulta esa idea a

[2]Al final de su obra fundamental sobre la lógica inductiva, MILL considera la lógica de la "práctica" o "arte": "Debe haber alguna norma mediante la que determinar la bondad o maldad, absoluta y comparativa, de los fines u objetos de deseo. Y, cualquiera que sea la norma, sólo puede haber una; porque si los principios últimos de conducta fuesen diversos, la misma conducta podría ser aprobada por uno de estos principios y condenada por otro; y sería necesario algún principio general, como árbitro entre ellos" (John Stuart MILL: *A System of Logic*. Nueva York: Harper, 1893, 8.ª edición).

Esto lleva a MILL a imponer una única norma universal mediante la que juzgar los asuntos prácticos, porque la única alternativa es "suponer un sentido o instinto moral" o "principios morales intuitivos". Los principios éticos generales sólo pueden conocerse por inducción. Como la certeza inductiva presupone la uniformidad de naturaleza, la psicología resultante es determinista. La moralidad es natural, dado que sólo una valoración naturalista permitirá métodos científicos de prueba. La base es el utilitarismo hedonista.

En cierto sentido, MILL impedía el desacuerdo respecto a las cuestiones morales, ya que siempre es posible llegar a conclusiones opuestas en ausencia de un acuerdo previo sobre el criterio. El resultado de este razonamiento es el cálculo utilitario que combina todos los deseos humanos en una única configuración y los satisface mediante el criterio del máximo de satisfacciones plenas derivadas. El juicio lo lleva a cabo un "espectador imparcial" que, en los tiempos modernos, demuestra su imparcialidad empleando técnicas "objetivas" de análisis.

© Ediciones Morata, S. L.

quienes se dedican a la investigación mediante encuesta y qué falible la lógica inductiva en la que se basa!

El procedimiento de razonar desde "una parte" hacia el "todo" constituye sin lugar a dudas una falacia lógica. Se supone que cada caso confirmatorio hace más probable una hipótesis. Sin embargo, si la hipótesis es: "todos los hombres miden menos de 30 metros", y encontramos uno que mida 29,9 metros, éste sería un caso confirmatorio que debilitaría considerablemente la hipótesis, en vez de reforzarla (GARDNER, 1976). Cada día que pasa en Los Ángeles sin que se produzca el gran terremoto previsto, ¿lo hace más o menos probable? En los estudios estadísticos, también es muy posible que dos trabajos independientes confirmen por separado una hipótesis y que, sin embargo, reuniendo los resultados de ambos, quede refutada (véase la paradoja de Simpson en: Martin GARDNER, 1976).

No obstante, a pesar de sus graves defectos lógicos, parece que la "ciencia" basada en la lógica inductiva funciona con relativo éxito. Sin embargo, el conocimiento procedente de la misma carece de certidumbre. Incluso los hechos científicos mejor establecidos deben tomarse como provisionales. Como afirma un científico:

> Seguro que el hombre de la calle cree que tales hechos científicos están tan bien fundados como probada está su propia existencia. Su seguridad es una ilusión. Ni siquiera el científico resulta inmune a tal ilusión. A fin de cuentas, en su práctica debe mantener en suspenso la desconfianza para hacer o pensar algo. Es, más bien, como un espectador de teatro que, para participar en lo que sucede en el escenario y comprenderlo, tiene que fingir para sí que está presenciando acontecimientos reales. El científico debe creer su hipótesis de trabajo, junto con su amplia estructura subyacente de teorías y supuestos, aunque sólo sea como tal hipótesis. Con frecuencia, el "discurso" abarca toda su vida. Poco a poco, se convierte en lo que, al principio, sólo fingía que era: un verdadero creyente. Utilizo a propósito la palabra "discurso" porque las demostraciones científicas, incluso las pruebas matemáticas, son fundamentalmente actos de persuasión.
>
> Los enunciados científicos nunca pueden ser ciertos; sólo pueden ser más o menos creíbles. Y la credibilidad es un término que pertenece a la psicología del individuo, es decir, una palabra que sólo tiene sentido con respecto a un observador individual. Decir que una proposición es creíble es, en último término, decir que un agente, con libertad para no creerla, la cree, o sea, que un observador, tras ejercitar el juicio y (posiblemente) la intuición, opta por aceptar la proposición como digna de ser creída.
>
> (WEIZENBAUM, 1976.)

La evaluación como persuasión

Si en las ciencias físicas las demostraciones son fundamentalmente actos de persuasión, aún más las investigaciones sobre la educación. El axioma de MILL acerca de que las ciencias sociales y naturales son idénticas,

© Ediciones Morata, S. L.

Scanned by CamScanner

desde el punto de vista metodológico, hoy día parece mucho más dudoso. CRONBACH (1974), por ejemplo, duda de la conveniencia de imponer a las ciencias sociales los ideales de las ciencias físicas. En el paradigma de las ciencias físicas, los hechos se explican y prevén mediante "una red de proposiciones que ponen en relación constructos abstractos".

Tras revisar veinte años de estudios de las interacciones entre aptitud y tratamiento, basados en aquel modelo, CRONBACH concluye que los fenómenos sociales están demasiado abiertos a las interacciones con otras variables para permitir generalizaciones estables. La estrategia positivista de establecer condiciones que permitan conseguir generalizaciones da por supuestos unos procesos estables que puedan dividirse en sistemas independientes a efectos de estudio de los mismos, supuesto muy frágil en relación con los sistemas sociales.

CRONBACH invita a interpretar los datos en su contexto, en vez de tratar de llegar a generalizaciones. Un observador, situado en un ambiente concreto, puede describir e interpretar los efectos en las condiciones locales. Mientras que el control experimental y la correlación sistemática plantean preguntas formales de antemano, la observación local es más abierta ante los imprevistos. El empirismo a corto plazo es sensible al contexto, y así el investigador puede prescindir de cierta capacidad de previsión. Deja de efectuar generalizaciones y construir teorías, elaborando, en cambio, "conceptos que ayuden a pensar a las personas". He ahí lo que defiende CRONBACH.

Yo estoy dispuesto a sostener que, por sí mismas, las evaluaciones no son sino actos de persuasión. Aunque, a veces, los evaluadores prometen pruebas cartesianas y utilizan los métodos inductivos de MILL, inevitablemente, las evaluaciones carecen de la certidumbre de prueba y del carácter concluyente que les otorga el público. Si en realidad existe, la evaluación definitiva es rara. Incluso un metodólogo científico tan minucioso como James COLEMAN ha de hacer frente a continuas y cáusticas críticas de su trabajo. Si se someten a un examen serio, las evaluaciones siempre resultan equívocas.

Esperar que la evaluación arroje conclusiones definitivas y necesarias es pedir más de lo que la evaluación puede dar. Sobre todo en una sociedad pluralista, la evaluación no puede llevar a proposiciones necesarias. Pero, aunque no pueda acabar en conclusiones definitivas, sí puede llevar a lo creíble, lo plausible y lo probable. Sus resultados no llegan a ser ciertos, pero todavía pueden ser útiles.

Probar algo supone satisfacer el conocimiento de una audiencia universal con respecto a la verdad, más allá de toda duda. Producir una prueba aceptable para una audiencia universal que comprenda a todos los hombres racionales exige superar las particularidades locales o históricas. La certeza requiere aislar los datos respecto de su contexto total como, por ejemplo, en los términos de un silogismo. La certeza lógica sólo puede conseguirse dentro de un sistema cerrado, completamente definido, como un juego.

Si la evaluación se limita al conocimiento cierto proporcionado por un razonamiento deductivo e inductivo estricto, tiene que dejar de lado gran

© Ediciones Morata, S. L.

parte de la capacidad de razonamiento que las personas utilizan de forma habitual para regir sus vidas. Esa limitación es el resultado de confundir la racionalidad con la lógica. No son idénticas.

Si convencer de manera absoluta a todos los hombres racionales constituye una carga excesiva para la evaluación, no lo es persuadir a determinadas personas. En lugar de las proposiciones necesarias derivadas de una lógica rigurosa, podemos emplear los razonamientos no obligatorios de la persuasión. En lugar de la necesidad de la autoevidencia, podemos valernos de la adhesión variable a las tesis cuando se presentan a destinatarios concretos. Las tesis pueden ser más o menos creíbles. Los destinatarios gozan de libertad para creer o no creer tras estudiar los razonamientos y ejercitar su propio juicio.

La evaluación aspira a persuadir a un público concreto del valor de algo o de que ese algo es lo que interesa, apelando a la razón y la comprensión del público[3]. A estos efectos, el conocimiento no probado es útil, aunque siempre sean discutibles las ideas que se exponen. Los métodos adecuados son los de la argumentación, que es el ámbito propio de lo "creíble, lo plausible y lo probable" y no de lo necesario (PERELMAN y OLBRECHTS-TYTECA, 1969).

La argumentación se opone a la demostración. Las demostraciones se basan en la lógica formal que evita la ambigüedad por medio de la coherencia interna de su sistema de símbolos. En la lógica deductiva, el origen de los axiomas es externo al sistema. Cuando pasamos de la deducción a la inducción, todos los tipos de cuestiones se convierten en discutibles, como la validez de la medida. Pero sigue buscándose el conocimiento "cierto".

En la evaluación, los contextos social y psicológico cobran especial relieve y el conocimiento se hace menos seguro. En tales condiciones, la argumentación orientada a lograr la adhesión y a aumentar la comprensión de destinatarios concretos resulta más apropiada. La persuasión sólo proclama la validez respecto a determinados destinatarios, y la intensidad con la que esos públicos concretos aceptan los descubrimientos de la evaluación constituye una medida de su eficacia. El evaluador no aspira a convencer a

[3] KELLY (1980) ha puesto de manifiesto las premisas ocultas presentes en mis razonamientos aquí expuestos. Está implícito en ellos que el evaluador actúe para persuadir a los destinatarios de cierto punto de vista porque es "verdadero" y tiene algún medio para establecer esto. No obstante, la verdad puede mantenerse con distinta intensidad. KELLY afirma también —correctamente, a mi modo de ver— que hago la distinción aristotélica entre el argumento teórico que conduce a la verdad y el argumento práctico que lleva a la acción razonada.

No estoy tan seguro respecto a su afirmación de que la evaluación persuada a alguien para que actúe en vez de persuadirlo de que algo constituye el problema. La acción es el objetivo final de la evaluación, pero en la acción están involucradas tantas otras consideraciones que parece difícil que el evaluador pueda valorar e, incluso, identificar las principales contingencias. Me parece que la evaluación persuade de la conveniencia de algo. En ciertas circunstancias, esto puede constituir un curso de acción, pero, por regla general, la acción supone otras consideraciones adicionales.

© Ediciones Morata, S. L.

Scanned by CamScanner

una audiencia universal, compuesta por todos los hombres racionales, por la fuerza de la necesidad de sus conclusiones.

La persuasión está directamente relacionada con la acción. Aunque la información de la evaluación no sea tan segura como la información científica dirigida a una audiencia universal, la persuasión es eficaz para promover la acción porque se centra en un público concreto y suministra la información que afecta a ese público. El conocimiento personalizado que induce a las personas a dejar de fumar puede ser distinto de las generalizaciones científicas que vinculan el fumar con las enfermedades del corazón o con el cáncer. Es más fácil que el conocimiento del ataque al corazón de un pariente próximo sea el que nos induzca a actuar, y no todos los gráficos y tablas. El razonamiento evaluativo es a la vez menos seguro, más particularizado, más personalizado y más tendente a la acción que la información derivada de la investigación.

En resumen, la evaluación persuade más que convence, argumenta más que demuestra, es más creíble que cierta, la aceptación que suscita es más variable que necesaria. Esto no significa que se reduzca a mera oratoria o sea completamente arbitraria. El hecho de que no se limite al razonamiento deductivo e inductivo no supone que sea irracional. La racionalidad no equivale a la lógica. La evaluación emplea otras formas de razonamiento. Eliminada la carga de la certeza, las posibilidades de una acción informada aumentan, en vez de disminuir.

Los destinatarios de la evaluación

Si la persuasión se convierte en el objetivo de la evaluación, los destinatarios a quienes se dirige la evaluación son importantes. Desde hace muchos años, se aconseja a los evaluadores que piensen en sus destinatarios y en el tipo de información que ellos necesitan. Lo que tiene importancia para un grupo, quizá carezca de relieve para otro. La argumentación parte de la base de que existe una "comunidad de mentes", hay un contacto intelectual y, al menos, hay acuerdo respecto a algunas cuestiones sobre las que comenzar a deliberar.

Debe existir un lenguaje común y el deseo del evaluador de persuadir a los destinatarios y de tener en cuenta sus preocupaciones. A menudo, estas condiciones no se cumplen. El concepto que se tiene de los destinatarios es erróneo o no se los toma en serio. No es raro que el evaluador facilite información apropiada para un público compuesto por psicólogos pero casi carente de sentido para un profesor o un funcionario del gobierno.

Es fácil que el acuerdo de una audiencia universal (todos los hombres de todos los tiempos) se consiga mediante el razonamiento lógico formal basado en conceptos de por sí evidentes. Por tanto, cuanto más estricto sea el diseño experimental, más convencida quedaría una audiencia universal de la relación entre causa y efecto, con independencia del contexto. Unos

© Ediciones Morata, S. L.

destinatarios más cercanos a la escena pueden aceptar la causa y el efecto sin necesidad de tal prueba. Por supuesto, la audiencia universal no es "reunible" en ningún momento determinado, aunque, de hecho, diversos grupos escogidos actúan como representantes de aquélla. Quizá los filósofos, más que nadie, representen este tipo de audiencia. Los argumentos que manejan éstos no siempre coinciden con los que esgrimen los profesores.

En la medida en que un razonamiento se dirija a una audiencia universal, menos "discutible" será. Poca discusión cabe en pura lógica deductiva. Con frecuencia, las técnicas de evaluación se presentan como indiscutibles, basadas, por ejemplo, en instrumentos válidos y fiables, que utilizan procedimientos estadísticos fundados, etcétera. En realidad, todos los enunciados basados en una evaluación pueden suscitar desacuerdos y son discutibles, si el desacuerdo se manifiesta adecuadamente. Cuanto más técnica y cuantitativa sea la evaluación, menos capaces de discutirla serán unos destinatarios no iniciados, pareciendo, en tal caso, que la evaluación es más incontrovertible de lo que en realidad es.

En las evaluaciones que utilizan metáforas estadísticas, podemos decir que los efectos del tratamiento difieren porque existe la probabilidad de que dos puntuaciones medias del test pertenezcan a poblaciones diferentes y, por tanto, que el programa experimental es mejor que el control. El uso generalizado de números en los procedimientos estadísticos y las puntuaciones de los tests proporcionan a las pruebas una apariencia de certeza e inequivocidad.

En realidad, detrás de los números se esconden muchos supuestos previos (como, por otra parte, detrás de toda evaluación). Casi siempre, es posible discutir la validez de los tests, la adecuación de los procedimientos estadísticos y el control del diseño experimental. La discusión no invalida la evaluación, pero, cuando se discuten las premisas, se pone en evidencia el carácter de la evaluación como discurso persuasivo. El evaluador puede o no tener éxito en la defensa de su estudio. En todo caso, para defenderlo ha de asumir un razonamiento no deductivo y más equívoco. Aunque, en apariencia, la evaluación apele a la racionalidad definitiva de la audiencia universal, termina en apelaciones directas a destinatarios concretos. Creo que es imposible estructurar de otra forma una evaluación.

Incluso una operación de evaluación de base muy amplia, como la del *Consumers Report*, que utiliza procedimientos "objetivos" y diseños experimentales complejos para evaluar productos de consumo, se orienta a unos destinatarios determinados. Sus argumentos, dirigidos a la clase media alta, no tienen mucho sentido para las clases inferiores o superiores, que no suelen tener muy en cuenta sus evaluaciones.

Por tanto, la situación a la que se enfrenta el evaluador consiste casi siempre en una llamada a públicos concretos que puede definir con cierta precisión. Si el evaluador no puede definir a sus destinatarios, la evaluación será indeterminada. Debe tratar problemas y estructurar argumentos que digan algo a públicos concretos. Más aún, es probable que los destinatarios constituyan un combinado de grupos diversos, lo que complica considera-

© Ediciones Morata, S. L.

Scanned by CamScanner

blemente su tarea. La orientación eficaz a determinados destinatarios modifica los límites de la racionalidad aplicable. No está confinado a las formas más restringidas de razonamiento. Cuando la evaluación se hace más equívoca, también aumenta su grado de posibilidad.

El diálogo socrático constituye un ideal de la argumentación entre dos partes. El diálogo se desarrolla como una cadena rigurosa de razonamiento entre alguien que pregunta y alguien que responde. El destinatario unipersonal es persuadido llevándole a aceptar paso a paso ciertos principios. En último término, en la interacción se ponen sobre el tapete las preocupaciones concretas del destinatario. Asimismo, el diálogo socrático encierra también gran fuerza respecto a terceros que puedan leerlo.

No obstante, el público al que se enfrentan la mayoría de los evaluadores no suele reducirse a una sola persona. Es mucho más corriente que esté formado por diversos grupos diferentes. Algunos teóricos han sugerido formas de evaluación en las que el evaluador mantenga frecuentes cambios de impresiones con los destinatarios durante el estudio. Con independencia de la forma de la evaluación, estoy convencido de que la que tiene éxito en su faceta persuasiva debe implicar a los destinatarios en el discurso fundamental, aunque ese discurso pueda desarrollarse de maneras muy diferentes.

El discurso que se desarrolla de este modo es más que un simple debate en el que los correspondientes partidarios expongan sus distintos puntos de vista. El diálogo debe consistir en una discusión en donde las partes busquen de manera seria y sincera respuestas mutuas. Esta restricción descalifica en gran medida el empleo de los métodos contradictorios en cuanto instrumentos persuasivos, dado que puede suscitarse un conflicto sin persuadir a nadie de nada.

KEMMIS (1976) pretende que el destinatario sea uno mismo: "la evaluación como autocrítica". Considera que los destinatarios primordiales han de ser quienes intervienen en el programa. Convencido de que el único camino para perfeccionar el ejercicio profesional consiste en la dialéctica entre el conocimiento y la acción, señala que las normas de la evaluación deben depender de los mismos participantes en el programa y que los datos consistan en los progresos que perciban los participantes. Así, la evaluación se convierte en autocrítica terapéutica. El objetivo final es que los participantes consigan un mayor conocimiento y una idea más clara que les conduzca a una acción eficaz.

Con independencia de quiénes sean los destinatarios, cuando se trata de argumentaciones, éstos deben compartir la responsabilidad. Como la información no es definitiva, los destinatarios gozan de libertad para establecer su grado de compromiso. Deben escoger activamente en qué medida están dispuestos a creer, lo que exige que comprueben la evaluación de un modo activo en vez de aceptarla o rechazarla de forma pasiva. Los destinatarios tienen que comprometerse de modo personal y compartir la responsabilidad. La decisión racional corresponde a los destinatarios y no al evaluador.

© Ediciones Morata, S. L.

Las premisas del acuerdo

La elaboración de la argumentación evaluativa parte de la base del acuerdo entre los destinatarios. Las premisas de la argumentación constituyen el principio de este acuerdo y el punto de partida de acuerdos más amplios. Del mismo modo que el sentido común admite verdades intocables e indiscutibles, algunas de las premisas principales de una evaluación son tácitas y no explícitas.

Según PERELMAN y OLBRECHTS-TYTECA (1969), hay dos clases de premisas: las "reales" y las "preferibles". Entre las reales están los hechos, las verdades y los supuestos básicos y, por regla general, la presunción de validez respecto a una audiencia universal. Por otra parte, las preferibles se identifican con unos destinatarios concretos y abarcan valores, jerarquías de valores compuestas y premisas de valor de carácter muy general, denominadas *"loci"*.

Los hechos y las verdades son datos e ideas que se consideran aceptadas por la audiencia universal, es decir, compartidas por los seres pensantes y, por tanto, sin necesidad de justificación. Que un dato constituya un hecho depende de la propia concepción de la audiencia universal. Si ésta cambia, también los hechos y verdades. No obstante, el mantenimiento de la categoría de hecho o verdad significa que, a los efectos del razonamiento, el dato es incontrovertible e indiscutible. Si se cuestiona el dato, pierde su categoría de hecho y pasa a ser tema de argumentación en vez de objeto de acuerdo.

Cuando existe acuerdo sobre las condiciones de verificación, como en la ciencia moderna, pueden darse muchos hechos. La ciencia moderna no concede la categoría de "hechos" a muchos datos. POLANYI (1958) pone de manifiesto cómo la ciencia protege de la incoherencia a sus propios sistemas de creencias mediante la negación de diversos datos que se oponen a otras creencias que tienen el carácter de hechos. Así, durante muchos años, la ciencia descartó que los efectos hipnóticos se produjesen de ningún modo. No se reconocía el carácter de hechos de los datos en cuestión porque entraban en contradicción con el sistema científico de creencias entonces vigente. Este sistema de creencias se modifica cada cierto tiempo, pero, con independencia de lo que excluya, los argumentos que se produzcan en el marco del sistema de creencias deben basarse en hechos y verdades indiscutibles.

Los argumentos también proceden de supuestos básicos que carecen de la autoridad y fiabilidad plenas de los hechos y verdades. Los supuestos básicos no pueden probarse, no obstante sean aceptados de modo bastante general como provisionalmente verdaderos. Muchos supuestos básicos están relacionados con el concepto de lo normal. En las evaluaciones que emplean modelos y metáforas estadísticos, se acepta de forma casi universal el supuesto básico que atribuye una distribución normal a una población.

La segunda clase de objetos de acuerdo es la de lo preferible. Los objetos preferibles sólo reclaman la adhesión de grupos concretos y no de

© Ediciones Morata, S. L.

Scanned by CamScanner

una audiencia universal. Los valores constituyen los ejemplos más sobresalientes. El acuerdo respecto a un valor supone admitir que existe una influencia específica sobre la acción o una disposición a la acción que el evaluador puede utilizar. Aunque sea importante para un grupo determinado, no se considera que el valor vincule a nadie en concreto.

En la ciencia, los valores aparecen ante todo al seleccionar los objetos de interés para la investigación, ya que es imposible investigar el mundo entero (POLANYI, 1958) y, quizá, en la aceptación de las conclusiones científicas por medio del juicio humano global (WEIZENBAUM, 1976). Pero, en el desarrollo de la mayor parte de la argumentación, sobre todo en las ciencias exactas, los valores se dan por excluidos. El análisis de las relaciones entre causa y efecto de ENNIS (1973) nos lleva a cuestionar esto mismo. En la evaluación, nada impide que los valores aparezcan a cada momento. Los valores se utilizan para persuadir a los destinatarios y para justificar las opciones que se hagan ante los demás.

Diversas combinaciones de argumentos pueden compendiarse en unas pocas agrupaciones generales denominadas *"loci"* (PERELMAN y OLBRECHTS-TYTECA, 1969). Los *loci* más comunes son los de cantidad y calidad. Los argumentos agrupados en torno a los *loci* de cantidad afirman que una cosa es mejor que otra por razones cuantitativas (mayor número, mayor grado, mayor durabilidad, etc.). Con frecuencia, la eficacia de los medios se justifica mediante *loci* cuantitativos. Las ideas de "normal" y "norma" también se basan en la cantidad.

En contraste con la cantidad, está la idea de calidad. Hay cosas que ostentan un valor elevado a pesar de su pequeño número. La valoración elevada de lo único se asocia a la calidad. Un individuo puede estar en posesión de la verdad mientras la multitud permanece en el error. Por ejemplo, SCRIVEN (1972) defiende que la noción de objetividad no tiene por qué estar vinculada a la cantidad de personas que sostienen determinada idea, ni la de subjetividad a la propia percepción de una persona, como se piensa a menudo.

Junto a los razonamientos generales sobre hechos y valores, existen acuerdos especiales restringidos a determinados destinatarios especiales y propios de cada evaluación. En la medida en que la evaluación se dirige a unos destinatarios técnicos, éstos compartirán determinados acuerdos y convenciones. Un grupo de investigadores de la educación constituye un conjunto de destinatarios técnicos de este tipo. Las evaluaciones que se dirigen a unos destinatarios inexpertos en la materia no pueden basarse en los mismos acuerdos.

Es posible que los acuerdos peculiares más importantes de una evaluación concreta se deriven de la negociación que con frecuencia precede a la evaluación (acuerdos entre patrocinadores, personal que participa en el programa y evaluadores). En esta importantísima negociación, pueden alcanzarse acuerdos sobre los criterios, métodos y procedimientos, acceso a los datos, divulgación de resultados, etcétera. Los desacuerdos sobre estos aspectos pueden acabar completamente con la credibilidad de la evaluación.

© Ediciones Morata, S. L.

En resumen, al principio de la evaluación el evaluador debe basarse en los acuerdos con los destinatarios. Estos acuerdos pueden ser tanto implícitos como explícitos. En realidad, sería imposible especificar con minuciosidad todos los aspectos, aunque es peligroso dar por supuesto el acuerdo sobre temas importantes cuando no existe. El evaluador debe partir de la postura de sus destinatarios, aunque las premisas iniciales no sean aceptables para terceras partes ni para el mismo evaluador siquiera. En caso contrario, la evaluación no será creíble ni persuasiva. Como mínimo, debe darse una comprensión compartida de la situación. Si discrepan en exceso los valores básicos, el evaluador puede optar por no llevar a cabo el estudio. Por supuesto, esta comprensión básica está sometida a las concepciones de decencia y justicia que prevalecen en la sociedad en conjunto, y el evaluador puede basarse en estas concepciones sociales más generales.

Todo esto no significa que el evaluador tenga que estar de acuerdo en todo con sus destinatarios. Es probable que existan áreas de desacuerdo, pues, si no, no harían falta argumentos. Es posible que los destinatarios quieran aprender algo nuevo o, de lo contrario, no haría falta evaluar. Pero la evaluación parte de áreas en las que existe el acuerdo, para alcanzar aquellas otras en las que el acuerdo es problemático.

La argumentación cuantitativa

El enfoque más popular de la evaluación es el cuantitativo. Algunos lo consideran como la auténtica esencia de la racionalidad y del método científico. Ha producido muchos buenos estudios de evaluación (y muchos malos también). Como este enfoque se enseña en los centros superiores y se promueve en la bibliografía, no necesitamos extendernos en sus virtudes, que son muchas. En este apartado pretendo mostrar que incluso la metodología cuantitativa consiste esencialmente en argumentos y está sometida a similares consideraciones. Utilizada de manera adecuada, puede ser una valiosa herramienta de análisis; empleada de forma inadecuada, es peligrosa.

La metodología cuantitativa constituye un cuerpo de métodos matemáticos y técnicas de medida a disposición del evaluador. La utilidad de la metodología depende de las semejanzas entre los problemas teóricos de los que ésta se ocupa y los problemas efectivos a los que se enfrenta el evaluador en un medio concreto. Por su parte, CRONBACH (1974) ya ha señalado que el ajuste entre los problemas teóricos y efectivos no es bueno. El contexto educativo es demasiado complejo.

Un matemático de la *Rand Corporation* (STRAUCH, 1976) examinó las dificultades que presenta la metodología cuantitativa cuando se aplica a estudios políticos, o sea, a cuestiones que surgen a partir de los procesos gubernamentales de toma de decisiones. Según STRAUCH, en la medida en que la metodología es matemática, la estructura de lo determinado por las

© Ediciones Morata, S. L.

Scanned by CamScanner

premisas que definen el sistema constituye un sistema cerrado*. El análisis matemático consiste en el estudio de esa estructura, tal como se deriva lógicamente de las premisas. Los resultados están conectados con las premisas mediante inferencias lógicas. En el sentido en que su validez puede determinarse sobre la base de esa cadena de razonamiento, los resultados son "objetivos" (no hace falta apelar a la competencia o al juicio de la persona que los obtuvo ni a sus destinatarios). Los resultados son necesariamente lógicos. En la argumentación, en cambio, no pueden separarse por completo de la persona que los obtiene.

La aplicación de la metodología cuantitativa a problemas efectivos utiliza un modelo matemático como representación simplificada del problema. Los resultados dependen, en parte, del análisis matemático, pero también del ajuste entre el modelo y el problema real. En las aplicaciones más sencillas, como en las ciencias físicas, los problemas reales son rigurosamente cuantificables. El control experimental refuerza la capacidad del evaluador para hacer que el problema real se adapte al modelo matemático, por ejemplo, la selección aleatoria en los modelos estadísticos. En esos casos, las conclusiones son "objetivas" en el sentido de que están sujetas a verificación independiente sobre la base de la lógica y del ajuste, sin referencia al juicio de la persona que las elaboró. Sin embargo, cuanto más conductual o político sea el problema real, más difícil resulta definirlo sin ambigüedad en términos matemáticos. Las relaciones entre la realidad y el modelo se hacen más débiles.

En un estudio cuantitativo del tipo mencionado, STRAUCH identifica los siguientes componentes: la *formulación* supone definir el problema formal a partir del problema real, estableciendo después un modelo matemático del problema formal. Se trata de un procedimiento de reducción. El *análisis* supone el cómputo en el contexto matemático definido por el modelo. Se traduce en enunciados matemáticos. La *interpretación* consiste en la reintegración de los enunciados al problema formal; y, por último, la interpretación de estas conclusiones depende *tanto* de la validez lógica del análisis como de la validez de las relaciones entre el modelo y el problema real. Aunque la validez lógica puede determinarse sin referencia al juicio subjetivo del analista, no así las relaciones. Estas últimas se fundan en los juicios subjetivos del analista. Tanto la formulación como la interpretación son procesos subjetivos. La formulación requiere reducir las dimensiones del problema real

* El término utilizado en el original es *"self-contained"*, traducido a veces como "autocontenido", "independiente", "que tiene sus propios recursos". Optamos por la expresión "sistema cerrado" porque refleja de manera más clara la intención del autor, sobre todo si tenemos en cuenta la importancia que, para el estudio de los sistemas sociales, tiene la "teoría general de los sistemas" de Ludwig VON BERTALANFFY *(General Systems Theory. Foundations, Development, Applications*, Nueva York: George Braziller, 1968. Trad. cast.: *Teoría general de sistemas*, México: Fondo de Cultura Económica) y, más en concreto, la "dinámica de sistemas" de Jay W. FORRESTER (*Industrial Dynamics*, M.I.T. Press, 1961). En este contexto, "sistema cerrado" es aquél que no interactúa con variables externas a sí mismo, como sucede en el sistema al que se refiere STRAUCH. *(N. del T.)*

© Ediciones Morata, S. L.

para hacerlas manejables mediante el análisis, añadiendo, quizá, algunos supuestos que hagan más fácil el análisis, aunque pueda resultar discutible en la perspectiva real, por ejemplo, la independencia de los acontecimientos.

La interpretación supone restaurar las consideraciones contextuales que habían sido eliminadas y, posiblemente, ajustarlas a causa de los supuestos simplificadores. Tanto la formulación como la interpretación requieren dosis considerables de juicio intuitivo. En consecuencia, las conclusiones no son, en realidad, tan "objetivas" como se afirma (véase el tratamiento de la objetividad que hacemos en un apartado posterior).

La forma usual de ocuparse del aspecto subjetivo de la metodología consiste en dejarlo de lado. En las ciencias naturales, en donde los métodos cuantitativos han cosechado tantos éxitos, esto no constituye un gran problema. La evidencia de "objetividad" en esos campos se toma como prueba de objetividad en otras áreas. Cuando se cuestiona esta relación, empieza a quedar claro que las premisas que le sirven de fundamento son muy discutibles.

A menudo, de los estudios cuantitativos se han derivado perspectivas interesantes, pero éstas suelen deberse más a los juicios intuitivos correctos del analista que a los cálculos bien hechos. Estos éxitos se atribuyen con facilidad a la metodología cuantitativa en vez de al juicio. Las críticas suelen centrarse más en la calidad técnica del análisis matemático que en la calidad de los juicios asociados a la formulación e interpretación. Cuando se discute la calidad del juicio, la justificación debe basarse en el tipo de razonamiento común a toda argumentación.

Un resultado de infravalorar el papel del juicio es lo que podemos denominar, con STRAUCH, "análisis orientado al método". El analista deja de lado las complejidades del contexto y se sumerge en su método preferido. Vista en una perspectiva superficial, la metodología se aplica de forma directa, como si no existiesen problemas de ajuste. Al final, se plantean unas cuantas advertencias, indicando que compete a los lectores decidir si el ajuste es bueno.

En su forma extrema, hay una escuela de pensamiento, que STRAUCH denomina "cuantificacionismo", que sostiene que la cuantificación es un valor positivo en sí misma. Una respuesta cuantitativa siempre es mejor que otra cualitativa. Cualquier problema puede reducirse a una solución cuantitativa y ningún problema llega a comprenderse de manera adecuada hasta que no se consigue. Por tanto, los métodos cuantitativos deben aplicarse a todos los problemas. Es probable que esta postura sea falsa, dado que, en realidad, pocas personas la suscribirían.

Esa actitud, que favorece la metodología "científica", se basa en un reduccionismo que trata un fenómeno como un sistema aislado, elabora un modelo cuantitativo para ese sistema y utiliza el modelo como sustituto del fenómeno. Como señalamos antes, el reduccionismo puede ser un elemento de las ciencias físicas intransferible a los fenómenos sociales.

La imagen de los proyectos cuantificacionistas es la de un proveedor de "hechos" objetivos basados en datos contundentes. No se responsabiliza personalmente de las conclusiones a las que puedan llegarse mediante su

© Ediciones Morata, S. L.

Scanned by CamScanner

metodología, dado que no son suyas. Sólo se limita a descubrirlas y a informar de los resultados de sus métodos objetivos. Desdeña los datos cualitativos como subjetivos.

Esta actitud se aproxima a lo que POLANYI (1958) describía como "objetivismo" en la ciencia. Es un intento para definir un método objetivo que elimine la responsabilidad del observador respecto de sus hallazgos. POLANYI defiende, en cambio, que una convicción requiere el compromiso y la responsabilidad personales incluso en la ciencia. El objetivismo ha tratado de representar el conocimiento científico como totalmente impersonal.

A menudo, el cuantificacionismo y el objetivismo convienen también al responsable de las decisiones en la medida en que éste puede justificarlas mediante la referencia a un descubrimiento "científico". Esto puede ayudarle a eludir su responsabilidad personal. Las tentativas de cuantificación de problemas no cuantificables y de pasar por alto los factores judiciales acaban deformando el proceso de decisión.

Indica STRAUCH que un modo de eliminar esa deformación consiste en utilizar los métodos cuantitativos como *perspectivas* respecto del problema real, en vez de como *sustitutos*. La aceptación del modelo matemático como representación válida del problema real supone utilizarlo como sustituto. El empleo del modelo para integrar hallazgos con el *conocimiento que ya se posee* significa utilizarlo como perspectiva.

Respecto a la mayor parte de los problemas reales, los destinatarios de la evaluación ya poseen imágenes bien elaboradas por sí mismos. El análisis cuantitativo puede proporcionar a los destinatarios una visión adicional, aunque no necesariamente mejor ni más válida, del problema. La interacción entre las propias imágenes y las perspectivas adicionales debe producirse en el interior de cada destinatario, de cada responsable de decisiones o de quien corresponda. Si sólo se utiliza la metodología cuantitativa como una perspectiva, se reduce el problema del ajuste entre el modelo y el problema real.

Por otra parte, tanto el evaluador como los destinatarios deben responsabilizarse de modo personal de los hallazgos, dado que no se derivan necesariamente del análisis. Las conclusiones no pueden justificarse por completo sobre la base de que se derivan, en buena lógica, de los supuestos de partida. La evaluación de cada uno de los supuestos debe completarse con la evaluación global del total.

Por tanto, la argumentación cuantitativa debe utilizarse siempre en conjunción con el juicio humano y éste debe ocupar una posición superior. Las consecuencias para el argumento cuantitativo en la evaluación son importantes. La metodología cuantitativa debe considerarse fundada en los juicios humanos y en el razonamiento intuitivo, justificándose en consecuencia.

La argumentación cualitativa

En su artículo sobre el conocimiento cualitativo, CAMPBELL (1974) indicaba que el conocimiento científico depende del sentido común y que los hechos

© Ediciones Morata, S. L.

concretos, sean de la ciencia o del sentido común, sólo llegan a conocerse en el conjunto de muchos otros hechos. "La proporción entre lo fiable y lo dudoso es siempre una fracción muy pequeña". En realidad, cualquier conocimiento de algo depende del contexto y, según CAMPBELL, el conocimiento cualitativo de "totalidades y pautas" proporciona el contexto necesario para interpretar los datos cuantitativos. Por ejemplo, la elaboración de hipótesis alternativas requiere conocer bien el medio local, un acto cualitativo.

CAMPBELL piensa que se ha dejado de lado el conocimiento cualitativo en beneficio de los métodos cuantitativos. Le gustaría que se utilizaran conjuntamente los métodos cualitativos y cuantitativos con el fin de efectuar validaciones mutuas entre ambos tipos. Cree que los métodos cuantitativos pueden proporcionar perspectivas que los cualitativos no son capaces de facilitar, a pesar de la mayor fundamentación de estos últimos. Asimismo, como todo conocimiento es esencialmente comparativo, cree que las técnicas cualitativas, como los estudios de casos, pueden mejorarse mediante consideraciones relativas al diseño experimental, que, a su modo de ver, no forman parte exclusiva de la metodología cuantitativa.

Al sopesar de nuevo la necesidad e, incluso, la prioridad del saber cualitativo, CAMPBELL (1975) revisa la "observación naturalista, anecdótica y de caso único". La generalización cuantitativa se opondrá en algunos aspectos a este conocimiento, pero sólo por confiar en un conjunto mucho mayor de observaciones de ese tipo. En su clásico documento sobre el diseño experimental, CAMPBELL y STANLEY (1966) mencionan el estudio de casos, considerando que carece de base de comparación y, en consecuencia, de justificación para extraer inferencias causales.

CAMPBELL ha modificado considerablemente su postura y ahora considera que el estudioso de casos hace muchas previsiones sobre la base de su teoría, que puede quedar invalidada. Se trata de un proceso de "ajuste del modelo" en el que los distintos aspectos de éste se confrontan con las observaciones del medio local. En la actualidad, CAMPBELL cree que el estudio de caso único constituye una base de conocimiento más segura de lo que pensaba antes.

Según CAMPBELL, ¿cómo podemos conocer algo? A partir de las dificultades epistemológicas actuales, busca un fundamento seguro del conocimiento. El esfuerzo para "eliminar la equivocidad, basando el conocimiento en datos sensoriales concretos y el espíritu del atomismo lógico se orientan en el mismo sentido de búsqueda de la certeza en los aspectos concretos" (CAMPBELL, 1966). La certeza se lograría mediante la definición de "aspectos concretos incorregibles". Esto se traduciría en términos especificables de modo inequívoco y en una "certeza de comunicación".

En la actualidad, CAMPBELL cree que esta forma de positivismo es insostenible, tanto desde el punto de vista de la filosofía como de la psicología. Las cosas fuera de contexto no son interpretables. Pero, ¿cómo podemos aún conocer algo partiendo de un conjunto de acontecimientos, cada uno de los cuales es de por sí indeterminado? CAMPBELL afirma que esto se consigue mediante el "ajuste del modelo".

© Ediciones Morata, S. L.

Scanned by CamScanner

En acontecimientos relativos a la cognición, como la visión binocular, los ojos reconocen los objetos corrientes mediante un proceso de triangulación. Cuanto más elaborado es el modelo, más improbable resulta un error de reconocimiento, desde el punto de vista estadístico. Por medio de la memoria, pueden compararse varios modelos. CAMPBELL considera el ajuste del modelo como un proceso de ensayo y error. Se trata esencialmente de pensamiento analógico, siempre presente, según CAMPBELL, en el proceso de conocimiento.

En realidad, la teoría científica constituye la instancia última del conocimiento y la relación entre la teoría formal y los datos es de ajuste del modelo, con un error asignado a la medida de los datos (puntuaciones "verdaderas" y puntuaciones "estimadas"), salvo cuando se está de acuerdo en la necesidad de revisar la teoría. Hay que ajustar dos modelos: el de la teoría y el de los datos. La aceptación o rechazo de la teoría está sometida a algún criterio de ajuste entre ambos. En realidad, nunca se rechaza una teoría a causa de su falta de ajuste, salvo si existe una teoría alternativa que la sustituya. Lo que hace que una teoría sea "correcta" es la falta de hipótesis rivales aceptables.

CAMPBELL piensa que estas consideraciones se aplican de forma directa a las cuestiones relativas a la evaluación de programas. "Creo que los problemas de equivocidad de las pruebas respecto a la eficacia de los programas son tan afines a los problemas generales de inferencia científica que nuestras extrapolaciones a recomendaciones sobre procedimientos de evaluación de programas pueden estar bien fundadas, con la adecuada crítica mutua."

Si entiendo bien su postura, CAMPBELL afirma que la evaluación constituye un aspecto de la investigación científica y está sujeta a problemas epistemológicos semejantes. Aunque pueda ser cierto, al menos, en este capítulo, he invertido de alguna manera la relación entre figura y fondo, considerando la ciencia como una argumentación dirigida a una audiencia universal que, por tanto, se ocupa del establecimiento de generalizaciones a largo plazo, y la evaluación como una argumentación que se dirige a destinatarios concretos, ocupándose de cuestiones relacionadas con un contexto determinado.

Podemos considerar que, en la evaluación, el ajuste del modelo no sólo se produce en la mente del evaluador, cuando estructura su estudio y examina el ajuste entre su descripción del programa y el programa en sí, sino también en las mentes de los destinatarios, cuando comparan el estudio de evaluación con su propia experiencia. Los destinatarios también tienen imágenes, recuerdos y teorías sobre el programa sometido a evaluación. Utilizando la evaluación como perspectiva (un modelo verbal, en este caso), los destinatarios comparan ésta con su concepción del programa. Los errores se atribuirán dependiendo de la capacidad de persuasión de la evaluación. De este modo, los destinatarios operan como hitos independientes de validación de la evaluación, debiendo asumir un papel activo en su interpretación y responsabilizándose personalmente de esa interpretación.

© Ediciones Morata, S. L.

Según CAMPBELL, el proceso básico de ajuste del modelo es más analógico que lógico (aunque no cabe duda de que el proceso lleva consigo muchas formas de razonamiento). En realidad, podemos llegar más lejos. En una epistemología basada en la eliminación de la equivocidad y en el establecimiento de la certeza del conocimiento mediante la definición de "aspectos concretos incorregibles", el razonamiento deductivo y el inductivo constituyen el modo adecuado de relacionar esos aspectos. La lógica formal depende de términos unívocos que operen en un sistema cerrado.

En la medida en que los términos sean ambiguos y el sistema abierto (o no reducible a subsistemas aislados), la lógica formal sólo puede aplicarse de manera argumentativa. El razonamiento ha de incluir otras formas de pensamiento, salvo que aceptemos la imposibilidad del análisis racional. En la evaluación éste es posible, pero sólo en raras ocasiones adoptará una forma de silogismo[4].

Revisión de la objetividad, validez e imparcialidad

¿Qué quiere decir que un estudio de evaluación sea "objetivo" o "válido"? Pocos conceptos se han confundido tanto y han provocado tanta desorientación. Muchas personas se muestran reacias a aceptar o a creerse las evaluaciones cualitativas sólo porque se basan en las observaciones de una sola persona. Se piensa que estas observaciones son en sí y de por sí subjetivas y, en consecuencia, carentes de justificación a efectos públicos.

La clave de la confusión radica en la errónea concepción de la "objetividad". SCRIVEN (1972) se ha ocupado espléndidamente de esta confusión, poniendo de manifiesto la desafortunada historia de la definición de la objetividad. El argumento fundamental de la mayoría de las definiciones de objetividad consiste en que hay algo externo a la mente que puede verificarse mediante el acuerdo público e intersubjetivo, y que podemos expresar o probar tales cosas sin la influencia de los sentimientos personales. Una evaluación que pueda operar así es objetiva. Pero, ¿acaso el punto de vista de una persona puede ser "objetivo"? La dificultad radica en la confusión de la objetividad con los procedimientos para determinar la intersubjetividad.

SCRIVEN (1972) sostiene que "objetividad" se utiliza en dos sentidos diferentes: el cuantitativo y el cualitativo. En el sentido cuantitativo del término, la opinión que una persona tenga de algo se considera subjetiva (se trata de la disposición de un individuo). La objetividad se consigue a través de las experiencias de un conjunto de sujetos u observadores. La experiencia común hace pública la observación mediante el acuerdo intersubjetivo. En un plano más formal, podemos decir que el conjunto de individuos nos

[4] Véase en el Apéndice A un análisis ampliado de la evaluación como argumentación. En el Apéndice B, puede examinarse un análisis de la "evaluación naturalista".

© Ediciones Morata, S. L.

Scanned by CamScanner

garantiza una representación más adecuada de la población (un problema de muestreo).

El sentido cualitativo de la objetividad es muy diferente. Se refiere a la calidad de la observación, con independencia de la cantidad de personas que la lleven a cabo. Se denomina objetiva la observación que se atiene a los hechos, mientras que se considera subjetiva la que de algún modo es tendenciosa. ¿Puede darse el caso de que las observaciones de una persona se atengan a los hechos concretos, mientras que las realizadas por un conjunto de individuos no lo sean? En efecto. Por tanto, una observación puede ser subjetiva, desde el punto de vista cuantitativo (la opinión de un individuo), y al mismo tiempo objetiva, desde el punto de vista cualitativo (realmente carente de sesgos y veraz).

En realidad, hemos de aceptar que los tipos de inclinaciones tendenciosas que influyen en la opinión de una persona difieren de algún modo de los sesgos propios de las opiniones grupales. Por ejemplo, un individuo puede sucumbir con mayor facilidad ante los puntos de vista idiosincrásicos, dado que sólo es capaz de tener una perspectiva. Por otra parte, existen sesgos sociales y culturales ante los que un grupo se muestra más susceptible que una persona concreta, p. ej., el chauvinismo. La objetividad cualitativa del individuo puede evaluarse por su trayectoria en relación con estas cuestiones y por sus intereses personales actuales. En todo caso, quien acepte total y exclusivamente la idea cuantitativa de objetividad nunca estará satisfecho con enfoques del estilo del estudio de casos.

¿Cómo llegó la idea cuantitativa de equiparar la cantidad de individuos que efectúan una observación con su veracidad para adquirir un ascendiente tal que excluyese la objetividad cualitativa? SCRIVEN remonta esta deformación hasta los intentos de la psicología para erradicar la introspección y las tentativas de la filosofía para eliminar la oscura metafísica. Ambas disciplinas procuraron conseguir esas metas por medio del principio de verificación. La intersubjetividad se operativizó como *el* criterio de objetividad. En su forma extrema, la equiparación de la objetividad con la idea cuantitativa de intersubjetividad se manifestó en el conductismo metodológico y en el operacionalismo. Pero la falacia del intersubjetivismo alcanza a todos los campos.

SCRIVEN menciona el caso de la evaluación de una antena de televisión en una revista de electrónica en la que el evaluador ve e informa de que la imagen procedente de una de las antenas probadas es mejor. No obstante, el evaluador pide disculpas por haber sido "subjetivo", *ya que no utilizó instrumento alguno para medir el incremento en decibelios.* En realidad, como señala SCRIVEN, sin utilizar instrumentos, es posible llegar a acuerdos intersubjetivos respecto al funcionamiento de equipos electrónicos, sin que la correlación de estos juicios de calidad reunidos con cualesquiera lecturas de instrumentos de medida sea elevada. ¿Por qué, entonces, para este evaluador confundido, la lectura de un instrumento es objetiva, mientras que el juicio de una persona es subjetivo?

La razón está en que el evaluador es la única persona que efectúa la observación y, aunque sabe que podría confirmarla apelando a sus colegas,

© Ediciones Morata, S. L.

cree que sería mejor disponer de un instrumento porque se llegaría a un acuerdo aún mayor entre observadores sobre la lectura del aparato de medida (aunque la medida que arroje el aparato no constituya un índice elevado de calidad). En este caso, la idea cuantitativa de intersubjetividad suplanta la calidad de la percepción.

En términos operacionales, según SCRIVEN, "la medida en una escala cuantitativa por medios mecánicos" se convierte en indicador de veracidad porque la fiabilidad entre jueces es superior. Al mismo tiempo, se sacrifica en realidad la validez en beneficio de la fiabilidad, porque la lectura del aparato de medida, aunque fiable, no es un buen indicador de la calidad de imagen. Éste es uno de los errores corrientes de la evaluación: la sustitución de la observación directa de la calidad por los instrumentos, la de la validez por la fiabilidad. Y es un error de primera magnitud.

A partir de esta idea —lo que otros no pueden experimentar directamente no puede tomarse como ciencia auténtica (intersubjetivismo)—, se ha desarrollado el concepto de objetividad como exteriorización de todas las referencias, de manera que pueda conseguirse una prueba múltiple; según SCRIVEN, una excesiva y grosera simplificación. En la investigación educativa, esto se ha manifestado en la equiparación de la objetividad con la capacidad de especificar y explicar del modo más completo todos los procedimientos de recogida de datos. La exteriorización y objetivación completas permiten la réplica, la garantía de la fiabilidad. En educación, ser objetivo ha llegado a significar disponer de un instrumento "válido" (exactamente, como un evaluador electrónico).

Lo que, en realidad, existe son instrumentos muy fiables cuya validez es cuestionable. Su correlación con los juicios de calidad no siempre es elevada. La deformación del principio intersubjetivista de verificación se ha traducido en la equiparación de la objetividad con procedimientos exteriorizados y replicables, aunque éstos puedan ser influidos por inclinaciones tendenciosas y, en consecuencia, ser subjetivos, desde el punto de vista cualitativo.

La identificación de la objetividad con un procedimiento externo completamente especificable tiene otro efecto importante. Libera al evaluador de la responsabilidad respecto de los resultados y consecuencias de la evaluación. En realidad, si estos instrumentos y procedimientos "objetivos" arrojan esos resultados, ¿cómo vamos a culpar al evaluador? Habrá que acusar a la ciencia. POLANYI (1958) denomina "objetivismo" a esta postura. En este sentido, la objetividad significa que las observaciones están sometidas a una verificación independiente sin referencia a la persona que las efectúa.

Ahora bien, no es posible especificar explícitamente todo el conocimiento ni verificarlo por completo mediante procedimientos externos independientes. SCRIVEN sostiene que, incluso en las pruebas matemáticas, cuyos pasos se reducen a lo evidente de por sí, la intuición desempeña un inevitable e importante papel. La verificación intersubjetiva no sólo no es garantía de veracidad, sino que no es necesaria. La verdad constituye un ideal al que

© Ediciones Morata, S. L.

UNIVERSIDAD DE PLAYA ANCHA
DE CIENCIAS DE LA EDUCACIÓN
SISTEMA DE BIBLIOTECAS
BIBLIOTECA CENTRAL

Scanned by CamScanner

sólo podemos *acercarnos* mediante la interacción entre introspección y verificación pública.

Dada su complejidad, nunca podrán explicarse del todo muchos juicios intuitivos. No obstante, las conclusiones no tienen por qué ser menos ciertas por nuestra incapacidad de explicarlas. El acuerdo entre muchos quizá sea necesario para explicar la verdad a terceros, pero no es preciso para la verdad en sí.

¿Cómo podemos establecer la validez de una afirmación si no podemos separarla por completo de la persona que la hace? Una forma consiste en comprobar la fiabilidad del observador en casos anteriores y constatar su carencia de sesgos. Estas comprobaciones no garantizan la consecución de la verdad, pero aquí no hay garantías posibles. Según SCRIVEN, hay afirmaciones relativas al conocimiento, híbridas de manifestaciones internas y externas, p. ej., enunciados de tendencias, analogías, aproximaciones, que son ciertas aunque no formen parte de los tipos de afirmaciones que solemos asociar con los enunciados científicos. Las llama afirmaciones de "conocimiento débil" y dice que representan el tipo de conocimiento del que pueden disponer las ciencias sociales.

Estos enunciados de conocimiento se manifiestan más como explicaciones que como previsiones. La explicación y la comprensión son funciones del modo de codificarse la información en la mente. La explicación supone que existe una persona que la entiende. No existe de por sí. En último extremo, o la comprensión es reducible a algo ya presente en la mente de sus destinatarios, que realizan el proceso de comprensión, o no habrá explicación.

De igual manera, salvo que la evaluación proporcione una explicación a unos destinatarios concretos, aumentando su comprensión por medio del contenido y la forma de presentación de las argumentaciones, no será una evaluación adecuada a esos destinatarios, aunque los hechos en los que se base puedan verificarse por otros procedimientos. Un indicador de la capacidad explicativa es el grado de persuasión de los destinatarios conseguido. En consecuencia, una evaluación puede ser "cierta" en el sentido convencional, pero no persuasiva para unos destinatarios concretos, a quienes no sirve de explicación. Por tanto, en su sentido más pleno, la evaluación depende tanto de la persona que efectúa el informe evaluativo como de la que lo recibe.

No es necesaria la previsión para demostrar la comprensión. La inferencia de un acontecimiento a partir de un coeficiente de correlación más algunas condiciones antecedentes no constituye un test necesario de validez u objetividad. En cambio, el modelo de razonamiento se acerca al de ajuste del modelo, de encontrar interpretaciones, explicaciones y comprensiones razonables *en un contexto dado.* El test de una explicación no es la exactitud de la previsión de un hecho, sino que los destinatarios puedan contemplar relaciones nuevas y responder "preguntas pertinentes nuevas".

Por último, respecto a la cuestión de la objetividad hemos de concluir dos cosas: o bien la objetividad no puede identificarse en exclusiva con un

© Ediciones Morata, S. L.

procedimiento exteriorizado, independiente por completo de las mentes que llevan a cabo las observaciones y las comprenden, o bien gran parte de la verdad es de carácter subjetivo. En el primer caso, la objetividad significa algo más que lo que suele tomarse por tal; en el segundo, significa algo menos.

¿Qué decir de la validez? Una definición de validez consiste en que se basa en procedimientos objetivos. La validez lleva consigo ideas sobre la concordancia entre lo que se intenta y lo que se lleva a cabo, la derivación correcta y la sanción por parte de la autoridad. En el sentido estricto de la objetividad cuantitativa, la validez se equipara con la previsión (comprobación de los datos en relación con un criterio). Pero ello supone una tentativa única y el intersubjetivismo como principio de verificación. Es un procedimiento demasiado estricto. En último término, dice CRONBACH (1971), la validez depende del uso que se haga de los datos y "la utilidad depende de los valores, no de las conexiones estadísticas de las puntuaciones".

Si no podemos conseguir una única puntuación que constituya un presunto indicador de validez, ¿cómo se determina ésta? Quizá la mejor respuesta consista en examinar las posibles causas de invalidez. Una evaluación puede ser no válida de diversos modos. Uno de ellos es que los "hechos y verdades" en los que se basa sean erróneos. Ambos se aceptan sin discusión. Otros datos han de determinarse mediante procedimientos aceptados de recogida de datos que, a su vez, están refrendados por una disciplina concreta y sometidos a examen público. A menudo, la validez se refiere a la utilización de tales procedimientos aceptados de recogida de datos, como señala el artículo de CRONBACH sobre validación de los tests.

Otro modo de cuestionar la validez tiene que ver con las relaciones que se establezcan entre las conclusiones e interpretaciones y los datos. Como afirma CRONBACH, la validez o no validez no se refiere tanto al test o a los procedimientos de recogida de datos como a las interpretaciones de los mismos. Se trata de la validez de una inferencia: ¿la inferencia se deriva correctamente de los datos y premisas?

También está la cuestión de si la interpretación es aplicable a situaciones distintas de aquélla de la que se ha derivado, pues las generalizaciones dependen del contexto. El diseño experimental se ocupa de forma sistemática de estas cuestiones en cuanto constituyen amenazas contra la validez interna y externa.

En los estudios cualitativos, es más difícil presentar pruebas de validez, lo que no significa que no existan. La demostración de la validez de los estudios naturalistas suele consistir en la confirmación de una clase de datos con otra. Al proponer los estudios de casos en la enseñanza de las ciencias, STAKE y EASLEY (1978) consideraban que las tendencias personales y la experiencia anterior constituían la principal amenaza a la credibilidad de los estudios de casos. Propusieron el uso generalizado de grabaciones magnetofónicas de entrevistas; de citas directas, en la medida de lo posible, e informar de los desacuerdos entre los entrevistados, si los hubiera. Las personas conocedoras de la situación local podrían leer el informe escrito y

© Ediciones Morata, S. L.

Scanned by CamScanner

juzgar la exactitud de la descripción. Habría que señalar los "indicios de incoherencia" de los trabajadores de campo con el fin de seguir su evolución posterior. También se documentarían y describirían los contextos de las observaciones. El hecho de garantizar las observaciones de diversos participantes sobre una cuestión o acontecimiento concreto era un modo de "triangular" lo que en realidad ocurriese.

La mayoría de las amenazas contra la validez se contempla desde la perspectiva de un destinatario universal. Pero hay otra forma de considerar la validez de una evaluación: si ésta es válida para destinatarios concretos. En realidad, la validez siempre está relacionada con los fines de alguien y la utilidad que le proporcione. Si la evaluación no se basa en valores propios de la mayoría de los destinatarios, éstos quizá no la consideren "válida", es decir, pertinente para ellos, en el sentido de bien fundamentada, justificable o aplicable. La evaluación puede pasar por alto, sin más, las principales cuestiones que se plantean unos destinatarios determinados. Al mismo tiempo, la evaluación puede ser válida, en el sentido de que los hechos en los que se fija son reales y las inferencias efectuadas a partir de los datos, correctamente derivadas. Desde el punto de vista de unos destinatarios concretos, el error puede estar en las premisas de partida.

Por otra parte, una evaluación también puede ser no válida en este sentido secundario si las formas de argumentación utilizadas son erróneas. Por ejemplo, en esta sociedad, las argumentaciones de "medios y fines", sobre todo las correspondientes a las relaciones entre coste y eficacia, tienen una fuerza singular. Si empleamos un argumento encaminado a favorecer la excelencia al máximo, en vez de escoger la mejor alternativa disponible, puede tener poco peso aunque, desde el punto de vista de un destinatario generalizado, ambas orientaciones sean igualmente verdaderas y válidas. Por tanto, la validez se aplica a la evaluación de formas muy diversas. (El debate entre GLASS y SCRIVEN que aparece en el Apéndice A se refiere, más que nada, a la forma del argumento.)

También ocurre que, cuanto más "naturalista" sea la evaluación, más se basará en sus destinatarios para extraer sus propias generalizaciones (validez externa). Por ejemplo, cada lector puede interpretar un estudio de casos de distinta manera, pues tiene presente su propio universo de casos a efectos de comparación. El lector puede ver semejanzas y diferencias basadas en su experiencia, pudiendo hacer sus propias interpretaciones.

Concibiendo de este modo el proceso de generalización, se modifica incluso el primer sentido en el que se utiliza la validez. El evaluador sigue siendo responsable de la confirmación y comunicación de los hechos y enunciados "verdaderos", pero parte de la interpretación se escapa de sus posibilidades. Dado que, como dice CRONBACH, la cuestión final consiste en la validez de la interpretación, que sólo el lector conoce con seguridad, los destinatarios deben asumir una considerable responsabilidad respecto de la validez de sus propias interpretaciones. En último extremo, el evaluador debe suponer que el pensamiento de sus destinatarios sigue procesos racionales.

© Ediciones Morata, S. L.

Como señalaba ENNIS (1973), la validez interna y la externa se refieren a fenómenos muy diferentes. La validez externa tiene que ver con la capacidad de generalización de enunciados causales generales. La validez interna versa sobre enunciados causales específicos que no supone la generalización a otros casos. La generalización implica siempre el conocimiento de las leyes pertinentes que intervienen en la extrapolación a ámbitos distintos. En cambio, un estudio de validez interna sólo alude a la causalidad en el pasado y en circunstancias específicas. No pretende la extrapolación, siendo, por tanto, menos dependiente de premisas externas.

Sin embargo, ni los enunciados causales específicos ni los generales se derivan lógicamente de un modo perfecto de las observaciones, ni siquiera en el caso de los mejores diseños experimentales. Incluso en los diseños más estrictos, son necesarias algunas premisas empíricas. Además, según ENNIS, la identificación de un hecho concreto con la causa supone inevitablemente un juicio de responsabilidad acerca de que un acontecimiento concreto es responsable del efecto. Esta adscripción de responsabilidad requiere un importante conocimiento de base y un juicio de valor. Lleva consigo la probable asignación de alabanza o censura y muestra un ámbito de intervención.

La mayoría de los evaluadores asume la responsabilidad de enunciados causales específicos del tipo "x causó y" en este estudio (validez interna), aunque, en sí, esto supone necesariamente un conjunto de premisas. Pero algunos remitirían la capacidad de generalización de los hallazgos a los juicios de los destinatarios, dado que ésta se basa en una información externa que pueden tener ellos, pero no el evaluador. Los destinatarios podrían efectuar algunas adscripciones de responsabilidad basándose en su propio conocimiento de base y en sus valores. Algunos evaluadores, sobre todo los naturalistas, podrían decir que, en último término, esto se traduce en generalizaciones superiores.

Hay aún otro problema relacionado con la objetividad. ¿Basta decir, en realidad, que el evaluador es objetivo? Si tomamos la objetividad en el sentido en que suele utilizarse, en cuanto procedimiento exteriorizado, especificable, que produce resultados repetibles, es, sin duda, un criterio insuficiente para la evaluación. La administración de tests normalizados de ejecución constituye un procedimiento completamente exteriorizado y especificable que produce resultados repetibles. Al mismo tiempo, parece que estos tests están muy sesgados de muchas formas, sobre todo en relación con grupos minoritarios. En este sentido, tenemos un instrumento objetivo, pero sesgado. De hecho, podemos elaborar un instrumento cuyo sesgo se incline en el sentido opuesto. (Para mayor confusión, si lo que pretende tal instrumento es la discriminación racial, podemos disponer de un instrumento objetivo, válido, para este fin.)

Una evaluación debe carecer de deformaciones y sesgos (debe ser objetiva, desde el punto de vista cualitativo) y el hecho de que esté exteriorizada, sea especificable y repetible no basta para conjurar posibles inclinaciones tendenciosas. Incluso, para la evaluación, no es suficiente la objeti-

© Ediciones Morata, S. L.

Scanned by CamScanner

vidad cualitativa porque lleva consigo el aura de neutralidad. Las personas sometidas a evaluación no quieren un evaluador neutral, que no se preocupe por los problemas. Una persona sometida a juicio no escogería un juez totalmente independiente de su propio sistema social.

La postura desinteresada no nos da derecho a participar en una decisión que determina en gran medida la suerte de alguien. El conocimiento de técnicas para lograr descubrimientos objetivos es inadecuado. En cambio, el evaluador debe ser considerado como miembro del grupo sometido a juicio o vinculado con el mismo, de igual modo que un acusado es juzgado por sus semejantes*. El evaluador ha de ser considerado como una persona preocupada, interesada y sensible a los argumentos pertinentes. Debe ser más imparcial que simplemente objetivo.

La imparcialidad del evaluador debe considerarse como la de un actor dentro de los acontecimientos, de alguien sensible a las argumentaciones adecuadas, en quien las fuerzas contendientes están equilibradas, en vez de no existir. El evaluador debe ser considerado como alguien que no toma partido de antemano a favor de una u otra postura.

El evaluador puede remitirse a criterios objetivos para resolver los problemas; pero, cuando se discute su propia imparcialidad, no basta con que dé pruebas de objetividad. Debe demostrar su imparcialidad, señalando cómo, en el pasado, ha actuado en contra de sus propios intereses.

El discurso evaluativo: la buena vida (junto a la falla de San Andrés)

Empecé este capítulo hace varias semanas. Todavía no se ha producido el gran terremoto de Los Ángeles. Se suceden días magníficos, muy parecidos unos a otros; por tanto, parece que mañana será como hoy, un sueño placentero que se extiende indefinidamente.

Cada día que pasa hace que el terremoto parezca menos probable que antes. Sin embargo, si ha de producirse en este año, debería ser *más* probable. Razono que el tiempo que he permanecido aquí es sólo una pequeña fracción del año en curso, de manera que las posibilidades de que se produzca el terremoto son ahora menores que eran respecto a todo el año correspondiente a la previsión. Razono que, aun en el caso de que el terremoto se produzca, los efectos no serán desastrosos. Además, el Medio Oeste está devastado por los tornados. Junto a esto, ¿acaso muchos de los hombres más despiertos del país, incluidos los sismólogos, vivirían aquí si el peligro fuese tan grande? Me siento mejor. Mi ansiedad disminuye.

Mientras tanto, en los últimos días, el *New York Times Magazine* reaviva el drama en su edición del bicentenario (4 de julio de 1976). Como elemento simbólico de *"America at 200"*, publica un reportaje sobre *"The Good Life (along the San Andreas Fault)"**. En portada aparece una ilustración de un fragmento de autopista zambulléndose en pleno océano, los restos de Los Ángeles tras el próximo terremoto. El artículo comienza con un panorama de seis párrafos sobre los efectos del terremoto previsto.

* El autor alude a la actuación del jurado en un proceso judicial norteamericano. *(N. del T.)*

© Ediciones Morata, S. L.

Quienes viven encima de la línea de fractura de nueve millas de profundidad tienen sus propias razones para vivir allí. Cuando su patio se desmorona a diario, un trabajador de correos, que tiene tres coches, preferiría mudarse, pero no puede vender su casa. El gerente de un rancho, para quien la vida en California es mejor que en cualquier otro sitio en el que haya vivido, explica: "No me voy. ¿Acaso existe algún lugar en el que no haya ninguna catástrofe?" Para algunos, la misma precariedad hace mucho más valioso estar aquí. Un consejero de inversiones en paro que vive en la falla dice: "Vives en un punto crítico. Puedes perder todo lo que tienes en cualquier momento".

No son las razones que yo daría, pero pueden ser correctas. Toda persona es libre para sopesar sus propias razones. Todos somos libres para hacer nuestras propias opciones. Y así debe ser cuando nos enfrentamos con esa incertidumbre. No podemos basar el juicio en una realidad irrefutable. Aun cuando los terremotos sean mucho más previsibles, sigue habiendo margen para escoger la respuesta. En el campo de la decisión social, la certeza parece remota, cuando no imposible.

Ante tales dificultades para llegar a una realidad irrefutable, hay quienes tratan de hacer sencillas las complejidades de la vida. Insisten en pretender que hay acuerdo cuando no lo hay, tanto respecto a hechos como a valores. Ocupando con frecuencia posiciones de poder, imponen definiciones simplificadas de la realidad en beneficio de la acción. Sin embargo, con independencia de la mayor o menor aceptación de la simplificación, la realidad sigue ahí. Crean lo que crean veintiún millones de californianos, el gran terremoto llegará.

La alternativa no tiene por qué ser una vuelta a la irracionalidad. Si las opiniones no pueden tener un fundamento irrefutable, tampoco deben considerarse completamente arbitrarias, como meros "juicios de valor". Esa clasificación circunscribe el conocimiento a lo claro, distinto y unívoco. Esta distinción crea un cisma entre el saber teórico objetivamente verdadero, por una parte, y la acción basada en motivos irracionales, por otra. En último extremo, lleva a tildar de irracionales a quienes no concuerdan con la propia perspectiva. La clasificación de ciertas personas como irracionales justifica el desprecio de sus opiniones y, quizá, de su dignidad e intereses; incluso, justifica la utilización de la sugestión y la fuerza contra ellas.

La alternativa consiste en tratar como racionales a todos los hombres. Entre el autoritarismo conservador de la tradición y el autoritarismo liberal del cientifismo, entre la certeza del fanatismo y la evasión de la responsabilidad del escepticismo se sitúa la deliberación racional. Debemos tomar en

* El autor reproduce en el encabezamiento de la sección el título de *The New York Times Magazine:* "La buena vida (junto a la falla de San Andrés)". *(N. del T.)*

© Ediciones Morata, S. L.

Scanned by CamScanner

serio las opiniones de los demás, introduciéndolas en un discurso riguroso. Éste es el ámbito de la argumentación y la esfera adecuada de la evaluación.

Partimos de la base de que los grupos de personas se adhieren a opiniones diversas con intensidad variable y que estas opiniones pueden ponerse a prueba en un discurso serio. Pueden considerarse incluso hechos concretos y valores. El discurso racional consiste en dar razones, aunque no razones indiscutibles. En el ámbito de la acción, en el que hay pocas cosas claras y distintas, la motivación puede ser racional. La práctica puede ser razonable.

El evaluador debe hacer participar a sus destinatarios en un diálogo en el que puedan emplear su razonamiento con toda libertad. Esto significa que los destinatarios deben asumir la responsabilidad personal de su interpretación de la evaluación, dado que el razonamiento que se les presenta no es del todo convincente ni completamente arbitrario. Esto supone que el evaluador asuma también la responsabilidad personal de sus juicios, ya que no puede ocultarse tras un método ciego. Ambas partes deben poner en práctica su razón natural.

© Ediciones Morata, S. L.

CAPÍTULO V

Coherencia y credibilidad
La estética

El conductor ebrio

> La humanidad permanece irremediablemente en la caverna de Platón, gozando todavía —su antiquísima costumbre— con meras imágenes de la verdad.
>
> Susan SONTAG: *On Photography*, 1977, pág. 1.

Consideremos dos imágenes distintas del conductor bebido. Imaginémonos al bebedor social corriente, al que suele tratársele con excesiva tolerancia, y a quien, tras desobedecer una señal de parada obligatoria, detiene la policía y, por tanto, se enfrenta con un problema. Imaginemos ahora al conductor que es bebedor habitual, siempre con unas copas de más, que va haciendo eses, dando tropezones y que, en la carretera, constituye un grave peligro para todos, incluso para sí mismo. La imagen que nos hacemos del conductor que bebe tiene mucho que ver con las recomendaciones que, orientadas a la acción, podríamos suscribir como medio para eliminar a los conductores bebidos.

Hasta la pasada década se ha venido considerando al conductor que bebe como un bebedor social, problema civil susceptible de corrección mediante acciones legales, como la imposición de castigos y multas. Sin embargo, desde hace unos diez años, un conjunto de estudios ha influido en la transformación de la imagen predominante del conductor bebido,

---

NOTA DEL AUTOR: Este capítulo fue publicado como: *"Coherence and Credibility: The Aesthetics of Evaluation"*, en *Educational Evaluation and Policy Analysis*, Vol. 1, n.º 5, septiembre-octubre 1979, págs. 5-17. Copyright 1979: *American Educational Research Association*, Washington, DC. Reimpreso con autorización. Gene GLASS, Jo FRIEDMAN, Donald HOGBEN, Rochelle MAYER, John NISBET y Helen SIMONS me hicieron valiosos comentarios sobre el mismo.

© Ediciones Morata, S. L.

Scanned by CamScanner

concibiéndosele más en la actualidad como bebedor habitual, patológico. Las recomendaciones para solucionar el problema derivadas de esa imagen son de tipo sanitario y están más orientadas hacia un reducido subconjunto de transgresores que a todos los conductores que beben.

El sociólogo GUSFIELD (1976) analizó la retórica empleada en estos estudios (la forma de presentación de los datos para persuadir al lector de las conclusiones). Lo que descubrió es realmente provocador. En las ciencias sociales, se ha considerado axiomático que, aunque un novelista pueda persuadir a un lector con la emoción, el científico sólo persuade con la lógica. GUSFIELD alude a esta idea de la neutralidad del lenguaje como la teoría del lenguaje de "cristal de ventana": una expresión que revela la realidad con absoluta transparencia y sin ninguna clase de adornos. El autor demuestra que, en la práctica concreta, la valoración de los estudios científicos se debe, en parte, a la utilización dramática que en ellos se hace del lenguaje.

GUSFIELD analizó con detalle el estudio más influyente sobre los conductores bebedores: *"Identification of Problem-Drinking Among Drunken Drivers"* (WALLER, 1967). Utilizando las categorías de BURKE (1945) de escenario, acción, agente, mediación y objetivo, GUSFIELD puso de manifiesto que el estudio de WALLER operaba como un instrumento retórico persuasivo. Como el estilo literario de las ciencias es de carácter neutral, el estudio científico debe parecer no literario, sin que medie en él punto de vista personal. En el estudio de WALLER sobre los conductores bebidos, la neutralidad del escenario viene dada por el rol del autor —investigador médico—, que escribe en el *Journal of the American Medical Association.* Establecida la neutralidad del escenario, desaparece la figura del autor en cuanto persona que interviene en el asunto.

La forma que reviste el estudio es narrativa, con una evidente estructura dramática. La tensión aumenta con la opción entre dos tipos de conductores bebidos, descargándose en el desenlace. La progresión dramática va desde el bebedor problemático como bebedor social hasta el problemático como conductor bebido. Según GUSFIELD, en esta progresión es importante que el cambio de perspectiva (resultado real del estudio) se considere como la consecuencia de datos externos derivados mediante el método científico. Por este motivo, el tono del informe es impersonal: "Recientes informes han indicado..."; "se va poniendo cada vez más de manifiesto que...". La forma impersonal sitúa la acción en la mediación externa de los datos y el método más que en el autor. Refuerza la idea de que las conclusiones surgen del universo de datos impersonales.

El estilo "científico" resultante es clínico, desinteresado, impersonal y carente de imágenes. El autor presenta el universo externo y permite que persuada al lector. El estilo sugiere que el observador se rige por el método y las reglas de la integridad científica. La atención prestada a los detalles y a la descripción meticulosa de los procedimientos, como la presentación de los números con siete cifras decimales, da impresión de precisión. Según GUSFIELD, este estilo refuerza la premisa epistemológica básica: utilizando

© Ediciones Morata, S. L.

el mismo método, distintos observadores tienen que llegar a las mismas conclusiones.

Aunque la exposición de la metodología se desarrolla con una neutralidad evidente, las consecuencias que se extraen para la acción se manejan de muy distinta manera. Los hechos de experiencia se convierten en imágenes. En su estudio sobre los conductores bebidos, WALLER menciona de manera reiterada la expresión "conductores bebidos", suscitando de ese modo imágenes poderosas, en vez de hacer una descripción con términos más neutros, como: "conductores que sufren accidentes después de beber". En el mismo título del artículo se utiliza la expresión "conductores bebidos". En la actualidad, la categoría de embriaguez se define operacionalmente por la cantidad de alcohol en sangre, definición legal cuya carga emocional es muy diferente de la que suscita la imagen del conductor ebrio.

Según GUSFIELD, cuando WALLER identificó a los "conductores bebidos" con los bebedores problemáticos mediante su revisión de archivos judiciales civiles, extrajo las consecuencias para la acción del conjunto de imágenes de su clasificación original. De este modo, el científico médico dio sentido a sus datos por reducción a imágenes. La representación del conductor bebido amplió el significado de los datos primarios de manera que abarcasen interpretaciones ya aceptadas y conocidas por sus destinatarios. No se trataba de una simple extrapolación de los datos.

En síntesis, la imagen del conductor bebido atribuye la responsabilidad al agente, al mismo conductor, y no al escenario en el que éste actúa. La conducción en estado de embriaguez se convierte en atributo del individuo que requiere tratamiento médico. Supone una conducta más patológica que normal y, en consecuencia, soluciones diferentes. Incluso se trata de un subgrupo diferente de personas. El delincuente civil se transforma en paciente.

GUSFIELD señala que este estudio es característico, en cuanto al estilo retórico, de los trabajos sobre la conducción en estado de embriaguez, aunque haya tenido una influencia especial. En un plano más general, afirma que la ciencia opera mediante esta clase de reducciones. El saber generalizable se crea mediante la vinculación de objetos específicos con categorías universales. "En esta empresa (la ciencia) están implícitas y son inherentes a ella la definición, descripción e interpretación de los datos a través de comunicaciones orales o escritas en la medida en que las conclusiones y generalizaciones suponen significados para la acción" (pág. 31). Y más adelante: "para que tengan relevancia o sean significativos, no sólo hay que seleccionar los datos, sino tipificarlos e interpretarlos también" (*ibid.*).

Aunque no todos los estudios de evaluación utilizan unas imágenes tan vívidas, en la evaluación surgen problemas similares. Las representaciones, la estructura dramática y la forma de exponer una evaluación constituyen cuestiones fundamentales para la importancia de la misma. Estos elementos, a menudo considerados como pura cosmética, pueden influir en la forma de pensar y actuar de las personas. ¿Cómo operan estos elementos y cuál es su relación con el contenido —el "valor de verdad"— de una evaluación?

© Ediciones Morata, S. L.

Scanned by CamScanner

Los evaluadores utilizan esos elementos, aunque no siempre de manera consciente y en la misma medida. Como señala GUSFIELD, "lo que se cuestiona, no obstante, es la necesidad de la interpretación y la íntima conexión entre esa interpretación y su forma de presentación, su elemento artístico" (pág. 32).

Del mismo modo, en su análisis del papel de las imágenes para definir los problemas sociales en la política social, SCHÖN (1979) afirmaba que el planteamiento de problemas sociales está mediado por lo que cuentan las personas sobre ciertas situaciones problemáticas. El encuadre de los problemas sociales depende de las metáforas subyacentes a los relatos. La forma de encuadrar los problemas es crítica respecto a las soluciones que se susciten. Por ejemplo, en todas partes se oye que los servicios sociales son "fragmentarios" y la solución implícita es que hace falta "coordinarlos". Pero los servicios que aparecen como "fragmentarios" pueden considerarse también "autónomos". La metáfora subyacente configura y orienta la solución de los problemas.

SCHÖN sostiene que nuestro pensamiento sobre la política social está orientado por imágenes omnipresentes y tácitas que denomina "metáforas generativas". Éstas se transfieren de un marco de referencia a otra situación. Por regla general, se inducen cuando nos sumergimos en la experiencia del fenómeno. Estas imágenes orientadoras son necesarias para nuestro pensamiento.

Por ejemplo, hay dos perspectivas muy distintas sobre la renovación urbana. Una considera que las barriadas baratas fueron en otras épocas comunidades saludables que se han estropeado. Un planificador social que tenga esa imagen pensará que la eliminación de la ruina urbana llevará consigo el rediseño y la reconstrucción completos. Otra visión muy distinta de las barriadas baratas las representa como comunidades naturales, viables y de renta baja que ofrecen a sus residentes importantes beneficios sociales. Esta segunda perspectiva supone recomendaciones totalmente distintas para la mejora de esas comunidades.

En la planificación social de los años cincuenta predominaba ante todo la imagen de la barriada barata como ruina. En los años sesenta, surge la de la barriada barata en cuanto comunidad natural como metáfora opuesta a la primera, compitiendo con ella para ganar la atención del público y de los expertos. Según SCHÖN, a partir de una realidad que es "ambigua e indeterminada", cada imagen selecciona características que constituyen los temas en los que se basan dichas imágenes. En la primera perspectiva, adquieren importancia expresiones como "ruina", "salud", "renovación", "ciclo de decadencia" y "plan integral". En la segunda, las ideas clave serán: "hogar", "modelos de interacción", "redes informales" y "desarticulación".

Cada imagen presenta una visión de la realidad social, seleccionando, nombrando y relacionando elementos pertenecientes al marco escogido. Según SCHÖN, los procesos clave son la "denominación" y el "enmarcamiento". Mediante la selección de ciertos elementos y la coherente organización de los mismos, esos procesos explican lo inconveniente de una

© Ediciones Morata, S. L.

situación determinada y sugieren una transformación. Los datos se convierten en recomendaciones.

La denominación y el enmarcamiento se llevan a cabo mediante la metáfora generativa. Los investigadores *ven* la barriada como ruina o como comunidad natural. Al ver *A* como *B*, la evaluación implícita en *B* se traslada a *A*. La primera metáfora es la de la enfermedad y la curación. La segunda es la de la comunidad natural (frente a la comunidad artificial). Las evaluaciones transferidas se basan en imágenes profundamente asentadas en nuestra cultura. Cuando vemos una situación compleja como de salud/enfermedad o de naturaleza/artificio, sabemos en qué dirección movernos.

La consideración de *A* como *B* facilita mucho nuestra capacidad de diagnosticar y prescribir. Por otra parte, puede llevarnos a pasar por alto otras características importantes de la situación que la metáfora no capta. Como las metáforas generativas suelen ser tácitas, podemos soslayar características importantes. SCHÖN dice que tenemos que ser más conscientes de nuestras metáforas generativas y que la mejor forma de conseguirlo consiste en analizar lo que decimos del contexto del problema.

La metáfora "profunda" explica por qué incluimos en lo que decimos unos elementos y otros no, por qué se dan como ciertas algunas premisas a pesar de las pruebas en su contra, y por qué parecen obvias algunas recomendaciones. La imagen del conductor bebido o la metáfora de la barriada como enferma configura el estudio y orienta nuestras propias acciones.

La trama

> O sea, lo primero que *hacemos* con las imágenes es construir una trama, del mismo modo que lo primero que hacemos con las palabras es decir algo; hacer un enunciado... Las imágenes y las tramas constituyen los activos de la mente.
>
> Susanne K. LANGER, *Philosophy in a New Key*, 1942, pág. 128.

Pensemos en una evaluación. En primer lugar, hay un conjunto de hechos, la realidad tal como es, que siempre será algo indeterminado. El evaluador se enfrenta con un universo ambiguo. Gracias a diversas lentes sociales y psicológicas, el evaluador representa esos hechos en un informe. El informe en sí es un producto artificial. Como tal, éste tiene coherencia y forma, cierta clase de estructura estética, aun en el caso de que la estructuración no sea del todo consciente.

Los lectores interpretan el informe y esa interpretación variará dependiendo de cada lector, de sus circunstancias y orígenes. Si a los lectores les parece creíble el informe, es más probable que asuman la correspondiente postura de valor frente al objeto evaluado. Es más, los lectores pueden ser impulsados a la acción, dependiendo de su disposición y otras circunstancias. El evaluador puede o no recomendar la acción.

© Ediciones Morata, S. L.

Scanned by CamScanner

Como artefacto artificial, el informe de evaluación tendrá cualidades estéticas, formas externas que susciten el interés, pero en un informe de evaluación hay aspectos más fundamentales que éstos. Toda evaluación debe tener un grado mínimo de coherencia. La coherencia mínima consiste en que la evaluación narre una trama. Debe existir una sucesión, explícita o tácita, de hechos (o, con mayor precisión, una interpretación de los hechos) para que el lector utilice la evaluación como guía para valorar. También pueden aparecer recomendaciones del evaluador, pero no son necesarias. La trama está presente.

En este sentido, las imágenes no pueden ser evaluaciones de por sí. Como indica SONTAG (1977) en su análisis de la fotografía:

> El deseo no tiene historia; al menos, en cada caso se experimenta todo él en primer plano, inmediato. Lo suscitan arquetipos y, en este sentido, es abstracto. Pero los sentimientos morales están inmersos en la historia, cuyas personas son concretas, cuyas situaciones siempre son específicas. Así, las reglas para utilizar la fotografía con el fin de despertar el deseo y la conciencia son casi las opuestas. Las imágenes que movilizan la conciencia siempre están vinculadas a una situación histórica concreta. Cuanto más generales sean, más difícil resultará que sean eficaces.
>
> (Págs. 16-17.)

Las fotografías pueden utilizarse para reforzar valoraciones e, incluso, presentarse en secuencia para contar una trama (TEMPLIN, 1978). La trama en sí es, no obstante, necesaria para interpretar los hechos. Éstos deben presentarse en un contexto histórico específico. La trama, más aún que la imagen y la metáfora, constituye la estructura básica subyacente a una evaluación.

Existen al menos dos formas convencionales de contar la trama. Una consiste en presentar al evaluador como observador neutral, científico. En este caso, lo que cuenta es el desarrollo de la trama. Sonaría más o menos así: "Soy un observador neutral, independiente, que ha efectuado determinadas medidas con arreglo a los cánones de la ciencia y ha descubierto que ciertas cosas son así. El programa se ha desarrollado tal como lo he descrito, con los siguientes resultados...". Al manifestarse de este modo, el evaluador especifica meticulosamente sus procedimientos metodológicos para realzar su credibilidad. Se espera que los lectores crean los resultados por la objetividad de la metodología y por su experiencia previa y su convicción respecto de tal metodología. Por regla general, el relato concluye diciendo que "se implementó el programa y los resultados fueron éstos y aquellos". Con frecuencia, escasean las descripciones de los hechos concretos. El estudio sobre los conductores bebidos es un ejemplo de esta línea de desarrollo argumental. La presentación habitual consiste en describir el proyecto o los objetivos del mismo, el tratamiento, los resultados o efectos y las conclusiones.

La segunda forma importante de llevar a cabo la trama consiste en que

© Ediciones Morata, S. L.

el evaluador permanezca próximo al desarrollo del programa, como refleja su "voz" de narrador, y cuente la trama describiendo los hechos con detalle. Para ello, el evaluador puede utilizar una expresión emocionalmente cargada y una presentación narrativa. El relato se parecerá a un informe periodístico. El primer enfoque de desarrollo de la trama se presta a la metodología cuantitativa y el segundo a la cualitativa, pero en ambos casos hay una ordenación de los hechos que configuran un relato, aunque muchos de ellos se den por supuestos y no se hagan explícitos. Cuanto más formal sea la presentación, más cosas se darán por supuestas.

Hemos de distinguir la "trama", que es una ordenación interpretativa de los hechos, básica en todas las evaluaciones y requisito previo para la valoración que haga el lector, de la forma dramática. Dicha forma puede variar, presentándose la misma trama de distintas maneras. Una trama evaluativa puede resultar más irrebatible, interesante y agradable por la forma dramática y otros elementos estéticos. Sin embargo, aunque la trama sea más básica, la estética con la que se presenta no es mera cosmética. El buen arte y la buena ciencia unidas conducen al lector a experimentar un hecho de modo parecido a como lo sintió el creador (BRONOWSKY, 1956). Como mostraba GUSFIELD en su análisis de los estudios sobre la conducción en estado de embriaguez, la estética que los envuelva puede producir un efecto importante sobre las recomendaciones y significado de un estudio político, incluso uno científico.

Tampoco puede sorprender a nadie que la "trama" de una evaluación sea más básica que su forma estética, dado que el desarrollo de la trama vincula específicamente unos hechos con otros, como en el caso de las relaciones entre causa y efecto. Los hechos de la trama pueden presentarse de distintas formas dramáticas, pero tanto ella como los elementos estéticos contribuyen a la coherencia general. En igualdad de condiciones, una mayor coherencia conduce a mayor credibilidad para los destinatarios. Las cosas se ajustan mejor. Sin embargo, si la coherencia del informe no es realista, la credibilidad se pierde.

Podemos pensar en la diferencia entre la trama y la presentación dramática como en la existente entre "contenido" y "forma" o, con términos de POLANYI (POLANYI y PROSCH, 1975), la diferencia entre la "trama" y la "estructura". Ésta puede consistir en imágenes, medidas u otras formas "artificiales" de presentación. Esto no significa que la "estructura" sea pura cosmética. El significado de una obra es la integración de su apariencia y su contenido. El significado de un poema no consiste sólo en el contenido del mismo, escrito en prosa.

La trama difiere de una cronología de hechos porque implica inferencias e interpretaciones de los mismos. Los acontecimientos se integran entre sí y las partes se ordenan en el conjunto. La trama misma puede estar integrada con mayor o menor rigidez, proporcionando la necesaria coherencia para la evaluación en su conjunto.

A estos efectos, un concepto clave es el de "coherencia". Ésta consiste en la conexión lógica, congruencia o "consistencia al razonar o relacionar,

© Ediciones Morata, S. L.

Scanned by CamScanner

de manera que una parte del discurso no destruya ni contradiga el resto; conexión armoniosa entre las distintas partes de manera que el conjunto 'sea lógico'", como en la coherencia de un razonamiento o informe *(Oxford English Dictionary)*.

De acuerdo con POLANYI (POLANYI y PROSCH, 1975), quisiera ampliar el ámbito de aplicación de la "coherencia" para que incluya también las imágenes, las estructuras dramáticas y otros elementos estéticos. Por ejemplo, podemos considerar la consistencia lógica del *concepto* de "conductor bebido", tal como se utiliza en la argumentación, pero también, podemos contemplar la *imagen* de dicho sujeto cuando se emplea de manera armoniosa con otros elementos.

En sí, la imagen sirve para condensar significados entremezclados, tanto emocionales como cognitivos, con su propia coherencia compacta. Tiene una relación con lo cotidiano, con el mundo "real" de la que carece el concepto. Podemos definir el concepto de manera explícita, pero la imagen sólo a través de la participación tácita. Los significados multiestratificados implícitos en las imágenes y demás elementos estéticos hacen posible la comunicación de evaluaciones complejas, elaboradas y muy perfeccionadas de los objetos sometidos a estudio. Los elementos estéticos transmiten intensidad, unidad y complejidad (BEARDSLEY, 1958).

Los elementos estéticos, como las imágenes y las estructuras dramáticas, se distinguen de las entidades lógicas porque aquéllos se captan de forma inmediata, sin recurrir a razonamientos formales. En este sentido, son como las percepciones. Los elementos estéticos pertenecen sólo a la apariencia y se captan mediante la influencia tácita, por la integración inconsciente de sus partes (POLANYI y PROSCH, 1975). Así, si leemos el estudio sobre el conductor bebido, asimilamos de algún modo inconsciente los significados tácitos transmitidos por las imágenes y la estructura dramática. En el aprendizaje hay un bagaje considerable de "percepción subliminal", aunque, si se indican, sea fácil detectar los elementos estéticos.

Es más, el lector participa activamente en ese aprendizaje y llega a adquirir una postura activa en la integración de las partes en significados tácitos y, con frecuencia, se involucra profundamente en el proceso, aunque la profundidad de su participación varíe entre unos y otros lectores. Más adelante, las inferencias tácitas y las explícitas se integrarán en una imagen global individual, la "estructura subjetiva de conocimiento". Se trata de la imagen del mundo del sujeto que, en último término, influye en la propia conducta (BOULDING, 1956).

Y sostengo, de un modo explícito y no tácito, que cuanto mayor sea la coherencia, más creíble y digno de confianza será el informe de evaluación. Un informe incoherente no es en absoluto creíble y, en igualdad de condiciones, cuanto más coherente sea (hasta cierto punto), más creíble parecerá.

Por regla general, también el evaluador procura ser persuasivo (para llevar a sus destinatarios a determinado punto de vista o, incluso, línea de acción concreta, mediante la razón y la comprensión). La mayor parte del peso de la persuasión recae en la argumentación explícita, pero también en

© Ediciones Morata, S. L.

este caso, la coherencia refuerza la persuasividad del argumento. Con menor frecuencia, el evaluador aspirará también a suscitar el compromiso, para vincular al lector con una postura determinada u obligarle a aceptarla. Estas evaluaciones son las más poderosas, pero tienen que pagar el precio de un conjunto reducido de destinatarios a cambio del compromiso más profundo de unos pocos. Desde la credibilidad hasta la persuasividad y el compromiso, avanzamos cada vez más hacia la acción. En la mayoría de las evaluaciones, la coherencia desempeña el papel más humilde de realzar la credibilidad.

La forma

> La ciencia no es sino el intento de descubrir la unidad en la enorme diversidad de la naturaleza o, con mayor exactitud, en la variedad de nuestra experiencia. La poesía, la pintura, las artes constituyen la misma búsqueda... de la unidad en la variedad. Cada una a su modo busca la semejanza bajo la diversidad de la experiencia humana.
>
> J. BRONOWSKI: *Science and Human Values*, 1952, pág. 16.

Si el contenido difiere de la estructura dramática de su presentación, ¿qué comprende la forma estética? En un artículo clásico sobre *"The Problem of Esthetic Form"**, PARKER (1960), siguiendo a ARISTÓTELES, decía que la forma estética puede analizarse en relación con seis principios: unidad orgánica, tema, variación temática, equilibrio, jerarquía y evolución. El principio fundamental es el de unidad orgánica y los demás son subsidiarios de éste de distintas maneras. Dicha unidad realza la coherencia plena.

El principio de unidad exige que la obra sólo contenga elementos necesarios. En cierto sentido, la forma estética está en contradicción con la vida, que contiene muchos elementos innecesarios, desde el punto de vista estético. Los otros cinco principios son manifestaciones del pricipio de unidad. En un informe de evaluación debemos disponer de un "tema" y, para conseguir el efecto completo, éste no sólo debe persistir, sino reaparacer en distintas variaciones. En los estudios de evaluación, se logra con frecuencia este efecto presentando datos que estructuran un tema común. Esta estructuración y abreviación de datos refuerzan la veracidad; movilizan al lector.

El "equilibrio" supone una igualdad de elementos opuestos que se balancean en la forma. Proporciona la tensión dramática. La "jerarquía" constituye un modo de organizar las distintas partes. Por último, PARKER añade el principio de "evolución", indicando cómo puede crearse el sentido pleno mediante una acumulación en la que las partes primitivas determinan lo que suceda después. La unidad evolutiva puede lograrse mediante un clímax

* "El problema de la forma estética". *(N. del T.)*

© Ediciones Morata, S. L.

Scanned by CamScanner

al final o asumiendo el mismo final mayor importancia que las demás partes. En la evaluación, el informe puede evolucionar hacia conclusiones surgidas de la presentación de los datos.

Quizá sea más fácil ver estos principios estéticos en una narración dramática. *The Final Days,* de WOODWARD y BERNSTEIN (1976), no es una evaluación, pero constituye un tratamiento interesante de materiales presuntamente reales. Se narran paso a paso los días finales del régimen de Nixon. Esta diaria cuenta atrás hasta el inevitable desenlace de su dimisión contribuye a la intensidad dramática de la obra. WOODWARD y BERNSTEIN no se limitan a contar uno detrás de otro los relatos de distintas personas sobre los hechos, tal como fueron recogiendo la información, sino que combinan diversas percepciones para constituir una representación general compleja, pasando de los pensamientos y acciones de una persona a los de otra. A partir de esta interrelación, surgen ciertos temas.

Distintas fuerzas convergen sobre un Nixon desafiante y recalcitrante. De diversas maneras, uno a uno, sus defensores más destacados van convenciéndose de que es culpable y debe dimitir. Estas pequeñas pinceladas personales en el marco de un drama más amplio se repiten en variaciones temáticas que contribuyen al tema global del continuo desgaste de la defensa de Nixon. Él mismo duda entre dimitir y continuar en el poder. Este equilibrio va perdiéndose día a día, hora a hora. A partir del tema del creciente aislamiento de Nixon, surge otro trascendente: su desintegración personal. Estos temas se entretejen con maestría hasta desembocar en el desenlace de su desgracia, su dimisión y su colapso como persona.

Principios estéticos semejantes se aplican al material cuantitativo, aunque a veces resulta más difícil discernir sus efectos. La simplicidad, economía y elegancia de un adecuado diseño experimental o análisis multivariado son manifestaciones del principio de unidad. De hecho, algunos científicos-filósofos, como POLANYI, creen que, en los dominios más elevados del pensamiento, el matemático se guía ante todo por su intuición estética.

Consideremos, por ejemplo, el atractivo estético de un análisis multivariado en comparación con una larga cadena de análisis bivariados de datos. Esa compresión suscita un atractivo que supera con mucho el mayor valor de verdad que pueda encerrar un análisis multivariado. La búsqueda de diseños más elegantes se basa, en parte al menos, en tales satisfacciones. O consideremos el atractivo que encierra una escala de Guttman. La parquedad de ese tipo de escalas y de las teorías jerárquicas basadas en ese ordenamiento apela a la sensibilidad estética. Siempre buscamos problemas y datos que se adapten a esos diseños.

El atractivo estético de los estudios cuantitativos se refleja en expresiones como "coherencia", "elegancia", "economía" y "potencia". En cierto sentido, las soluciones "elegantes" son compactas, sencillas, rápidas. Proporcionan una sensación de control, utilidad y dominio semejantes a la búsqueda y hallazgo de la "palabra exacta" para describir algo. Las soluciones elegantes evitan pasos y procedimientos innecesarios. Son irresistibles.

No sólo encontramos esta coherencia en los estudios cuantitativos. Ni

© Ediciones Morata, S. L.

los elementos estructurados tienen por qué ser cifras o símbolos. Una de las razones primordiales del carácter persuasivo de la evaluación de las grabaciones educativas de AERA efectuadas por GLASS (1972) consiste en la forma en que los argumentos se conjugan en una representación general (Apéndice A). Los elementos son las argumentaciones. GLASS consigue integrar las partes lógicas de la evaluación en una estructura retórica global coherente. En este caso, la elegancia significa que pocas partes son prescindibles respecto a la lógica global de la evaluación. Hay una estructura orientadora clara, aunque compleja, en la que las partes se ajustan con exactitud y pulcritud en el conjunto.

Además de la elegancia y la coherencia, hemos de afirmar que el trabajo de GLASS resulta creíble por su "voz", la actitud del narrador en relación con su objeto de estudio. GLASS proyecta una "persona" en calidad de narrador convencido por los datos, con independencia de cómo se manejen. La imagen del narrador es la de un hombre "racional" dispuesto a dejarse persuadir por las pruebas. La identificación de esta característica como un efecto del estilo no cuestiona su autenticidad.

*"Outward Bound" (Salida)**

> Evocar en uno mismo un sentimiento ya experimentado y evocado después en uno mismo por medio de movimientos, líneas, colores, sonidos o formas expresadas en palabras, para transmitir ese sentimiento de manera que otros experimenten el mismo sentimiento: ésta es la actividad artística.
>
> TOLSTOY: "La comunicación de la emoción".

La evaluación que SMITH, GABRIEL, SCHOTT y PADIA (1976) hicieron de *Outward Bound* resulta poco habitual por utilizar tanto métodos cuantitativos como cualitativos. Las dos partes se trataron por separado y ambas se incluyeron en el informe final. Las líneas que abren el estudio sitúan el problema y la tensión dramática:

> ¿Qué le ocurre al joven que opta por afrontar las dificultades físicas relacionadas con las rocas, peñascos, campos nevados y torrentes de las Montañas Rocosas y las dificultades interpersonales de un pequeño grupo aislado en un medio extraño? Los defensores del programa *Outward Bound,* alumnos y profesores, sostienen que los efectos psicológicos son exclusivos en cada participante sin que puedan preverse. En cambio, otros muchos afirman que el programa

* Como observará el lector, el programa *Outward Bound* consistía en cursillos de verano cuyo objetivo es el fortalecimiento de la personalidad de quienes participan en él por medio de un ejercicio físico inhabitual, la permanencia en un ambiente imponente y hostil, en relación con el medio ordinario, y la vida en un grupo reducido de personas aisladas en el medio físico descrito. La denominación del programa, cuya traducción al castellano es "salida" es lo bastante alusiva al respecto. *(N. del T.)*

© Ediciones Morata, S. L.

Scanned by CamScanner

causa cambios profundos y previsibles en los sentimientos de los participantes respecto a sí mismos y a los demás.

(Pág. 401.)

La tensión introducida en estas líneas proporciona ambiente dramático, significado e interés. La cuestión básica se trata de dos maneras: mediante la medición cuantitativa y por observación cualitativa participante. Ambas reflejan la historia global de la evaluación.

En la parte cuantitativa, se escogieron cuatro variables como criterios de resultados: "autoestima", "autoconciencia", "autoafirmación" y "aceptación por los demás". Se elaboraron instrumentos psicométricos para medir las cuatro variables. La "voz" del apartado cuantitativo era la del narrador neutral, científico. La oportunidad para la elegancia provenía del diseño del estudio.

Trabajando con grupos de participantes que desarrollaban su entrenamiento en diferentes ocasiones durante el verano, los evaluadores asignaban al azar estudiantes a grupos, administrando los cuestionarios sólo una vez a cada grupo en distintos momentos. Así, pudieron utilizar un diseño complejo de series temporales en el que cada grupo servía de control de los demás. Se recogieron y analizaron 44 medias en cada uno de los cursos de junio, julio y agosto.

La unidad se conseguía mediante la representación gráfica precisa de las 44 medias, de manera que pudiesen compararse entre sí, tanto de manera visual como matemática. Una tabla de cuatro por tres entradas resumía de forma sucinta la considerable masa de datos. El único defecto estético consistía en que las cuatro variables de resultados no estaban relacionadas de ninguna manera entre sí, rompiendo la unidad del análisis. La observación participante se ocupó también de la misma cuestión básica. La tensión dramática quedó establecida a través de un relato de primera mano. El estilo era personal y participativo, en vez de neutral y desinteresado.

> El viaje de seis horas de Denver a las montañas de San Juan permitió que se incubaran nuestros temores y esperanzas. Mis propios temores se centraban en la diferencia entre mí mismo y mis compañeros de *Outward Bound*. Calculé que doblaba la edad de los demás y me preguntaba si su fortaleza y aguante superarían los míos. La perspectiva de fallar físicamente y constituir un lastre social era desalentadora. Como los niños de mi alrededor, sabía algo de lo que se suponía era y me procuraba *Outward Bound* para mí. Era consciente de sus orígenes militares y de los rumoreados peligros físicos del programa. Era también la indicación de que experimentaríamos la entrada mística en la edad adulta. Aunque ya hacía mucho tiempo que yo había entrado en dicha edad, ciertos aspectos de mí mismo suscitaban dudas respecto a los próximos veintitrés días: mi temor a las alturas, mi inexperiencia en situaciones competitivas, mi intransigencia cuando me obligaban. Las dudas eran cuestiones mentales repetidas: ¿me dejarán atrás? ¿Me entrará pánico a mitad del ascenso por una roca escarpada? ¿Fracasaré en el papel de adulto, me haré cargo de demasiada

© Ediciones Morata, S. L.

> responsabilidad y seré rechazado? ¿Abandonaré? El silencio que reinaba en el autobús me llevó a pensar que los demás también iban meditando sobre las experiencias que se avecinaban. Cada uno de nosotros tenía sus propias visiones y pesadillas íntimas.
>
> (Págs. 411-412.)

En el estudio, el drama personal y la expresión de sentimientos acrecentaban el interés. Los encabezamientos de las distintas secciones indicaban la progresión dramática: "Esperanzas", "Competiciones", "Expediciones", "A solas", "Desafío social", "Expedición final", "Maratón". El clímax de la narración estaba constituido por el ascenso a una montaña peligrosa y una carrera agotadora. La redacción era a la vez penetrante y vívida.

> Pensé que *Outward Bound* estaba especialmente diseñado para personas como Chris. Nunca había trabajado, nunca había tenido que luchar, en realidad, sólo había vivido esta experiencia de forma indirecta a través de sus padres y de los medios de comunicación. El curso la obligaba a realizar una serie de experiencias compendiadas, provocando la aceleración de su maduración. Casi podía ver el aumento de su vigor y de su confianza a medida que progresaba el curso. Ella luchaba con fuerza, pero nunca dejaba de quejarse. Era incapaz de abandonar su juego de proclamar sus debilidades y ocultar sus capacidades. La estratagema debe haberle dado resultado, ¿acaso no le valió un paseo gratuito por el Río Grande a mis espaldas?
>
> (Pág. 417.)

Se abandona la voz neutral para dar paso a un intenso relato de los hechos en primera persona. Se dejaba constancia de resultados como: la autoafirmación, la autoestima y la autoconciencia, en manifestaciones personales. Esta versión de los hechos proporcionaba otro nivel de comprensión y de significado, diferente de los hallazgos relativos a que el cursillo *Outward Bound* tenía una influencia positiva en la "autoafirmación" y "autoestima" de los participantes, sin afectar de forma mensurable a la "aceptación de los demás" o a la "autoconciencia" de los participantes, que fueron los resultados de la medida de las cuatro variables en la parte cuantitativa de la evaluación. No obstante, ambas secciones mostraban una estructura estética. Juntos, los dos componentes proporcionaban una respuesta muy cualificada, pero también creíble y persuasiva, a la pregunta planteada en el primer párrafo del estudio.

Los dos análisis juntos sólo cuentan una misma historia de lo ocurrido, y esa misma historia podría haberse presentado en distintas formas dramáticas. Por ejemplo, la estructura general podría haber consistido en una narración dramática completa, con los datos cuantitativos repartidos a lo largo de la misma para apoyarla. El informe podría haber consistido primordialmente también en un análisis cuantitativo con explicaciones de los efectos que se incluyeran en la sección de "discusión". También podría haberse redactado

© Ediciones Morata, S. L.

Scanned by CamScanner

el informe como un discurso socrático o como una obra teatral. SMITH y cols. (1976) optan por otorgar un peso igual a lo cuantitativo y lo cualitativo.

Con independencia de la forma, el argumento de la evaluación sería el mismo, aunque difiriera en cuanto a su fuerza estética. La credibilidad de la evaluación depende, en cierta medida, de la correspondencia: de la parte cuantitativa con los principios de la medida; de la parte cualitativa con la experiencia personal de los destinatarios y de las partes entre sí: su coherencia. Si estos elementos se ajustan adecuadamente entre sí y la evaluación adopta una forma estética, toda ella se considerará muy creíble. En este caso, las partes muestran una coherencia mutua y es fácil que los destinatarios se convenzan de la eficacia del entrenamiento de *Outward Bound.* El único fallo está en que la parte cuantitativa y la cualitativa no están unidas explícitamente, por lo que la unidad global no resulta tan intensa como podría haber sido.

Con independencia de que una evaluación abunde en metáforas o en coeficientes, estructura dramática o economía matemática, toda evaluación debe cumplir algo fundamental: contar una trama. Ésta puede ser explícita, truncada o implícita, pero siempre tiene que existir para poder extraer conclusiones. Una evaluación debe interpretar hechos sucedidos en el transcurso del tiempo.

Con frecuencia, en muchas evaluaciones cuantitativas la trama se da por supuesta. Por ejemplo, en la evaluación cuantitativa de *Outward Bound,* se presume la existencia de un programa de algún tipo, una serie de acontecimientos, y que los instrumentos de la evaluación miden los resultados de ellos. El objetivo del diseño experimental consistía en descubrir si los presuntos acontecimientos eran las causas de los cambios observados en las medidas. En tal caso, se derivan en todo ello determinadas inferencias. La estética agradable de la evaluación depende del uso apropiado del diseño experimental, la estadística, etcétera.

En la sección cualitativa de la evaluación, los acontecimientos se representan de forma personal, en vez de darse por supuestos. La estética está en las imágenes y en la estructura dramática. Los datos procedían de la observación directa en relación con categorías cotidianas, en vez de derivarse de interpretaciones de instrumentos sobre la base de categorías científicas. La trama presentada moviliza al lector y, en igualdad de condiciones, cuanto mejor sea la presentación estética de la misma, más se movilizará al lector.

Autenticidad

> Una obra de arte no es auténtica ni veraz en virtud de su contenido ni por su "pura" forma, sino por el contenido convertido en forma.
>
> Herbert MARCUSE: *The Aesthetic Dimension,* 1978, pág. 8.

¿Por qué son tan importantes los elementos estéticos como las imágenes y la estructura dramática? Hay muchas explicaciones, en su mayoría extraídas de teorías de la estética o del arte, pero no existe una explicación única aceptada (DICKIE, 1971). Es obvio que contemplamos nuestras vidas en forma de escenarios y argumentos, que las imágenes están configuradas en el lenguaje de nuestros pensamientos cotidianos. Las imágenes concretas constituyen la trama que compone nuestra mente y nuestra memoria.

© Ediciones Morata, S. L.

El antropólogo TURNER (1973) afirmaba que el afecto es "en sí mismo concreto, particular y está asociado con la relación única del yo con su ambiente objetivo" (pág. 354). Los principios abstractos son inadecuados para actuar en el mundo real. Esa integración en el ambiente se consigue mediante condensaciones concretas, afectivas, que sólo pueden proporcionar las imágenes, los símbolos, los rituales y los mitos. Estos instrumentos transmiten "fuerza afectiva y motivadora" en el marco del significado subjetivo de la persona.

De igual modo, SULLIVAN (1977) indicaba que el papel de la imaginación consiste en proporcionar imágenes y mitos integradores. El individuo actúa siempre en un contexto personal e histórico concreto. Las teorías de la acción y de la conducta muy formales y abstractas pasan por alto este nivel más profundo de la motivación contingente[1].

Aunque tales análisis son perspicaces, quizá subestimen la fuerza mítica de la misma ciencia. A veces, las personas actúan sobre la base de resultados que consideran extraídos de forma "científica". La ciencia tiene sus propios mitos y su propia autoridad mítica. Y la necesidad de una imaginería concreta no explica los efectos de las representaciones estéticas más abstractas.

Por ejemplo, ¿por qué es tan eficaz la forma dramática? Según BROUDY (1972), aunque la forma dramática no equivale al significado moral o metafísico, es su primer "indicio". La verdad literal suele ser confusa e insignificante. Debe convertirse en "ficciones plausibles". Las ilusiones son necesarias para el significado humano. Las interpretaciones estéticas se convierten en búsquedas de significado de los acontecimientos humanos. Es más, con frecuencia, las personas juzgan la credibilidad de los hechos que se les co-

[1] En estética existen teorías similares. El objetivismo de la teoría de PIAGET tiene sus análogos en filosofía moral y en estética. La teoría de la justicia de RAWLS (1971) se predica sobre la base de una "situación original" en la que un "velo de ignorancia" limita el conocimiento y las motivaciones que deben tener las personas cuando escogen sus principios de justicia. Los sujetos que se hallan en las situaciones originales hipotéticas desconocen su posición social o sus particulares talentos. Además, son racionales y no altruistas. En otras palabras, el interés individual se reduce en cuanto motivación y podemos presumir que las personas actuarán de manera imparcial. Como en la psicología de PIAGET, se concibe al actor como ser impersonal, abstracto, carente de motivaciones contingentes.

En la filosofía de la estética, la postura correspondiente recibe la denominación de objetivismo fenomenológico (GOTTSHALK, 1962). Un individuo debe ver un objeto desde la adecuada actitud estética, que incluye el desplazamiento de lo práctico y la pérdida de la conciencia del yo. Asumiendo esa actitud, el individuo es capaz de mostrar objetivamente cómo ciertas características de la obra provocan determinadas emociones (BEYER, 1974). En contraste con esta perspectiva está la postura que sostiene que el significado de un objeto está relacionado con el interés del individuo al ver el objeto (BEYER, 1977).

© Ediciones Morata, S. L.

Scanned by CamScanner

municanan sobre la base de criterios estéticos como la viveza y la unidad intrínseca, sobre las apariencias.

Por tanto, una experiencia estética es *una* experiencia y no mera experiencia. El artista selecciona acontecimientos cuyo impacto es acumulativo. La tensión que surge de la estructura dramática rige la atención del público. Por ejemplo, en el estudio del conductor ebrio, el investigador mantiene la atención entre las dos imágenes del conductor borracho que compiten entre sí. Desde el punto de vista de BROUDY, la alternativa de la representación imaginativa es la insignificancia. La presentación estética trata de captar la "esencia" de los acontecimientos más que su mera realidad literal.

También POLANYI considera que la cuestión subyacente es la significación:

> Nuestras vidas carecen de forma, sumergidas en un centenar de encrucijadas. Las obras de arte son representaciones imaginativas, labradas en modelos artificiales; y estos modelos, cuando se integran juntamente con una constante importante, producen un significado de una calidad característica. Estos modelos artificiales hacen... que las obras de arte queden aisladas del flujo de la experiencia personal y de la vida pública. Ellos hacen que las obras de arte sean algo desligado, en muchos casos transportable y reproducible y, en potencia, inmortal.
>
> (Pág. 101.)

Desde una perspectiva más radical, MARCUSE (1978) hacía hincapié en la importancia del sentimiento personal y de la forma artística en la reforma social. La subjetividad —la historia interna y personal del individuo— puede ser una fuerza liberadora si se expresa en una obra de arte. La "realidad dada", la realidad configurada y verificada por las instituciones sociales dominantes, puede reconfigurarse en la obra de arte. Ésta constituye una realidad alternativa en oposición a la realidad "dada". Así, la obra de arte desafía el monopolio de las instituciones establecidas respecto a la definición de la realidad.

En este proceso, la forma estética convierte el contenido en un conjunto nuevo con entidad propia. Esta nueva visión revela las dimensiones de la realidad reprimidas en una "realidad ficticia" que capta sólo la esencia de los hechos. Por medio de la concentración, la exageración, el interés por lo esencial y la reordenacion de los hechos, se reestructura la conciencia de la audiencia. La experiencia se intensifica.

El contenido de la obra es un sentimiento subjetivo y la forma subjetiva convierte el contenido en un conjunto con entidad propia, que permite que la obra se enfrente a la "realidad dada". Cuando el contenido se convierte adecuadamente en forma, la obra adquiere el carácter de autenticidad. Por eso MARCUSE puede decir: "La auténtica utopía se funda en la memoria" (pág. 73).

© Ediciones Morata, S. L.

La belleza del conductor ebrio

> La sabiduría, la habilidad para llegar al núcleo de la cuestión en situaciones concretas, se adquiere despacio; es una disciplina de experiencia, imaginación e historia, no de inteligencia desnuda.
>
> M. NOVAK, *Ascent of the Mountain, Flight of the Dove,* 1971, pág. 58.

El realizador de documentales Frederick Wiseman, muy conocido por sus películas sobre instituciones sociales, como el sistema de bienestar y la industria de envasado de carnes, describía sus documentales para la televisión como "ficción de la realidad", como una "estructura ficticia con la ilusión de la verdad" (O'CONNER, 1976). A partir de la "masa informe" de cincuenta horas de película, Wiseman da forma al documental en el proceso de montaje, en el que introduce los elementos de ritmo, estilo y estructura. Según Wiseman, aunque la selección se hace arbitrariamente, el realizador trata de captar una parte de la "verdad", tal como él la ve.

Sin embargo, la evaluación no es ficción o, al menos, no debe serlo. Las afirmaciones que se hacen respecto a las artes, incluso al arte documental, no tienen por qué ser las que den forma a la evaluación. De un informe de evaluación esperamos mayor veracidad que de una novela o película. Volviendo a la pregunta original, ¿cómo podemos justificar una presentación artística de los datos que afecta a la misma interpretación del estudio en cuestión?

LANGER (1942) decía que hay, al menos, dos tipos de significados: el literal y el artístico. El literal se refiere al contenido explícito y se expresa mediante proposiciones. Pueden ser verdaderas o falsas. El significado artístico versa sobre la forma tácita y se expresa como sentimiento. Puede ser adecuado o inadecuado. En consecuencia, la referencia concreta y la significación poética son dos relaciones muy diferentes entre símbolos y sus significados, pero ambas pueden identificarse.

En la teoría de LANGER, la ciencia avanza gracias al pensamiento discursivo que puede reducirse a conceptos analíticos y a "hechos". LANGER indica que el uso discursivo de los símbolos, que alcanza su clímax en la ciencia, sólo es una forma de conocimiento. Hay formas no discursivas de conocer, como las visuales; formas en las que los elementos del conjunto se presentan de modo simultáneo, como en las pinturas, imágenes, metáforas y mitos. En cuanto tal, el significado artístico no está sometido a tests de falsificación similares a los que existen en los enunciados proposicionales.

En cambio, la forma artística es más comprensible en términos de consistencia, coherencia, economía y elegancia. No obstante la utilidad de los símbolos discursivos, no tenemos más remedio que acudir a la forma artística para expresar el sentimiento. La teoría del arte de LANGER es una teoría de la "imitación", en la que la forma artística actúa como análogo de la emoción, del mismo modo que el lenguaje discursivo opera como análogo lógico del mundo físico (DICKIE, 1971). Con independencia de la fuerza de la afirmación,

© Ediciones Morata, S. L.

Scanned by CamScanner

es innegable que el significado artístico puede oponerse, a veces, al literal. Es posible, sin duda, que los usos inadecuados de la metáfora, los hechos erróneamente reconstruidos o los datos omitidos pueden desorientar al público.

Por ejemplo, *The Final Days*, de WOODWARD y BERNSTEIN, suscitan cuestiones sobre su veracidad. Se levantó una polémica acerca de la veracidad de los hechos mencionados e, incluso, sobre hasta qué punto podían conocer los autores determinados sucesos íntimos. Algunos participantes dijeron que los autores habían dejado de lado detalles discordantes con el fin de intensificar el drama sobre la desintegración de Nixon. La credibilidad de los autores se resintió porque el libro resultaba demasiado dramático. Con independencia de sus méritos en este caso concreto, no es difícil apreciar cómo puede sacrificarse, a veces, la verdad literal a la forma dramática. A pesar de que John KEATS opine lo contrario, en ocasiones, la verdad y la belleza son incompatibles.

Más allá de la veracidad de la obra se sitúa la pregunta de si es equitativa respecto a las personas que representa. En otro lugar he tratado la cuestión de la justicia relacionándola con el atentado contra el propio derecho a la autoestima (HOUSE, 1976). No cabe duda de que la relación entre verdad y justicia en *The Final Days* es discutible. Incluso columnistas liberales han criticado que los autores fueron demasiado lejos en su representación de Nixon como persona en proceso de desintegración.

Una evaluación puede verse gravemente amenazada si se comprueba que uno de sus elementos es falso o de origen discutible. ¿Qué ocurriría, por ejemplo, en la evaluación de *Outward Bound* de SMITH y cols. (1976), si descubriéramos que quien escribió el excelente relato en primera persona no ha experimentado en realidad la situación? La autenticidad de la obra se resentiría si el redactor hubiera reconstruido las aventuras de un tercero.

¿Y qué decir si el viaje a las Montañas Rocosas se basara, en realidad, en tres viajes efectuados en tres ocasiones por tres personas distintas, y, por razones estéticas, el autor hubiese combinado las tres experiencias diferentes en una única narración dramática? La ventaja es evidente. De ese modo, la narración transmitiría el tono del sentimiento de la experiencia de la montaña —veracidad de sentimiento— con una fuerza que no sería posible alcanzar con tres relatos distintos. Esto resultaría positivo para la coherencia de la obra, su credibilidad y su capacidad de persuasión.

Por otra parte, esa narración no sería verdadera, en sentido literal. ¿Estaría justificado que el autor se tomara tales libertades? Debería darse una correspondencia entre la veracidad literal y la de sentimientos, entre verdad y belleza. Aún con la importancia que la belleza tiene en la evaluación, si se produce un conflicto entre verdad y belleza, la primera debe predominar. En último término, la evaluación no es ficción, aunque las imágenes y la forma sean necesarias para transmitir significados complejos. Es posible el equilibrio entre verdad y belleza. Podemos omitir ciertos hechos para que el informe sea más coherente, más legible. Pero sólo sacrificaríamos una pequeña

© Ediciones Morata, S. L.

verdad a cambio de una belleza de gran magnitud. La verdad es relativamente más valiosa.

Para mí, la solución gira en torno a la idea de la trama evaluativa. Pueden permitirse libertades en la estructura dramática siempre que no se modifique la trama evaluativa. O sea, ¿unas personas fueron a las montañas y experimentaron estas cosas, aunque no exactamente como aparecerían en un planteamiento literal? Si esas personas no experimentaran cosas similares o si el lenguaje figurativo sugiriese estados sentimentales diferentes de los inducidos por el viaje a las montañas, la licencia sería excesiva en una evaluación. Los destinatarios de la misma podrían verse inducidos a error de forma significativa y tendrían una representación falsa del programa. Asimismo, tendrían que decirnos que los acontecimientos relatados habrían sido reconstruidos. Podríamos comprobar la veracidad por la consistencia interna de la narración, por la coherencia con el análisis cuantitativo y por su congruencia con la propia experiencia del lector.

Tampoco los estudios cuantitativos se ven libres de este conflicto. Las técnicas de medida y análisis se basan a menudo en criterios estéticos y, a veces, se traducen en aplicaciones inadecuadas. Es más, las secciones dedicadas a la metodología de la mayoría de los estudios están redactadas desde una perspectiva reconstructiva de lo ocurrido más que desde la forma real de recoger y analizar los datos. Esta reconstrucción debería tener un límite fijado en el punto en que la historia de la evaluación se modifica de manera significativa. Por ejemplo, los destinatarios pueden suponer que se hayan adoptado las precauciones habituales en la recogida de datos cuando, en realidad, no se hubieran tomado. O esos mismos destinatarios pueden creer que se ha efectuado algún tratamiento cuando, de hecho, su implementación sería cuestionable. Como en un estudio cuantitativo los destinatarios dan por supuestas tantas cosas, no aclaradas por los autores, la reconstrucción de lo ocurrido constituye un problema permanente.

Volviendo al ejemplo inicial del estudio del conductor bebido, ¿se justifica que el investigador utilizara imágenes fuertes del conductor ebrio? La respuesta es un "sí" condicionado, si aceptamos que el propósito de una evaluación o estudio político consiste en informar la opinión y conducir a la acción. La condición se basa en que la imagen del conductor ebrio debe ser congruente con el resto del estudio, pues no es la única versión que hubiera podido escoger el autor, y que la perspectiva adoptada refleje adecuadamente el mundo real. Estas versiones son esenciales para la acción moral. Emplean el "vocabulario de la acción" que sigue constituyendo una obligación del evaluador (HOUSE, 1973). No obstante, en los casos concretos en los que verdad y belleza se oponen, la primera es más importante que la segunda, y la justicia, más importante que ambas.

© Ediciones Morata, S. L.

Scanned by CamScanner

CAPÍTULO VI

Justicia

> La justicia es la primera virtud de las instituciones sociales, como la verdad lo es de los sistemas de pensamiento.
>
> John RAWLS: *A Theory of Justice.*

Consideremos una situación típica del evaluador en ejercicio. Un gobierno contrata sus servicios para evaluar un programa educativo dirigido a alumnos desventajados. ¿Cómo actúa? ¿Qué busca? Las posibilidades son mayores de lo que podríamos pensar. Puede administrar un test normalizado de rendimiento y comparar las puntuaciones con las de los alumnos que no participan en ese programa. Puede elaborar un test especial, basado quizá en objetivos, para evaluar determinadas áreas de "deficiencia" académica. Puede medir las actitudes, opiniones y autoconceptos de los estudiantes.

Por otra parte, puede: solicitar las opiniones, actitudes y juicios de los profesores acerca del programa; observar cómo se comportan en clase, en los pasillos, en la calle alumnos y profesores; registrar y analizar lo que los profesores dicen de los alumnos, cómo los clasifican, qué niveles fijan; preguntar a los padres sobre el programa, los profesores, las escuelas; examinar la participación de los padres en las actividades escolares o los esfuerzos de las escuelas para hacer participar a los padres.

Puede: pedir las opiniones de los patronos sobre el contenido o los resultados del programa; comparar el contenido del programa con las destrezas necesarias para las actividades laborales o estudiar las oportunidades de trabajo de los alumnos, o el salario, o la actuación después del programa; examinar las notas de cursos escolares posteriores; pedir a expertos en las áreas correspondientes o a líderes políticos que juzguen el programa; hacer sus propios juicios; montar un pseudojuicio y hacer que juzguen el programa personas del público. Podría hacer muchas otras cosas.

¿Qué hace el evaluador? Influyen en ello un conjunto de factores: su experiencia, su formación, lo que quiere el gobierno, lo que pretenden las personas que participan en el programa, etc. Uno de estos factores influyentes es la concepción de justicia, tácita, por regla general, que tengan el evaluador y el patrocinador de la evaluación. Lo que crea correcto el evaluador y la concepción de justicia que predomina influirán de manera significativa en la evaluación.

Hemos de admitir que la influencia es compleja y nada directa. La evaluación no es una deducción de la concepción de la justicia. La conexión es más vaga, más implícita, pero en absoluto menos influyente. La concepción dominante de justicia limita los enfoques que se adopten, las actividades que se consideren justificadas, los argumentos significativos. Las concepciones de la justicia operan más en calidad de marcos de referencia que como máquinas internamente coherentes para deducir conclusiones. Distribuyen el peso de los argumentos de determinadas formas.

Las técnicas y las prácticas de evaluación no forman, en sí mismas, parte de una determinada teoría de la justicia, como el utilitarismo o la justicia como equidad de RAWLS. Por ejemplo, la regresión múltiple no es, de por sí, utilitaria, aunque un evaluador que creyera que son pocos los indicadores cuantitativos de éxito acabaría utilizando una y otra vez la regresión múltiple. Quien pensara que en una evaluación deberían estar representados de forma directa los intereses de diversas clases de público emplearía la entrevista mucho más que un utilitarista. En sí, la técnica no es utilitarista, sino el razonamiento que lleva a ciertas opciones que ha de hacer el evaluador.

En parte, el evaluador actúa por intuición y la filosofía puede agudizar las intuiciones y sensibilidad de aquél. Las intuiciones del evaluador no sólo se refieren a la justicia, sino al propio interés, y la filosofía puede hacer más agudas las mejores intuiciones mediante la reflexión sobre las teorías de la justicia. Las personas pueden justificar las evaluaciones en relación con las concepciones y las teorías de la justicia, y así lo hacen muy a menudo, aunque de la forma intuitiva aquí indicada.

El evaluador tiene que escoger. Si piensa que una mayor cantidad de puntuaciones de tests normalizados se traducirá en mayor escolarización, un trabajo mejor, más dinero y, quizá, mayor satisfacción de los estudiantes —en pocas palabras, que está convencido de las ventajas de los tests— los utilizará como indicadores de éxito del programa. Si cree que la participación de los padres es fundamental o que los intereses de los estudiantes están mejor representados por las opiniones de sus padres, llevará a cabo una evaluación muy diferente. Quizá todavía utilice tests normalizados. La

NOTA DEL AUTOR: Este capítulo constituye una amplia modificación de las ideas contenidas en "Justice in Evaluation", en: G.V. GLASS (ed.): *Evaluation Studies Review Annual,* Vol. 1, Beverly Hills: Sage Publications, 1976; y "The Role of Theories of Justice in Evaluation", *Educational Theory,* Vol. 29, n.º 4, 1980. Este último es una respuesta a la crítica de STRIKE (1979).

© Ediciones Morata, S. L.

© Ediciones Morata, S. L.

Scanned by CamScanner

forma de representar los intereses del alumno y de quiénes sean los intereses registrados se traducirá en evaluaciones significativamente distintas. El hecho de que se representen intereses diversos es una cuestión que atañe a la teoría democrática, pero hay muchas concepciones de la justicia coherentes con un punto de vista democrático. Por ejemplo, un evaluador neomarxista quizá busque desigualdades presentes en el medio escolar, mientras que un pluralista político procurará representar las opiniones y valores de muy diversos grupos.

La evaluación es una actividad política por naturaleza. Presta sus servicios a los responsables de las decisiones, se traduce en redistribuciones de recursos y justifica que unos reciban ciertas cosas y otros otras. Está íntimamente vinculada con la distribución de los bienes básicos en la sociedad. Es algo más que un enunciado de ideas; se trata de un mecanismo social para la distribución, que aspira a adquirir una categoría institucional. La evaluación no sólo debe ser veraz, sino también justa. Sus esquemas al uso, con independencia de su valor de verdad, reflejan la justicia en grados muy diversos. Y la justicia constituye una norma importante con respecto a la cual juzgar la evaluación.

En este capítulo, presento tres concepciones de justicia —la utilitaria, la intuicionista pluralista y la justicia como equidad de RAWLS—, relacionándolas con los enfoques actuales de la evaluación. Trato de señalar algunos puntos fuertes y débiles de las distintas concepciones y de llegar a establecer un juicio global sobre la que es preferible. Como señala STRIKE (1979), una teoría de la justicia propia del sujeto no determina un único enfoque de la evaluación. No obstante, creo que la teoría del sujeto limita los enfoques posibles de la evaluación y se utiliza para justificar esos enfoques (HOUSE, 1979).

Utilitarismo

Según RAWLS (1971), la ética utilitarista estipula que una sociedad es justa cuando sus instituciones están organizadas de manera que se consiga el máximo balance neto de satisfacción, teniendo en cuenta el conjunto de los individuos. El principio de utilidad consiste en elevar al máximo ese balance de satisfacción. El utilitarismo requiere la existencia de una medida o índice común de satisfacción que haga posible efectuar cálculos cuantitativos. En educación, esta medida común está constituida casi siempre por las puntuaciones obtenidas en tests normalizados. Es el índice sustituto (o vicario) de satisfacción.

La estructura básica de una teoría ética viene dada por la forma de establecer la relación entre dos conceptos básicos: el bien y el derecho. En las teorías teleológicas —entre las que figura el utilitarismo— el bien *se* define de modo *independiente y antecedente* respecto al derecho. Después, el derecho se convierte en lo que maximiza el bien (ya sea la satisfacción, la excelencia, o cualquier otra cosa). La misma racionalidad significa elevar

© Ediciones Morata, S. L.

el bien al máximo, un concepto con una gran atracción intuitiva. Se considera que no procurar el máximo bien es, como mínimo, ineficaz, poco inteligente e incluso irracional. En la ética utilitarista, es *inmoral*.

Por ejemplo, en el enfoque del análisis de sistemas aplicado a la educación, se toman las puntuaciones en los tests como índice del bien. Aquello que las maximiza es, a la vez, lo que se debe y lo que conviene hacer. Los mejores programas educativos son los que, entre las alternativas disponibles, consiguen los mayores incrementos en las puntuaciones obtenidas en los tests. El diseño experimental, la instrumentación y la estadística son medios para determinar eficazmente las "mejores" alternativas, mediante el establecimiento de la relación lógica y cuantitativa entre lo ajustado a derecho y el bien. La conveniencia social se convierte en el tema de maximizar las puntuaciones obtenidas en los tests. La mayor parte de las discusiones sobre la evaluación giran en torno a la eficacia de la misma como medio para determinar la veracidad y magnitud de ese incremento.

El efecto de una teoría de este tipo consiste en la división de los juicios morales en dos clases diferentes. El bien, los "juicios de valor", constituyen una clase independiente de juicios que puede especificarse de antemano sin referencia a lo correcto. En sentido tradicional, el "derecho" abarca el problema de la distribución de los bienes básicos. En la economía moderna, muy dependiente del pensamiento utilitarista, se hace gran hincapié en la producción de bienes básicos, prestándose mucha menos atención al modo de ditribución de esos bienes. La atención se centra en el incremento de la producción total y no en la distribución de los bienes. En la evaluación utilitarista, se resalta el incremento de las puntuaciones de los tests y no la distribución de esas puntuaciones.

La justicia utilitarista significa llevar al máximo la satisfacción total. Cada uno de los deseos o esperanzas de los individuos ha de equipararse con el de los otros, con independencia de que uno tenga mayores aspiraciones y otro menores. En cierto sentido, si alguien tiene mayores aspiraciones, tiene derecho a más. Es posible suprimir satisfacciones a quienes tienen menos en beneficio de alguien que "necesita" más. Este tipo de transacción es una de las causas del fracaso de la economía moderna en los países en vías de desarrollo. El producto nacional bruto de un país puede aumentar sustancialmente por medio de nuevas industrias. Sin embargo, el dinero o la misma industria se ajusta a menudo para satisfacer las aspiraciones crecientes de la nueva clase media, mientras las correspondientes a las clases inferiores siguen como estaban. El incremento total de la riqueza no supone la mejora de la parte que a cada uno le corresponde. La estabilidad de las satisfacciones de las clases inferiores en beneficio del aumento de las deparadas a las superiores es coherente con la justicia utilitarista. La satisfacción global se maximiza.

Tanto en educación como en otras áreas de tipo social, a menudo las pretensiones (necesidades, objetivos, etc.) se dan por supuestas. Preparamos "evaluaciones de necesidades" para determinar éstas lo más objetiva y cuantitativamente posible. En nuestra teoría de la medida, basada en las

© Ediciones Morata, S. L.

Scanned by CamScanner

diferencias individuales, las necesidades son intrínsecas y permanecen a la espera de ser medidas, clasificadas por orden de prioridad y equilibradas en relación con otras necesidades. Si las personas de clase social elevada albergan muy grandes aspiraciones respecto a sus hijos, todas ellas entrarán en el análisis de necesidades y tendrán que ser satisfechas. Su cumplimiento quedará reflejado en la medida común del rendimiento y en todos los bienes y valores que en la sociedad dependen de esas puntuaciones. La distribución entre las personas no es tan importante como la suma total. El lema es elevarlo al máximo en todas partes.

Como sólo se tiene en cuenta la puntuación neta final, la disminución correspondiente a una persona se compensa con el incremento de otra. Los estudiantes de clase social superior tienen que incrementar al máximo sus puntuaciones de rendimiento y elevar su posición social, aunque sea a costa de los alumnos de clase inferior. Todos tienen libertad para aumentar su propio bien y, como lo justo consiste en elevar al máximo el bien, las acciones de la clase social superior son al tiempo buenas y justas.

Como todas las aspiraciones se valoran en cuanto a su apariencia inmediata y las esperanzas y deseos son mucho mayores en las clases sociales y económicas superiores, el utilitarismo, en cuanto sistema de justicia, lleva consigo la tendencia a favor de la clase alta. El deseo de un Rolls Royce que tenga una persona es tan digno de satisfacción como el que otra experimente respecto de un Ford. Las mayores exigencias de los estudiantes de clase alta o media se satisfacen a costa de las clases inferiores.

En el utilitarismo, es esencial que todo se encuadre en un sistema de aspiraciones, de manera que pueda cumplirse la justicia. Todo debe reducirse a una unidad de medida que haga posibles las comparaciones. El método básico de comparación es el del clásico "espectador imparcial" utilitario. Al ser imparcial y simpatizar con todas las partes, el espectador (evaluador de necesidades) puede organizar todas las aspiraciones (necesidades) en un sistema coherente de aspiraciones (conjunto de objetivos). Sopesa y equilibra la intensidad de las aspiraciones y les otorga el peso adecuado en el conjunto del sistema (evalúa). Después, el administrador o legislador (el omnipresente responsable de las decisiones) ajusta los medios limitados con el fin de maximizar la satisfacción de aquellas aspiraciones. El resultado es una administración eficaz con la máxima satisfacción posible.

La semejanza del evaluador utilitarista con el espectador imparcial es grande. En la economía y en la evaluación modernas, el espectador imparcial funda en alto grado su imparcialidad en medidas cuantitativas. Los económetras modernos reclaman datos unívocos para medir la eficacia de los programas sociales (RIVLIN, 1971). Estos datos pueden convertirse en el diseño más eficaz para la sociedad. El evaluador ideal también registra sin animosidad las aspiraciones mediante las evaluaciones de necesidades y las definiciones de objetivos. Asigna pesos y mide los resultados de un modo imparcial.

Mediante técnicas de investigación conocidas, el evaluador combina

© Ediciones Morata, S. L.

todas las necesidades educativas en un sistema. Dice RAWLS: "El utilitarismo no toma en serio las diferencias entre las personas, confundiendo la impersonalidad con la imparcialidad." Pero este reduccionismo es esencial para que la justicia utilitarista determine lo que se debe y conviene hacer.

Como indiqué en el Capítulo III, los modelos de evaluación de análisis de sistemas, de evaluación por objetivos, de decisión y el modelo de objetivos independiente parecen fundarse en una concepción utilitarista de la justicia. Todos ellos tratan de conseguir la mayor utilidad elevando al máximo un número reducido de variables, un conjunto de objetivos, las preferencias del responsable de las decisiones o sopesando resultados. Procuran llegar a un juicio de la utilidad social general que, a su vez, conduce a la máxima felicidad en la sociedad. Como en la economía, el coste constituye una consideración primordial, dado que el dinero gastado podría utilizarse en otras cosas para aumentar la felicidad; o sea, podría mejorarse la utilidad.

El utilitarismo asume que la esencia de la conducta racional consiste en elevar al máximo las satisfacciones o utilidades individuales (MACPHERSON, 1966). La esencia humana consiste en la acción racional que maximiza la utilidad. Como las aspiraciones de las personas respecto a toda clase de satisfacciones son ilimitadas y los medios para realizarlas siempre son escasos, el problema consiste en descubrir el sistema que pueda emplear los medios escasos para lograr la máxima satisfacción. El bien último consiste en conseguir la mayor utilidad.

El utilitarismo se basa en la idea de que todo el mundo procura conseguir el máximo posible. La mayor utilidad supone hallar una fórmula mediante la cual las personas consigan satisfacciones con el mínimo esfuerzo. Esto plantea problemas en relación a cómo identificar y reunir las satisfacciones o utilidades que diferentes cosas procuran a las personas, y cómo compararlas mediante una única escala de medida. Esto es necesario para poder decir que una determinada ordenación de utilidades es la mejor, para afirmar que un conjunto de ellas contribuye a una utilidad total mayor (MACPHERSON, 1966).

Entre los enfoques utilitaristas de la evaluación, el análisis de sistemas trata de identificar un conjunto crítico de variables en cuanto indicadores sociales; el enfoque por objetivos parte de los objetivos del programa; el de decisión señala como elementos críticos las decisiones clave de los responsables, y el evaluador que prescinde de los objetivos se centra en los efectos concretos del programa. Cada modelo adopta un enfoque ligeramente distinto para lograr la máxima utilidad.

El utilitarismo ha constituido una forma brillante y productiva de pensamiento; pero, como parte de las demandas tal como están e insiste en una medida común de bienestar, favorece con frecuencia a las clases sociales elevadas en detrimento de las inferiores. Cuando la medida comun se convierte en indicadores como los tests normalizados de aprovechamiento, el problema se complica, porque esos mismos tests pueden estar sesgados desde los puntos de vista social y racial; se basan en una teoría de las diferencias individuales; desde el punto de vista histórico, se han utilizado

© Ediciones Morata, S. L.

Scanned by CamScanner

con el propósito explícito de seleccionar personas, apartándolas de los grupos sociales (KARIER, 1973), y pueden emplearse para justificar las desigualdades. La justicia utilitarista no coincide con muchas sensibilidades morales.

Pluralismo/Intuicionismo

La segunda teoría de la justicia es la pluralista/intuicionista. Esta teoría afirma que existe una familia irreductible de primeros principios que deben sopesarse entre sí buscando el equilibrio más justo entre ellos (RAWLS, 1971). No hay un principio fundamental, como el de la utilidad del utilitarismo, que puede servir para determinar los pesos que se asignen a los principios últimos. Por tanto, se da una pluralidad de primeros principios por medio de los cuales se efectúan los juicios sobre la justicia y no existe ningún método o conjunto explícito de reglas de prioridad para sopesar entre sí dichos principios.

La mayor parte de las ideas de justicia de "sentido común" son de este tipo, como también la mayoría de las doctrinas filosóficas formales. Por regla general, hay distintos grupos de preceptos específicos que se aplican a problemas concretos de justicia, como los impuestos o la conducta criminal. Estos preceptos específicos se aplican de manera más o menos intuitiva dentro de áreas definidas, equilibrando unos los principios fundamentales de una manera y otros de otra. Los pesos relativos de los principios específicos pueden llevar a los más intensos debates (como ocurre con la libertad de expresión) aunque las personas estén de acuerdo en los principios fundamentales. La costumbre y los intereses de las partes implicadas conducen a otorgar pesos distintos.

Sin embargo, el equilibrio intuitivo entre los principios no proporciona medio alguno de juzgar la justicia de las prácticas cotidianas. El intuicionista espera que las personas se pongan de acuerdo en los principios, una vez identificados, o que al menos acuerden algún procedimiento que permita otorgar pesos. Al mismo tiempo, el pluralista/intuicionista sostiene que no existe ninguna concepción ética suprema en la que se funden esos pesos. La complejidad de los hechos morales es demasiado grande para que un principio de ese estilo dé cuenta de ellos.

Para las escuelas de evaluación que he denominado pluralistas/intuicionistas, el origen y el contenido de los principios de justicia son distintos. En el modelo de evaluación crítico artístico, el crítico educativo aplica principios a los que ha llegado a través de su propia experiencia y preparación en los programas educativos. Podemos suponer que cada crítico tendrá su propio conjunto de principios que aplicará a una situación dada de manera intuitiva. Como el crítico de arte, el educativo llega a un juicio complejo del programa o normativa.

El modelo profesional utiliza un conjunto de principios a menudo aceptados de antemano. Por ejemplo, *Evaluative Criteria* menciona muchos cri-

© Ediciones Morata, S. L.

terios aceptables para juzgar los programas de inglés, otro conjunto para juzgar los de matemáticas, etc. Los expertos en inglés o en matemáticas aplican estos criterios establecidos -y quizá otros- de forma intuitiva. Cada experto equilibra los principios de acuerdo con sus propias disposiciones.

De un modo similar, el *Council on Program Evaluation* (COPE) presenta una larga lista de criterios que emplea para juzgar los departamentos universitarios. En realidad, estos criterios pueden resumirse en unos pocos principios, como la calidad de investigación y de enseñanza. A partir de ahí, el *Council* establece un conjunto de juicios mediante el equilibrio intuitivo entre esos principios. Por regla general, estos juicios otorgan gran peso a la investigación. Tanto en el modelo de tipo crítico artístico como en el profesional, los criterios, los principios orientadores últimos y los pesos proceden de los profesionales de la educación.

En cambio, los modelos de evaluación cuasijurídico y de estudio de casos suelen derivar sus principios y valores de la participación de las personas implicadas en el programa o en las decisiones sobre el mismo. En el modelo cuasijurídico que emplea un procedimiento parajudicial, el jurado y el juez determinan los criterios y principios que se utilicen para llegar a una decisión. Los principios concretos que se empleen dependen de la forma de seleccionar el jurado, el juez o el tribunal de audiencia. Puede tratarse de miembros del público en general, como en un jurado auténtico, o responsables de decisiones encargados del programa en cuestión.

En muchos casos, como en un tribunal de audiencia *"blue-ribbon"* (tribunal de expertos)*, éste puede estar compuesto por miembros prestigiosos del público o de un subgrupo profesional, a quienes se considere en general con autoridad en la cuestión de que se trate, como un grupo de expertos en temas fiscales, evaluadores o responsables de decisiones. Por ejemplo, el *Dissemination Review Panel*, en las evaluaciones de revisiones de programas federales del departamento de Salud, Educación y Bienestar, para determinar si estos programas deben ser "validados" para su divulgación por el resto del país.

Por último, entre los pluralistas/intuicionistas hay quienes valoran la participación de las personas implicadas en el programa. En general, ésta consiste en la recogida de puntos de vista y opiniones de diversas personas respecto al programa o normativa en cuestión. El evaluador registra y refleja fielmente sus puntos de vista. De este modo, los principios, criterios y valoraciones de las personas participantes se utilizan para juzgar el programa. Por supuesto, al seleccionar y resaltar algunos aspectos y puntos de vista en detrimento de otros —algo inevitable—, entran en juego algunos principios del propio evaluador.

---

* Se trata de un jurado formado por expertos para juzgar asuntos muy especializados. Es frecuente (por ejemplo, en Estados Unidos) en la resolución de conflictos en los ámbitos empresarial y médico. *(N. del R.)*

© Ediciones Morata, S. L.

Scanned by CamScanner

Al reflejar los juicios de distintos grupos relacionados con el programa, este último enfoque no sólo es pluralista en el sentido filosófico de juzgar sobre la base de principios diversos, sino también en el sentido político de representar intereses políticos diferentes. Es obvio que los juicios de los participantes y del evaluador se equilibran casi siempre de manera intuitiva.

La teoría pluralista/intuicionista de la justicia es la que más se acerca al sentido común, pero deja sometida la evaluación a cualesquiera principios y valoraciones que se les ocurra utilizar a los jueces, con independencia de quienes se trate. En un extremo, existe el peligro de que el proceso de juicio se rija por principios profesionales y, en el otro, de un completo relativismo en el que se considere tan buena la opinión de uno como la de cualquier otro. Ninguno de los dos extremos coincide con la idea de justicia de la mayoría.

Justicia como equidad

La más reciente de las teorías modernas sobre la justicia es la de RAWLS (1971). Este autor menciona dos principios de justicia mediante los cuales pueden ser juzgadas las instituciones y convenios sociales. RAWLS denomina su concepción "justicia como equidad". A diferencia del utilitarismo, la "justicia como equidad" asume que siempre existe una pluralidad de fines y una distinción de personas que impiden combinar todas las aspiraciones en un sistema. En primer lugar, hay que llegar a un acuerdo que permita repartir el bien y saldar las disputas que surjan. Más tarde, acordados los principios de reparto, los individuos tienen libertad para determinar su propio bien y tratar de conseguirlo, pero siempre de acuerdo con los principios aceptados que determinan lo que se ajusta a derecho.

Al definir cómo se llega a estos principios, RAWLS utiiza el constructo hipotético de "la posición original". Suponiendo que todas las personas sean iguales, desde el punto de vista moral, y se les conceda un tratamiento justo (y desconozcan qué ventajas y desventajas tendrán en la sociedad concreta), ¿qué principios estarán dispuestas a aceptar para definir la justicia?

La concepción general de justicia de RAWLS es la siguiente: "Todos los valores sociales —libertad y oportunidad, ingresos y salud y las bases del respeto a sí mismo— se reparten por igual, salvo que la distribución desigual de alguno, o de todos estos valores beneficie a alguien" (RAWLS, 1971, página 62).

Los principios específicos de justicia para las instituciones son:

Primer principio:

Toda persona ha de tener igual derecho a los más amplios sistemas de libertades básicas iguales, compatible con un sistema semejante de libertad para todos.

© Ediciones Morata, S. L.

Segundo principio:

Las desigualdades sociales y económicas han de organizarse de manera que:

a) beneficien a los menos agraciados, salvado el principio de justicia, y
b) estén subordinadas a servicios y empleos abiertos a todos en condiciones de equitativa igualdad de oportunidades (págs. 302-303).

Teniendo en cuenta estas premisas, resulta obvio que la justicia como equidad restringe otros fines. Pone límites en relación con lo que podamos hacer. En esa perspectiva, no podemos tomar como indiscutibles los deseos y aspiraciones, pues quedan sometidos a los principios básicos.

La justicia como equidad excluye específicamente la imposición de inconvenientes a unos pocos en beneficio de muchos. Las reglas de prioridad de los dos principios de la justicia especifican que el primero siempre antecede al segundo. Se impulsan al máximo las libertades básicas con independencia de los beneficios económicos y sociales. Sólo entonces pueden abordarse las desigualdades sociales y económicas. De acuerdo con la segunda premisa, *sólo* pueden permitirse esas desigualdades si benefician a los *menos* favorecidos de la sociedad. No pueden aceptarse las desigualdades si no se benefician de ellas los peor situados, ni tampoco pueden aceptarse transacciones que afecten a las libertades básicas a cambio de beneficios sociales y económicos.

La segunda premisa, el "principio de diferencia", singulariza la posición de los menos favorecidos de la sociedad para juzgar si pueden permitirse las desigualdades. Por ejemplo, si son necesarios salarios elevados para atraer a personas a ocupaciones que beneficien a los menos favorecidos, puede aceptarse la desigualdad. (Este juicio se opone al "principio de eficacia" del utilitarismo. El principio de eficacia sólo permite la redistribución de bienes primarios en la medida en que el beneficio de un grupo social no se produzca en detrimento de otro grupo. Esto refuerza de forma considerable el orden establecido. Una estructura será "eficiente" al máximo, o en grado óptimo, cuando sea imposible mejorar algo sin pejudicar a otros. En el esquema de RAWLS, la premisa de la justicia como equidad *precede* al principio de eficacia. Se impulsan al máximo las aspiraciones a largo plazo de la posición menos favorecida, con las restricciones impuestas por la primera premisa.)

Suponiendo por ahora que aceptamos la teoría de la justicia de RAWLS, ¿qué importancia tiene evaluar (si acaso tiene alguna)? Es evidente que ambas premisas son lo bastante abstractas como para no determinar la evaluación. Por otra parte, en cuanto mecanismo redistributivo importante, parece razonable que la evaluación preste atención a la justicia como equidad. Las libertades básicas que protege el primer principio son las libertades políticas, la de expresión y reunión, la libertad de conciencia y la de pensamiento, libertad de personas y propiedades, etcétera.

© Ediciones Morata, S. L.

Scanned by CamScanner

Además, según RAWLS (pág. 440), el bien primario más importante es el respeto a sí mismo. Sin él, la persona no podrá creer que merezca la pena tratar de dar cumplimiento a su propio plan de vida (cualquiera sea éste). En tal caso, su vida carecerá de sentido. Un argumento fundamental contra el utilitarismo consiste en que algunos pueden verse obligados a abandonar sus aspiraciones en beneficio de la utilidad general. La rebaja involuntaria de éstas en provecho de otros reduce el fundamento de su propio respeto a sí mismo. La justicia como equidad no admite esta situación.

El primer principio puede aplicarse a la evaluación de dos maneras; según si tiene en cuenta lo que busque el evaluador o la forma de llevar a cabo la evaluación. El primer principio garantiza las libertades básicas, incluido el derecho a la autoestima. Por ejemplo, una queja fundamental de quienes se oponen a los tests, en concreto, consiste en que las puntuaciones obtenidas en los mismos reducen la autoestima de muchos niños en la escuela. No sólo los críticos dicen que los tests están sesgados en los planos racial y de clase social; los economistas radicales (BOWLES y GINTIS, 1972-1973) sostienen que los tests se utilizan precisamente con el fin de justificar la estructura jerárquica del sistema de clases sociales, de manera que los miembros inferiores de la sociedad crean que merecen lo que son. Llevan consigo la reducción de las aspiraciones.

Con independencia de la veracidad de estos cargos, en la justicia como equidad estas cuestiones son motivo de grave preocupación y para el evaluador constituyen el problema principal. Con frecuencia, es más fácil determinar la conveniencia de una práctica *concreta* y más cercana al cometido real del evaluador. Por ejemplo, mandar a su casa a un niño con su puntuación de CI plantada encima (práctica corriente en algunas grandes ciudades) puede resultar repulsiva incluso a los evaluadores más fervientes partidarios de los tests. Otras prácticas, como la de pegar a los niños para que suban sus puntuaciones en los tests suelen descartarse de antemano (incluso aunque provoquen un ascenso espectacular de tales puntuaciones). Hay cosas que no pueden hacerse en nombre del bien común.

Es posible que la aplicación de ambas premisas reduzca, aunque de ninguna manera elimine, los juicios intuitivos. Tras decidir que el respeto a uno mismo es un derecho que precede a cualquier otra consideración, con independencia de las pretensiones del programa o de sus patrocinadores, todavía tenemos que aplicar el principio. Esto lleva consigo una serie de juicios intuitivos y el cometido del evaluador de aplicar los criterios a circunstancias específicas sigue siendo difícil.

No obstante, la selección de criterios primarios para aplicarlos a la mayor parte del resto de los criterios (por artificial que pueda parecernos a quienes nos hemos formado en una tradición empírica) proporciona un fundamento moral más útil de lo que a primera vista pudiera parecer. A veces, la visión resultante es informativa. Por ejemplo, si todo el mundo tiene derecho a la autoestima, de ello también se desprende que, en una sociedad justa, los menos favorecidos no pueden violar la autoestima de quienes están en mejor posición. No se permiten las actividades denigrantes

© Ediciones Morata, S. L.

para el individuo. Es obvio que ello presupone que la autoestima de los más favorecidos no se basa en la disminución de la estima de los peor situados.

Si la información lesiona el respeto a sí mismo de la persona implicada, no debe incluirse en el informe de evaluación, con independencia de su influencia en la calidad del proyecto. Han de protegerse los derechos de los individuos antes que el beneficio del proyecto. Sólo podría violarse este principio en el caso de una transacción entre este derecho y otro de igual importancia. Por supuesto, el problema básico de un proyecto está constituido por el *sistema total* de derechos iguales de todos los implicados. No habría que tratar con tanto miramiento a un director despótico que lesione la autoestima de otros. Se permiten las transacciones entre derechos básicos.

Aplicado por completo el primer principio, hemos de prestar atención al segundo. La esencia de la segunda premisa consiste en que las desigualdades sociales y económicas sólo son justas cuando benefician a los menos favorecidos de la sociedad. Un "personaje representativo" del sector menos afortunado debe preferir las perspectivas que se le ofrecen con las desigualdades más que sin ellas. Se juzga la estructura social desde una posición social determinada: la de los menos favorecidos. Las desventajas de quienes están situados en el escalón inferior de la estructura social no pueden justificarse por las ventajas que de ello se deriven para los demás, como acepta el utilitarismo. Por supuesto, este juicio ha de ser coherente con el primer principio. No pueden sacrificarse las libertades básicas a los beneficios sociales y económicos.

Como en el caso del primer principio, el segundo puede aplicarse de dos maneras: como criterio del programa o como criterio de la misma evaluación. En la *Illinois Gifted Evaluation,* nos mantuvimos en contacto directo con el personal que trabajaba en el programa, que fue muy equitativo en relación con el primer principio. Visto en retrospectiva, el principal déficit consistió en que no se investigaron posibles efectos secundarios de la agrupación uniforme de los niños bien dotados. Analizamos los efectos que se produjeron en estos mismos niños, pero no lo hicimos con los que no pertenecían a ese grupo ni consideramos el impacto social sobre el sistema de clases. Hemos de admitir que estos problemas no son fáciles de resolver, en parte porque a menudo no se plantean lo suficiente. Sin embargo, deberían haberse tratado con mayor profundidad.

En la justicia como equidad, no se considera que las desigualdades en los talentos naturales tengan que dar derecho a mayores aspiraciones materiales en la sociedad, salvo que se empleen en beneficio de los menos dotados. Por ejemplo, se podrían conceder mayores recompensas a los médicos por la preparación a la que han de someterse y esperando que el atraer a la medicina a individuos mejor dotados beneficie a los menos favorecidos, así como a cualesquiera otros. Pero el hecho de que alguien esté dotado con mayores talentos naturales no le da derecho a mayores aspiraciones. En cierto sentido, el contrato que surge de la "posición original" consiste en un acuerdo para compartir la suerte y la desgracia, como en

© Ediciones Morata, S. L.

Scanned by CamScanner

una familia: sus miembros no aumentan su buena suerte al máximo a expensas del menos favorecido de la misma.

¿Qué diferencias podría producir el segundo principio de justicia en una evaluación? Consideremos el estudio sobre la igualdad de oportunidades educativas de COLEMAN (1966) que, esencialmente constituye un enfoque liberal de un problema social. Aunque no se trate de una evaluación, en sentido estricto, el estudio de COLEMAN constituye un intento de valoración de la falta de igualdad de oportunidades educativas, un tipo de evaluación de necesidades sobre el que se estructura la política gubernamental. A COLEMAN se le planteó el problema sin una concepción clara del significado de la igualdad de oportunidades. Como señala ENNIS (1975), la concepción de alguien depende del significado que tenga para él la "educación" y de cómo interprete "tener una oportunidad". COLEMAN optó por dar a "educación" el significado de "educación académica" y, entre cinco concepciones de "oportunidad", optó por "igualdad de resultados", tal como los mide un test de vocabulario.

El estudio suponía una única medida de utilidad. El análisis adaptó la forma de regresión múltiple, refundiendo las variables de origen social y escolar en relación con las puntuaciones obtenidas en el test. A partir de esto, COLEMAN llegó a la conclusión de que las escuelas no corregían las desigualdades que reflejaban las puntuaciones obtenidas en el test. Podía presumirse que lo que necesitaban los negros para acceder a un plano de igualdad eran las destrezas básicas reflejadas en las puntuaciones del test. Se estimó que el problema primordial era el aislamiento socioeconómico y racial, considerándose que la respuesta sería la integración. La táctica sería llevarlos en autobús*.

Tal como lo analiza ENNIS (1975), gran parte de ello estaba implícito en la carta de transmisión del encargo, que identificaba el aislamiento racial como el principal impedimento en contra de la igualdad de oportunidades educativas y en la concepción de igualdad que compartían COLEMAN y las autoridades gubernamentales. Yo añadiría que la conclusión también estaba implícita en la metodología del estudio y en la teoría de la justicia en cuyo ámbito se formuló. ENNIS (1973) señalaba que una cadena causal de razonamiento sólo es empírica parcialmente; la otra porción es axiológica. Los valores participan en la determinación del punto en el que es posible interferir la cadena, estableciendo el carácter diferencial. Los liberales podían aceptar el traslado en autobús como respuesta al problema, rechazando, sin embargo, la interferencia familiar, mientras que los conservadores se opondrían a ambas posibilidades. Dice ENNIS: "Lo que hizo, en apariencia, fue seleccionar los factores que estudiar, cambios que pensaba 1) no violarían

* La táctica del *"busing"* se utilizó para llevar en autobús escolar a los niños negros a escuelas de barrios blancos y a niños blancos a escuelas de barrios negros, con la pretensión de conseguir la integración racial, estructurándose un medio escolar de integración cerrada, independiente del ambiente familiar, perteneciente a medios socioeconómicos distintos del escolar, segregados y, por tanto, también cerrados. *(N. del T.)*

© Ediciones Morata, S. L.

ningún conjunto de valores (al menos, los suyos, en la mayoría de los casos) y 2) podrían indicar una diferencia. Después, llevó a cabo un estudio empírico de estos factores, tratando de descubrir cuáles establecían la diferencia".

Mediante este enfoque, los investigadores determinaron la medida de utilidad y los factores que les parecieron aceptables para provocar una interferencia que influyera en aquélla. La parte empírica del estudio se dedicó a determinar qué factores de los predeterminados afectarían a los fines también establecidos de antemano. Incluso el análisis se sometió a juicios de valor al determinar qué variables entrarían en primer lugar en el proceso de regresión múltiple. Al incluir en primer término los factores debidos al medio de origen, la covarianza atribuible a la interacción entre ese medio y la escuela no podría atribuirse a ésta. Además, el error residual podría atribuirse a cualquier otra cosa, p. ej., a la "suerte", en el análisis de JENCK (LEVIN, 1972). No creo que el estudio de COLEMAN estuviese mal hecho, pero, aunque era en gran medida cuantitativo, estaba *necesariamente* plagado de supuestos ocultos de valor.

El enfoque adoptado por COLEMAN eliminó otras diversas opciones políticas que podrían haberse adoptado. Dejó completamente de lado las actitudes —si no los intereses— de los grupos más preocupados por conseguir una medida general de utilidad. En este caso, la eliminación de la segregación condujo a llevar a los niños negros en autobús a escuelas de grupos blancos no anglosajones, cuyo nivel social y económico era sólo algo superior al de los negros y a transportar a niños blancos a las escuelas del *ghetto*. Esta situación se convirtió en una amenaza a la ya precaria posición de los blancos no anglosajones. Más tarde, COLEMAN (1975) ha cuestionado la táctica del transporte en autobús, comentando sobre su estudio original: "Es evidente que, para lograr este propósito, se debería haber prestado mucha más atención a las reacciones de los blancos no anglosajones con medios económicos suficientes para trasladarse". En el estudio no se tuvo en cuenta el punto de vista de estas personas.

Sin embargo, había una omisión aún mayor: tampoco se consideró el punto de vista de los negros, que eran las personas a quienes se suponía beneficiarían las políticas de integración. Es posible que los puntos de vista de los negros —o los de los blancos no anglosajones— hubieran sugerido otras políticas alternativas en relación con la justicia social. No sólo las políticas gubernamentales prescindieron de la influencia de los más afectados, sino que el estudio de evaluación no reflejó sus puntos de vista e intereses. A pesar de las nobles intenciones de los reponsables de las decisiones, el enfoque estuvo muy lastrado por el paternalismo.

En este caso, quienes pagaron el mayor precio de la reforma social fueron los negros y los blancos no anglosajones (los dos grupos menos favorecidos en este ambiente). En cierto sentido, se procuró beneficiar a los negros a costa de sus antagonistas más naturales y que más tenían que perder. Los grupos sociales que formularon y pusieron en práctica esta política estaban básicamente desligados de los afectados por ella.

© Ediciones Morata, S. L.

Scanned by CamScanner

Quedaba otra injusticia adicional. Dentro del sistema educativo, los costes de la integración recayeron más sobre los menos favorecidos: los niños y, sobre todo, los niños negros que quedaron a merced de las fechorías de los racistas. Incluso en los términos en los que se planteó, quedó demostrado que el estudio original era inadecuado, pues, en donde se llevó a cabo la integración, las puntuaciones en los tests no se elevaron necesariamente (COLEMAN, 1975). Parece ser que se excluyeron del análisis causal factores importantes. Por último, y sobre todo, el enfoque mantuvo el control fuera del alcance de las bases.

Un enfoque pluralista/intuicionista del problema habría recogido los puntos de vista de todos los implicados. No habría establecido unos fines predeterminados ni eliminado posibles causas alternativas sin examen previo. Además, un enfoque de justicia como equidad se aseguraría de que los puntos de vista de los menos favorecidos tuvieran prioridad respecto a los demás. Examinado el problema desde la perspectiva de los menos favorecidos, la prioridad pasaría a quienes ocupaban el segundo lugar entre los menos favorecidos: los grupos blancos no anglosajones, en este caso.

La utilización de las teorías de la justicia

Ninguna de estas teorías predominantes de la justicia son completamente satisfactorias como fundamento de la evaluación. El utilitarismo se sitúa muy cerca de la filosofía oficial del gobierno. Siempre que es posible, los funcionarios gubernamentales tratan de reducir todo a medidas monetarias comunes. El análisis de coste/producto constituye el máximo hito de este enfoque, y la preocupación por el dinero es con frecuencia un indicio de que el evaluador adopta esta teoría de la justicia. Hay mucho que decir a favor del utilitarismo. Proporciona un criterio bien fundamentado para adoptar y justificar las decisiones públicas. Su principal debilidad estriba en que tiende a favorecer a las clases superiores en detrimento de las inferiores y conduce a juicios que no siempre concuerdan con la propia sensibilidad moral. El empleo de medidas utilitaristas lleva a menudo a una excesiva simplificación de los resultados, como ocurre en la evaluación *Follow Through.*

La teoría pluralista/intuicionista de la justicia es más coherente con las ideas cotidianas, de sentido común, de la justicia. La mayoría de la gente utiliza principios múltiples para emitir juicios y los equilibra de forma intuitiva. Los principios utilizados varían de una persona a otra y de una a otra situación. Los empleados para juzgar la justicia de las políticas fiscales no tienen por qué ser los mismos que los utilizados para juzgar la justicia de las tablas salariales. Esta situación concuerda con los usos cotidianos y de sentido común, pero, ¿es suficiente para la evaluación de los programas públicos?

¿Nos daríamos por satisfechos sabiendo que un evaluador utiliza un conjunto de principios y otro emplea un conjunto diferente para evaluar un

© Ediciones Morata, S. L.

determinado programa, sobre todo si ese programa fuese nuestro? La auténtica piedra de toque de la justicia es la idea de consistencia, que implica que los casos similares se tratan de forma semejante. Este enfoque parece amenazar la consistencia. En un extremo aparece la amenaza del relativismo (los principios utilizados por uno son tan buenos como los empleados por cualquier otro). En consecuencia, cualquier juicio es tan bueno como el que más. Esto parece que sería el fin de la evaluación. Sobre los enfoques políticamente pluralistas que utilizan los juicios de los grupos involucrados como criterios últimos de evaluación, se cierne la amenaza de esta posibilidad.

En el otro extremo de los enfoques pluralistas/intuicionistas, es difícil que los principios totalmente derivados de los profesionales consigan la plena aceptación del público. La idea que una profesora tenga del trato justo dispensado a los alumnos quizá no coincida con la de los padres. ¿Cuál empleará el evaluador? Pocas personas estarían dispuestas a dejar que quienes dirigen los ferrocarriles tengan el control exclusivo de lo que se haga. Del mismo modo, los enfoques profesionales de la evaluación, en los que unos expertos evalúan a otros, sean profesores, cirujanos o ingenieros, han ido perdiendo credibilidad durante la pasada década. El público ha perdido confianza en los profesionales que se regulan a sí mismos. Da la sensación de que la gente no está dispuesta a dejar en manos de otros los juicios sobre la justicia, que esencialmente son decisiones respecto a la distribución.

En la teoría más moderna sobre la justicia, RAWLS presenta una más igualitaria que el utilitarismo y más absoluta y determinada que el intuicionismo/pluralismo. RAWLS piensa que concuerda mejor con el sentido común. Pero la justicia como equidad tiene sus propias dificultades en cuanto teoría. Se han levantado contra ella muchas críticas. Una de las más corrientes consiste en que la prioridad absoluta que se concede al principio primero sobre el segundo significa, en la práctica, que nunca se aplique el principio de diferencia. Otros dicen que aún el más mínimo beneficio que se pretendiera para el menos favorecido justificaría apropiarse de todo lo que tiene el más favorecido.

Tampoco la justicia como equidad es siempre coherente con el propio pensamiento intuitivo sobre la justicia. Sólo mira al menos favorecido, sin preocuparse de su participación en las decisiones relativas a su suerte, aspecto que considero muy deseable. La crítica lógica más importante para la evaluación de programas públicos consiste en que el bien colectivo no puede derivarse de las pretensiones individuales del modo que pretende RAWLS. El bien individual no equivale al colectivo, idea que desarrollaré en el siguiente apartado.

No obstante, estoy convencido de que la justicia como equidad es, en cuanto teoría de la justicia, superior al utilitarismo, lo que no significa que éste carezca de valor. La justicia como equidad muestra una sensibilidad moral más profunda. Yo prefiero una concepción intuicionista/pluralista de la justicia a ambas, pero sólo aquella en la que se especifiquen de antemano

© Ediciones Morata, S. L.

Scanned by CamScanner

determinados valores. En el siguiente apartado defenderé que una teoría adecuada de la justicia ha de tener en cuenta los valores de igualdad moral, autonomía moral, imparcialidad y reciprocidad, así como el principio sumativo de utilidad.

Este juicio no constituye una denuncia universal de todo el pensamiento utilitarista en la evaluación. En muchos casos, este pensamiento es adecuado sin que nos lleve por mal camino. Por ejemplo, la evaluación de un plan destinado a lograr la mejor utilización de las camas hospitalarias puede girar en torno a la idea de eficacia, con exclusión de la mayoría de las consideraciones adicionales. El pensamiento utilitarista comienza a ser inadecuado cuando entran en juego otras consideraciones apremiantes.

Por regla general, como indicaba al comienzo de este capítulo, la teoría de la justicia que podamos mantener no determina el tipo de evaluación que efectuemos. Hay otros muchos factores que influyen en ella. La teoría propia actúa de manera subrepticia y sutil. Si el evaluador es utilitarista, volverá una y otra vez sobre las condiciones monetarias y de utilidad. Si es pluralista político, reflejará las opiniones de los distintos grupos en relación con el programa. Si es rawlsiano, considerará los derechos de los individuos y los intereses de los menos favorecidos. Si es neomarxista, pondrá de manifiesto las desigualdades sociales y económicas de la situación. A largo plazo, la propia concepción implícita de la justicia resulta significativa.

Sospecho que la concepción que cada persona tenga de la justicia será mucho más variada, idiosincrásica, dependiente de la situación e incoherente que la mayor parte de las grandes teorías aquí expuestas. Sólo muy pocas personas tienen la capacidad de articular una teoría coherente de la justicia; sus nombres figuran en las portadas de los libros. El evaluador no necesita alinearse con ninguna de las grandes teorías (todas las cuales son deficitarias) u organizar la suya propia. El evaluador sí es profesionalmente responsable de utilizar estas teorías como mojones que nos indican dónde estamos y adónde queremos llegar. Sirven como orientaciones para nuestro rumbo.

© Ediciones Morata, S. L.

TERCERA PARTE

Los principios de la evaluación

> *Por tanto, respetar al otro como persona moral supone tratar de comprender sus aspiraciones e intereses desde su punto de vista y presentarle consideraciones que le hagan posible aceptar las limitaciones de su conducta.*
>
> John RAWLS.
> *A Theory of Justice*, 1971, pág. 338.

Scanned by CamScanner

CAPÍTULO VII

La democratización de la evaluación

La política de la elección

Desde el punto de vista histórico, en los países occidentales, el estado liberal ha precedido al gobierno democrático. La sociedad liberal quedó conceptuada como un mercado competitivo en el que el individuo gozaba de libertad para hacer la mejor adquisición para sí mismo en casi cualquier materia. La idea liberal fundamental era que el sistema social se organiza según el principio de la libertad de elección. "La democracia liberal consiste en la política de la elección" (MACPHERSON, 1965, pág. 33).

> La esencia del liberalismo... es la visión de una sociedad constituida por unidades independientes y autónomas que sólo cooperan cuando las bases de colaboración favorecen los fines de las partes. Las relaciones de mercado constituyen el paradigma de esa cooperación, recogido con precisión en la idea de que el paso del feudalismo a la era liberal de mediados del siglo XIX fue el cambio "de la categoría social al contrato" y que los desarrollos posteriores invirtieron el proceso una vez más. El contrato constituye incluso el modelo de relaciones poco prometedoras como las políticas, en las que las leyes benefician a unos a expensas de otros. Se dice que el sistema en su conjunto beneficia a todos, de manera que todos estarán de antemano de acuerdo con su existencia.
>
> (BARRY, 1973, pág. 166.)

En realidad, el liberalismo fue liberador. Desde el siglo XVII hasta el XIX liberó a las personas de las costumbres y la autoridad. Sustituyó las relaciones basadas en la categoría social por las del contrato impersonal y de mercado. Aunque el proceso liberador trajo, como subproducto, grandes desigualdades de riqueza, éstas se consideraban como un paso inevitable y razonable. Las desigualdades no eran nuevas.

En estas sociedades liberales, incluso el gobierno tuvo que hacerse

© Ediciones Morata, S. L.

Scanned by CamScanner

más responsable al quedar en una situación de mercado. Se le concebía como proveedor de bienes políticos de entre los cuales los consumidores seleccionaban sus preferencias. Diversos partidos políticos se hicieron responsables ante sus votantes mediante el procedimiento de las elecciones. De este modo, el gobierno tuvo que responder a los intereses del electorado, aunque éste no era en absoluto democrático. Al principio, estaba constituido por una reducida elite de varones ricos y de clase alta. La opción de gobierno sólo se ofrecía a esta elite y el gobierno respondía a los intereses de la misma.

La democracia surgió con vacilaciones. A través de los mecanismos de libre expresión y reunión, las clases inferiores exigieron el derecho al voto. Sobre la base de la igualdad de derechos individuales y de oportunidades, reclamaron derechos y libertades semejantes a las alcanzadas por las clases superiores. Dentro de la lógica del liberalismo, les parecía injusto carecer de opción de gobierno y, por tanto, de representación de sus propios intereses. Una vez franqueadas las barreras, las clases inferiores no utilizaron el voto para derrocar a las clases superiores, sino para adoptar una postura competitiva en beneficio de sus propios intereses.

La forma liberal de democracia difería notablemente de las no liberales surgidas más tarde en los países comunistas y en vías de desarrollo. Tanto unos como otros se oponían a la idea de la sociedad de mercado. Las sociedades comunistas consideraban la democracia como un tipo de gobierno por o para las personas corrientes. El proletariado fijaría las reglas hasta llegar a una sociedad sin clases. En tal marco, no cabían las libertades liberales.

En cambio, las sociedades en vías de desarrollo se oponían a menudo tanto a la idea de mercado como a la de clases, considerando la democracia como el gobierno según la "voluntad general" de todas las personas de un país. Con frecuencia, un único partido político, en vez del triunfador en unas elecciones, dominaba el gobierno para llevar a efecto la voluntad general, sin que fuese posible la competición entre partidos políticos en una sociedad tradicional y no mercantil. El significado de la elección, en estos tres tipos de sociedades políticas, es diferente.

Así, la igualdad era una idea fundamental pero, en las democracias liberales, se combinaba con las ideas de elección, competición y sociedad de mercado. Aunque las ideas liberales se democratizaron más, ampliando su vigencia a grupos diversos y extendiendo el ámbito de las decisiones públicas, el principio de elección siguió siendo fundamental. Propongo, como fundamento orientador de las evaluaciones en estas democracias liberales, el principio de *igualdad de elección*, premisa que combina la igualdad y la elección.

Aunque coherente con las dos ideas fundamentales de la democracia liberal, la igualdad de elección difiere significativamente de la pura postura liberal de la que se deriva, de manera tácita, una gran parte de las políticas y evaluaciones actuales. El principio liberal estricto de elección exigiría que se favoreciese al máximo la elección con independencia de su

© Ediciones Morata, S. L.

distribución. El principio democrático liberal de igualdad de elección no sólo requiere que se maximice la elección, sino que se distribuya por igual. La democracia liberal pasa a ser no sólo la política de favorecer al máximo la elección, sino también la política de distribuirla.

Evaluación democrática liberal

El concepto de un procedimiento de evaluación formal y público para ayudar a elegir políticas y programas públicos se deriva en sí de la idea liberal de elección. En la medida en que hace referencia a las clases inferiores o a todos los ciudadanos, también es democrática. Con frecuencia, la elección de programas y políticas se hace sobre la base de las preferencias e intereses privados. Nada en la evaluación liberal (que favorece la elección al máximo) impide que se utilice para fines privados.

Tan fácil es evaluar para un rey como para el público y no hay nada en las ideas liberales que lo impida. En realidad, habría que estimular la evaluación en beneficio de los intereses privados. En su forma más extrema, el evaluador actuaría en beneficio de los intereses de quien pagara la evaluación. La liberal se juzgaría según la utilidad que tuviera para sus destinatarios, con independencia de quiénes fueran. Pero la evaluación de un programa público transforma la opción en una *decisión pública* deliberada. El servicio a los intereses privados no puede constituir el criterio último de la evaluación. La utilidad no puede ser el valor primordial. La evaluación democrática liberal tiene que considerar el interés social y público por encima de los intereses privados de los individuos.

Los principales enfoques modernos de la evaluación dan por supuesta la libertad de elección. También presuponen una metodología individualista, una orientación empírica y un mercado de ideas libre, en el que los consumidores "comprarán" las mejores. Todos son liberales en la medida en que se basan en el concepto de la sociedad de mercado, competitiva e individualista. Sin embargo, difieren considerablemente respecto a las elecciones que se hagan, a quién se elija y al fundamento en el que se base la elección. En otras palabras, difieren en cuanto a sus tendencias democráticas.

En la práctica, la evaluación puede democratizarse aún más ampliando las opciones evaluativas a todos los grupos y extendiendo la evaluación pública a todas las opciones públicas. Esto puede lograrse expandiendo el tipo de datos que se recojan, centrando la evaluación en los niveles más altos de decisión, ampliando el ámbito de destinatarios y grupos de referencia, y extendiendo las opciones hasta incluir el mismo método de evaluación. Estos cambios estarían orientados a ampliar el ámbito de aplicación del principio de igualdad de elección en la evaluación, del mismo modo que la igualdad de oportunidades ha ampliado su ámbito en la sociedad en general.

La idea fundamental de igualdad consiste en tomar a cada uno (o al grupo designado) como único grupo de referencia. Si las personas reci-

© Ediciones Morata, S. L.

Scanned by CamScanner

ben un tratamiento diferente, hemos de justificar esto con principios o razones de peso (BARRY, 1965). La evaluación se desarrolla necesariamente a partir de un punto de vista que incluye algún grupo concreto de referencia. Este grupo está constituido por el conjunto de personas que el evaluador tiene en cuenta al realizar su evaluación. De modo más preciso, el grupo de referencia es el conjunto de personas cuyos intereses no pueden pasarse por alto.

El grupo de referencia debe identificarse con anterioridad a la evaluación. Puede o no coincidir con los destinatarios de la misma. Por ejemplo, el evaluador puede dirigir la evaluación a un responsable clave de las decisiones gubernamentales, aunque el grupo de referencia —aquél cuyos intereses se consideran— no sea independiente. Éste no tiene por qué tener noticia del programa de evaluación llevado a cabo en su beneficio.

Dicho grupo pueden ser muy pequeño —uno mismo o la propia familia—, como cuando compramos un coche, o muy grande —los consumidores y usuarios—, como en algunas evaluaciones de programas públicos. Aunque la idea fundamental de igualdad llevaría a tomar a cada persona como grupo de referencia, esto plantearía un grave problema en cada evaluación. Creo que el grupo de referencia adecuado para una evaluación debe estar constituido por los afectados por el programa o política en cuestión. En el terreno práctico, esto reduce el ámbito sometido a consideración y permite prestar especial atención a los grupos afectados especialmente por un programa o política.

Intereses

La evaluación moderna no constituye de por sí un procedimiento de decisión social, sino que forma parte de uno para la asignación de recursos. Prevé algún tipo de situación en la que se toman decisiones sociales. En este proceso, el papel de la evaluación se limita a la persuasión racional sobre la base de principios y valores comunes.

La evaluación lleva consigo la manipulación de hechos y argumentos con el fin de valorar o determinar la conveniencia de algo. Un conjunto de principios o valores sirve de fundamento para juzgar. Los criterios usados pueden limitarse a uno, como el utilitarista de elevar al máximo la utilidad, o una mezcla de criterios sin prioridades determinadas de antemano, como en la postura pluralista.

Al hacer juicios políticos en una democracia liberal, hay dos tipos de principios o consideraciones: premisas relacionadas con la pretensión y premisas relativas a los ideales. Las primeras dan por sentados los deseos o pretensiones de las personas e indican cómo se satisfarán o se distribuirá la satisfacción de los mismos, sin juzgarlos en sí. Por otra parte, las premisas relativas a los ideales especifican que algunas pretensiones son mejores que otras y la acción pública debe estimularlas. Por ejemplo, la pretensión a favor de las artes puede considerarse más importante que la relativa a los

© Ediciones Morata, S. L.

deportes, por lo que deberá proporcionarse apoyo público a la primera (BARRY, 1965).

Una característica del liberalismo es que las decisiones públicas se toman sobre todo teniendo en cuenta las premisas referidas a las pretensiones. A efectos públicos, éstas se toman tal como se plantean, sin juzgar si una es más conveniente que otra. En el plano público, sólo se hacen juicios respecto a las pretensiones. Por supuesto, los individuos tienen plena libertad para actuar en sus vidas privadas según sus ideales.

El liberalismo sólo considera legítimas las pretensiones que surgen fuera del ámbito de influencia del Estado. Como no puede considerarse que nadie tenga mejor gusto ni juicio que cualquier otro, la opinión de uno vale tanto como la de cualquier otro a efectos políticos. El único criterio de bondad se convierte en la satisfacción de los deseos (BARRY, 1965). En cambio, las posturas no liberales juzgan que determinadas pretensiones son más importantes y dignas del apoyo público que otras. Un perfeccionista o un marxista cree que hay que dar cumplimiento a ciertas pretensiones humanas pero no a otras. Los ideales para juzgar se introducen desde el primer momento. A veces, se cuestiona la misma autenticidad de las pretensiones expresadas, lo que indica la posibilidad de la evaluación no liberal.

En la tradición liberal, sin embargo, las pretensiones se toman tal como se expresan y la cuestión se plantea en términos de agregar en bruto la satisfacción de la pretensión o distribuirla. Los principios de agregación en bruto se aplican a la consecución de la máxima satisfacción de las pretensiones, cuya idea clave consiste en el logro de los propios "intereses". Los principios distributivos suponen conceptos como el de justicia, equidad, igualdad y libertad.

El principio básico de agregación es el de la máxima satisfacción de las pretensiones. Decir de alguien que está incluido en el grupo de referencia de una evaluación es decir que en ella se tienen en cuenta sus "intereses". Un programa o política forma parte de los intereses de una persona cuando aumenta su oportunidad de conseguir lo que pretende. En otras palabras, los "intereses" representan medios generalizados para cualquier fin que pueda plantearse una persona. Es posible que podamos proteger o incrementar los intereses de una persona, p. ej., su riqueza o poder, sin conocer en qué consisten sus fines últimos (BARRY, 1965).

Según esto, las personas también pueden equivocarse respecto a sus propios intereses. Pueden querer un programa o política que no produzca, en realidad, el resultado previsto, situación en la que la evaluación puede resultar especialmente útil; o pueden elegir deliberadamente un programa o política opuesto a sus propios intereses. En general, el concepto de "intereses" sirve de orientación útil respecto a la magnitud y distribución de la satisfacción de las pretensiones y, por tanto, constituye un índice utilizable en la evaluación. Con frecuencia, se emplea implícitamente de este modo.

El concepto de "intereses" también es comparativo. Un programa o política sólo forma parte de los intereses de alguien cuando se compara con otro programa o política o con el *status quo*. Las personas pueden

© Ediciones Morata, S. L.

Scanned by CamScanner

estar de acuerdo con los resultados de determinado programa o política y no respecto a que forme parte de los intereses de un grupo concreto. Como las personas pueden comparar deliberada o inconscientemente el programa con otros competidores, la norma o clase de comparación cambia y los juicios últimos sobre el éxito del programa pueden variar.

Gran parte de las disputas en el campo de la evaluación versa sobre la clase de comparación respecto al programa o política, y no sobre los resultados empíricos concretos. (Véase en el Apéndice A la disputa entre GLASS y SCRIVEN sobre la utilidad de ciertas grabaciones magnetofónicas educativas.) El que un programa o política forme parte de los intereses del grupo de referencia depende de las opciones de una comparación, elección que, con frecuencia, encubren las discusiones sobre la metodología.

Muchos téoricos importantes de la evaluación presuponen implícitamente que el propósito de la evaluación consiste en ayudar a los responsables de las decisiones o a los mismos grupos de referencia a determinar los intereses pertinentes de los diversos grupos (CRONBACH, 1979). Para que la evaluación pueda hacerlo, es preciso que los intereses del grupo de referencia estén representados de alguna manera en la evaluación. El principio general consiste en que en ella deben estar incluidos los intereses de todos los afectados por el programa o política, aunque la forma de hacerlo sea una cuestión abierta.

En una evaluación liberal pura, las pretensiones se dan por sentadas, contando cada una como candidata a la satisfacción. No se presta atención a las derivaciones que de estas pretensiones hagan determinados grupos o clases sociales. Así, tenemos el aparato de las "evaluaciones de necesidades", que suelen ser conjuntos de pretensiones. En la medida en que requieren que las personas juzguen directamente los programas, exigen juicios inmediatos de intereses. Estas encuestas presentan casi siempre sesgos de clase, a causa del muestreo de quienes responden y de su metodología. Solicitan información sobre la base de instrumentos que contienen categorías propias de profesionales de clase media (HOUSE, 1973).

En cambio, las evaluaciones no liberales se basan en juicios respecto a determinadas pretensiones e intereses. La desviación más común es la de los profesionales que albergan fuertes sentimientos en relación con lo que debería ser una persona educada en sentido ideal. A veces, este ideal entra en conflicto con las pretensiones manifestadas públicamente sobre la educación. Muchos de los juicios de profesionales relacionados con los ideales pueden incorporarse a la evaluación; o el evaluador puede guiarse de forma exclusiva por la idea de los sentimientos públicos. Por supuesto, en muchas cuestiones el juicio de profesionales puede ser superior al del público.

Otra posibilidad en relación con los ideales consiste en promover los intereses de determinadas clases, como, por ejemplo, las inferiores. Una evaluación de este tipo se basaría en una concepción ideal explícita de lo que debería ser la sociedad y en la evaluación se integrarían valores y criterios derivados de este especial punto de vista sobre la sociedad. De

© Ediciones Morata, S. L.

antemano, se considerarían importantes ciertos aspectos. Los enfoques no liberales, relacionados con ideales, no tienen por qué promover la elección.

El punto de vista al que hemos dado preferencia aquí es el de la igualdad de elección. Las personas deben tener la oportunidad de elegir, de manera que no les determinen desde fuera las opciones, aunque favorezcan sus propios intereses, pero las elecciones deberían distribuirse de modo que todos los grupos y clases sociales tuvieran las mismas oportunidades para elegir. Habría que facilitar a los grupos sociales inferiores la oportunidad de determinar las opciones que más les favorezcan.

El interés público en la evaluación

Cuando nos ocupamos de los intereses de una persona o de un grupo reducido, éstos son "privados". Aunque nada se opone a que la evaluación sólo represente intereses privados, es difícil que la evaluación de programas y políticas públicos pueda restringirse de ese modo.

Casi siempre, los programas y políticas públicos tienen que ver con los intereses de un gran grupo de personas, como el de los minusválidos, los menos favorecidos o los superdotados. A nivel local, puede tratarse de los de una determinada escuela o de un grupo concreto de alumnos de una ciudad. Son "intereses especiales". Aunque muchos programas gubernamentales se ocupen de grupos de interés especial, son más los programas o políticas que afectan a los intereses de dos o más grupos a la vez. Un programa para los minusválidos no sólo afecta a los estudiantes, sino también a sus padres, profesores y patronos. Podemos decir que estos grupos tienen "intereses comunes". En una democracia liberal, la mayor parte de los programas gubernamentales gozan del apoyo de grupos de intereses especiales, pero son el resultado de la coalición de intereses comunes. Los evaluadores suelen tener que ocuparse de un programa o política que refleja diversos intereses.

La evaluación en sí puede concebirse como una política resultante de un interés común. Las partes implicadas en ella están de acuerdo en que todas saldrán ganando con la evaluación ya que determinará si el programa en cuestión cumple ciertos criterios. Aunque los resultados de la evaluación no satisfagan a una de las partes y sea posible que, en realidad, dañen sus intereses, aún podemos decir que en su conjunto sirve al bien común.

Por último, tenemos el "interés público". Se trata del que tienen en común las personas, en cuanto miembros del público, entendiendo por "público" una cantidad indefinida de individuos "no identificables" y no unas personas concretas (BARRY, 1965). El interés público puede o no ser más fuerte o digno de atención que el privado, especial o común, pero se comparte de forma diferente.

Supongamos dos intereses especiales, p. ej., los fabricantes de automóviles y los sindicatos llegan, en beneficio mutuo, a un acuerdo que exige la subida de precio de los coches. ¿Qué ocurre con los intereses de las

© Ediciones Morata, S. L.

Scanned by CamScanner

personas que no participan del contrato, o sea, los consumidores? El interés público abarca los intereses de estos miembros no identificados del público a quienes afectará aquella iniciativa. En este sentido, la pertenencia al público no se establece de antemano, sino que varía según la cuestión y el contexto. Con frecuencia, se identifica el público con los consumidores, quienes, por regla general, quieren más y mejores bienes a menor precio. Sin embargo, no es difícil imaginar problemas en los cuales los consumidores puedan entrar en conflicto con otros segmentos del público, como los ecologistas.

Siendo así, ¿cómo podemos definir el interés público? La solución no es tan sencilla como cuando un individuo sopesa y decide su propio interés, pues están involucrados individuos diferentes. En este caso, el interés público se define a menudo como la suma de todos los intereses involucrados, lo que equivale a conseguir el máximo de satisfacción sin contar con la distribución de la misma. Una alternativa consiste en contemplar sólo los intereses que todas las personas tienen en común (cuando existen intereses compartidos) y dejar de lado los divergentes.

Los distintos enfoques de la evaluación reflejan de diversos modos los intereses. Ya antes clasifiqué los ocho principales enfoques de la evaluación en dos grupos: los utilitaristas y los pluralistas/intuicionistas. Los primeros tratan de llegar a un juicio general sobre la utilidad social, basado en un único principio o criterio dominante, que es el de agregación de la máxima satisfacción bruta de las pretensiones.

El grupo utilitarista se subdivide en un subgrupo *de gestión,* que considera a los directivos como sus primordiales destinatarios, grupo de referencia o ambas cosas, y un subgrupo que considera a los consumidores y usuarios como destinatarios, grupo de referencia o ambos. En el subgrupo de gestión, el enfoque del análisis de sistemas establece indicadores sociales, como las puntuaciones normalizadas de los tests, en calidad de medidas vicarias de utilidad social, que se equiparan al interés público. Se supone que la satisfacción de los intereses de cualquier grupo social aumenta cuando se incrementan sus puntuaciones. Cuando los intereses se oponen, se suman todos ellos. Es posible que la mayor elevación de las puntuaciones en los tests maximice la satisfacción de las pretensiones. Es más, el interés público se estructura en torno al máximo incremento en las puntuaciones por la mínima cantidad de dinero. El "mejor" programa será el que consiga esto mismo.

El enfoque por objetivos conductuales representa los intereses en su definición de objetivos. Si éstos se cumplen al máximo, o se logran los mínimos establecidos, se satisfará el interés público. El evaluador mide los objetivos. En el enfoque de decisión, se presume que la postura oficial del responsable de la decisión representa el interés público. El cometido del evaluador consiste en proporcionar al destinatario información que le permita perfeccionar sus decisiones. Todos estos enfoques identifican el interés público, ya se traduzca en índices sociales, en objetivos formulados de manera pública o en decisiones de funcionarios públicos, con la estructura oficial de gestión.

© Ediciones Morata, S. L.

Las evaluaciones basadas en estos enfoques difieren entre sí según tomen a los directivos como destinatarios o también como grupo de referencia. Si los directivos constituyen el grupo de referencia, la evaluación irá en beneficio de los intereses de éstos más que del público. Por desgracia, esto se repite con demasiada frecuencia. Las evaluaciones se llevan a cabo para satisfacer intereses de gestión (especiales), lo que resulta inaceptable aún en el marco de la teoría de estos enfoques.

El subgrupo de usuarios considera que el público está dividido entre productores y consumidores o usuarios. Sus destinatarios pueden ser tanto directivos como usuarios, pero el grupo de referencia estará constituido por éstos. El evaluador refleja los intereses del consumidor. El paradigma de esta postura es la *Consumers' Union.* De nuevo, el interés público se identifica a menudo con el mejor producto al mínimo coste.

El grupo pluralista/intuicionista evalúa sobre la base de muy diversos principios, sin especificar las prioridades entre los mismos. Se divide en un subgrupo profesional y otro participativo. Los profesionales creen que quienes más conocimientos tienen y mejor informados están sobre un campo concreto son los que más pueden decir al respecto. El interés público se satisfará mejor si deciden los expertos. En cierto sentido, el juicio de decisión se delega en el saber superior de ese grupo.

En la medida en que los subgrupos profesionales consideran que ellos hacen avanzar los medios para conseguir fines públicamente definidos, no tiene por qué producirse un conflicto. Pero, por supuesto, los profesionales poseen sus propios ideales muy caracterizados acerca de cómo debe ser una persona educada. Como las normas de evaluación son con frecuencia profesionales, esas normas pueden diferir de las públicas. En este sentido, las evaluaciones profesionales pueden desviarse, para bien o para mal, de los principios estrictos de satisfacción de pretensiones, al emplear las consideraciones de los profesionales sobre los ideales. Además, es difícil desligar estas normas de los propios intereses profesionales. No hace mucho, se han producido diversos intentos, desde los programas de tests de competencia hasta los tribunales médicos de revisión, para reafirmar las demandas públicas en estas evaluaciones.

Por último, el grupo participativo cree que la mejor manera de contribuir al interés público consiste en que las personas participen en algún grado en la evaluación. Los enfoques cuasijudiciales solicitan las opiniones de quienes están involucrados en el programa, incorporándolas, incluso al pie de la letra, al informe. Los enfoques cuasijudiciales facilitan la participación haciendo que las partes afectadas presenten pruebas y argumentos en procedimientos cuasijudiciales. Estos enfoques no son pluralistas en la medida en que se utilizan diversos criterios para juzgar, pero sí lo son políticamente en tanto en cuanto en la evaluación están representados distintos intereses.

© Ediciones Morata, S. L.

Scanned by CamScanner

Concentración de poder frente a difusión de poder

Es instructivo comparar dos enfoques de evaluación que son extremadamente distintos en cuanto a la definición del interés público; el primero, de concentración de poder, y el segundo, de difusión del mismo. El paradigma del enfoque de la concentración de poder es la política de evaluación de la *U.S. Office of Education* durante la pasada década (McLaughlin, 1975; House, 1978b). Implica que el gobierno central puede definir mejor el interés público. Los representantes son elegidos y ellos nombran una burocracia. Ésta define la política y toma determinaciones autorizadas. Puede trabajar con un grupo de expertos que ayude en esta tarea (Barry, 1965).

En realidad, el gobierno federal actúa a menudo como si decidiera qué programas sociales son mejores, utilizando una variante del enfoque del análisis de sistemas para evaluar programas públicos. Los indicadores sociales —en educación, puntuaciones obtenidas en tests normalizados, casi siempre— se utilizan como índice del interés público. El poder y las decisiones se concentran en los funcionarios, evaluadores e índices del gobierno central.

Por otra parte, la idea de la difusión de poder se remonta, al menos, hasta Hume, que decía: "debe suponerse que todo hombre es un *bellaco,* sin otro fin, en todas sus acciones, que el interés privado." Según esta tesis, a nadie debe confiarse tal poder. Debe haber comprobaciones y equilibrios, tal como aparece en la Constitución de los Estados Unidos. En educación, la mejor representación de la tesis de la difusión de poder quizá sea la "evaluación democrática", de MacDonald, el enfoque más explícitamente democrático de los participativos.

En la evaluación "democrática", el evaluador recoge citas de los participantes en el programa y, en su informe, redactado como estudio de casos, refleja con fidelidad sus puntos de vista. El evaluador es un "intermediario" en los intercambios de información accesible a los no especialistas. Representa un conjunto de intereses y no es posible que haga un mal uso de la información (MacDonald, 1974). El evaluador no hace recomendaciones, sino que presenta la información a los destinatarios para que la utilicen como crean conveniente. En teoría, el evaluador presenta el informe de evaluación a las personas de las que ha recabado los datos, permitiéndoles vetar la información que no deseen se incluya.

La evaluación de MacDonald parece tener presente una situación de decisión similar a la democracia directa, en la que los ciudadanos discuten cara a cara los problemas y deciden sobre ellos. A su vez, esto se basa en el punto de vista clásico del liberalismo, en el que los individuos se asocian directamente para conseguir sus fines individuales, sin recurrir a las instituciones o al gobierno. Se facilita al máximo la elección del individuo. El gobierno y, en este caso, el evaluador, sólo es árbitro.

La comparación de esta visión de la evaluación con el enfoque del gobierno federal, sobre una base política y moral, resulta favorable a la primera. Yo he criticado con severidad la política federal de evaluación por

© Ediciones Morata, S. L.

no incluir intereses divergentes, por su tratamiento injusto y por deficiencias técnicas en su ejecución (House y cols., 1978). Incluso desarrollada adecuadamente, la suposición de que los indicadores sociales representan en realidad el interés público es dudosa. Es más frecuente que las evaluaciones representen intereses especiales. Tampoco las acciones del gobierno, unilaterales y a menudo arbitrarias, refuerzan la calidad moral de las evaluaciones.

El enfoque de evaluación de MacDonald incluye de forma intencionada intereses diversos, permite que las personas reflejen los suyos propios y se basa en la idea del consentimiento mutuo. El concepto de democracia de MacDonald se acerca mucho al de gobierno con el "consentimiento del gobernado" o, en este caso, del evaluado. Por supuesto, hay otros conceptos de democracia, sobre todo los que tienen que ver con la distribución de bienes entre clases sociales. En comparación con la política federal, con su forma de imponer acciones sin autorización ni consulta, un enfoque basado en el consentimiento tiene considerable atractivo.

Por otra parte, un enfoque de difusión de poder (porque la necesidad del consentimiento difunde el poder) no carece de problemas. Es bastante dudoso que la democracia directa sea posible en una sociedad industrializada y de masas, compuesta por grupos fragmentados. Las decisiones suelen tomarse a nivel central o de grupo, lo que puede requerir información diferente. Por supuesto, puede afirmarse que las decisiones no *deberían* tomarse en esos niveles, que se traducen en malos resultados de las mismas. También podemos plantear si solicitar el consentimiento de cada persona respecto de los resultados de la evaluación no llevará a un interés común a los participantes en el programa, pero dejando de lado el interés público. ¿Quedaría excluida información valiosa? ¿Presentaría la evaluación un sesgo que favoreciese el *status quo* y no el comienzo de acciones?

En la práctica, garantizar el consentimiento de cada participante plantea unos costes de negociación muy elevados. Efectuar ésta con cada persona puede exigir enormes cantidades de tiempo y energía. De hecho, a causa de esta dificultad, varias evaluaciones de este tipo se han retrasado o no se han terminado (Simons, 1977). Si los costes de negociación son demasiado altos, las personas se niegan a participar y dejan sin examinar el trabajo. La negociación sobre los resultados introduce también la posibilidad de que personas poco escrupulosas hagan mal uso de la información (Elliott, 1977). Una forma de reducir los costes de negociación consiste en llevar a cabo ésta con representantes aunque, por supuesto, éstos vuelven a concentrar el poder.

Una dificultad final de estos procedimientos consiste en el coste de la información (Barry, 1965). Al recoger y transmitir datos desde los grupos hacia los responsables de las decisiones, pueden tener más ventajas los pequeños grupos organizados. Para los grandes, carentes de forma, desorganizados, como el público en general o las clases inferiores, será muy costoso formular sus puntos de vista. Por tanto, aunque se incremente la información total disponible a efectos de tomar decisiones, esa información

© Ediciones Morata, S. L.

Scanned by CamScanner

puede inclinar a favorecer intereses particulares y no a los del público en general. En realidad, la difusión de poder puede facilitar la influencia de intereses especiales (BARRY, 1965).

Es obvio que todos los enfoques de evaluación presentan dificultades. Insisto en estos problemas para demostrar que no existe una panacea, siquiera en una evaluación concebida democráticamente. La evaluación "democrática" de MACDONALD no es lo bastante potente como para reflejar de forma adecuada los intereses de clase social, pero introduce en la evaluación una idea moral extremadamente importante: la del consentimiento mutuo. Éste se manifiesta en la elección. No obstante, yo preferiría que el consentimiento se ejerciera en el acuerdo de evaluación más que en los resultados.

Principios distributivos

Es inevitable que las pretensiones e intereses de las personas se opongan entre sí y es preciso que estos conflictos se resuelvan de algún modo. Mientras los principios de agregación sólo tienen en cuenta la cantidad bruta de satisfacción para un determinado grupo de referencia, los principios distributivos, como la igualdad, la justicia y la equidad, se emplean para juzgar la forma de distribuir la satisfacción de las pretensiones entre los miembros del grupo de referencia (BARRY, 1965). O, expresado de modo más preciso, los principios se utilizan para juzgar los *procedimientos* mediante los que se toman las decisiones de asignación.

En una sociedad liberal, los procedimientos de decisión social para la resolución de los conflictos son críticos. En el liberalismo hay pocas cuestiones esenciales en las que todo el mundo esté de acuerdo, de manera que no puede esperarse que las decisiones se tomen sobre la base de los resultados o de algún modelo ideal de distribución, p. ej., mediante una equidad absoluta. En cambio, deben justificarse sobre la base de los procedimientos utilizados, entre los que se cuenta la evaluación pública. Los principios distributivos suelen aplicarse a los métodos de decisión más que a las decisiones mismas. Concibo la evaluación como un procedimiento de descubrimiento de hechos o "descubrimiento de valores", preparatorio de otro referido a la asignación real. Sólo es un procedimiento de decisión en la medida en que determina públicamente un estado de cosas sobre el que pueden basarse otras decisiones.

La idea fundamental de igualdad especifica que se considere que todas las personas pertenecen a un único grupo de referencia. Suponiendo que se haya definido éste de modo amplio o estricto, ¿cómo han de distribuirse las cosas entre ellas? La igualdad distributiva constituye una cuestión más contundente. La interpretación "fuerte" especifica que un bien ha de dividirse por igual, con independencia de cualesquiera características personales de los individuos del grupo de referencia. La interpretación "débil" consiste en que sólo han de ser iguales las *oportunidades* para satisfacer las pretensiones. El ideal que abrazan las sociedades liberales consiste en la forma débil, a favor de la igualdad de oportunidades más que del reparto efectivo. Por tanto, es más fácil defender que se incluyan en la evaluación los intereses de todos los grupos pertinentes que mantener que aquellos se satisfagan por igual. Sólo se facilita la oportunidad.

© Ediciones Morata, S. L.

La desigualdad de oportunidades surge a partir de la injusticia de los procedimientos o de las condiciones de origen. Cuando no se siguen los métodos adecuados o cuando factores no pertinentes, como la raza o la clase social, influyen en las decisiones, el procedimiento es injusto. Es posible que procedimientos y condiciones de origen justos llevasen a la igualdad de oportunidades.

En las evaluaciones, los grupos no sólo se quejan con razón de que sus intereses no quedan reflejados en ellas, sino también de los sesgos dependientes de instrumentos y análisis inadecuados, etc. La justicia de los procedimientos y de las situaciones de origen cobra extrema importancia en una situación social en la que el individuo tiene la oportunidad de participar en una competición para la satisfacción de pretensiones, en la que hay que luchar frente a otros intereses para ganar.

La aplicación de los principios distributivos depende de los procedimientos concretos de decisión social que se utilicen. BARRY (1965) identifica siete tipos "puros" de procedimientos de decisión social: concurso de méritos, combate, suerte, votación, regateo, competición y determinación de autoridad. Las distintas evaluaciones anticipan tipos diferentes de procedimientos de decisión social.

No es fácil ver cómo puede llevar la evaluación a un procedimiento de decisión como el combate. Podemos imaginar éste como una metáfora de la maniobra y el regateo políticos, pero sería raro que se diese el auténtico combate (la imposición de la voluntad de un contendiente por la fuerza). Por regla general, se convierte en regateo o competición. De igual modo, es raro que se dé la suerte como procedimiento de determinación. Podemos imaginar la votación, pero no suele utilizarse. Quedan el regateo, el concurso de méritos, la competición y la determinación de autoridad.

Tanto el regateo como el concurso de méritos pueden formar parte de las negociaciones; pero, en sentido estricto, el regateo es una situación en la que una parte ofrece un beneficio o desventaja a la otra a cambio de que ésta lleve a cabo alguna acción. Gran parte de las decisiones sociales dependen del regateo, pero ése no es el papel de la evaluación. Podemos imaginar que ésta se utilice como amenaza en una situación de regateo, pero es difícil pensar en algún fundamento moral de la evaluación que utilice la amenaza y los incentivos materiales en su diseño. El papel de la evaluación debe limitarse a la persuasión en el proceso de decisión social. Es una cuestión crítica que marca los límites entre la política y la moralidad, sobre la cual volveremos. Para que la evaluación tenga alguna autoridad moral, no puede concebirse como regateo. Sus resultados no pueden obtenerse mediante amenazas o incentivos.

Todo ello hace que nuestra discusión se centre en: el concurso de

© Ediciones Morata, S. L.

Scanned by CamScanner

méritos, la competición y la determinación de autoridad como procedimientos legítimos de decisión que la evaluación podría anticipar. El concurso de méritos aspira a alcanzar un acuerdo sobre la división de los bienes moralmente correcta. El acuerdo se logra sobre la base de lo que favorece el interés público, lo que produzca la máxima satisfacción de pretensiones, etc. No hay amenazas ni incentivos. Si se llega al acuerdo, las partes habrán cambiado su forma de pensar sobre lo que pretenden. Incluso si una parte tiene poder para modificar las cosas, no debe pretenderlo. En cambio, en el regateo, cada parte trata de conseguir todo lo que pueda en virtud de su poder. Mediante el concurso de méritos, cada parte se convence de que la solución adoptada es la correcta. Parece, pues, que es el procedimiento ideal de decisión social para la evaluación.

A veces, el concurso de méritos dependerá de la cuestión de quién es mejor para algo. Para solucionar el problema, conviene sustituir la cuestión original por una competición, como puede ser una oposición. La citada cuestión original sobre el mérito no queda saldada, sino que se reemplaza por la competición, que puede solventarse de un modo más fácil y, quizá, más objetivo. Las evaluaciones que emplean experimentos comparativos y variaciones planificadas constituyen estas competiciones. Como la resolución depende del rendimiento comparativo, todas las partes deben saber de antemano cuál es el criterio de rendimiento y, por supuesto, para que sea justo, el resultado debe constituir un índice preciso de la calidad que se supone mide la competición (Barry, 1965). Los evaluadores denominan a esto "validez". Por ejemplo, en la evaluación *Follow Through,* planteada como competición, se suscitaban considerables dudas sobre si la competición medía la calidad que se esperaba. En este sentido, era injusta (House y cols., 1978).

Por último, está la determinación de autoridad. Cuando las partes no pueden ponerse de acuerdo, pueden llamar a alguien para que resuelva la situación, alguien a quien todas las partes reconozcan como legítimo juez. Las partes nombran un árbitro. En la determinación de autoridad, el árbitro influye en el resultado basándose en el mérito de los candidatos, mientras que en la competición, los mismos contendientes deciden la cuestión sobre la base de las destrezas y competencias en litigio. El procedimiento de determinación necesita un árbitro que decida el resultado, mientras la competición precisa de un árbitro que compruebe que se respetan las reglas del juego. De nuevo, en *Follow Through,* el gobierno federal actuó primero como legislador, determinando las reglas y como árbitro después, declarando al vencedor. Son papeles incongruentes.

Aunque la mayoría de los evaluadores sostendría que el concurso de méritos, en el que todas las partes llegan a un acuerdo sobre la base de los resultados, constituye el procedimiento ideal de decisión social, es probable que la mayoría de las evaluaciones implique la determinación de autoridad. El evaluador, los funcionarios gubernativos o un grupo de profesionales declara las decisiones. Es evidente que, aun en estas situaciones, el concurso

© Ediciones Morata, S. L.

de méritos suele preceder a tal declaración, aunque sea llevado a cabo por un grupo selecto de participantes.

Las evaluaciones contradictorias, que incorporan un jurado o juez, se estructuran, en realidad, por sí mismas, tras desarrollar procedimientos de determinación de autoridad legal, que obligan, por tanto, a la resolución. Otras evaluaciones, como algunos de los enfoques de estudios de casos, tratan de prolongar y ampliar el concurso de méritos, prohibiendo la determinación de autoridad a cargo de los evaluadores. Otras estimulan al evaluador para que introduzca su interpretación explícita de los hechos en el procedimiento de decisión social. Los procedimientos reales de decisión son combinaciones de estos tipos puros.

Para que se tomen como legítimos y establezcan la igualdad de oportunidades, estos procedimientos de decisión social han de considerarse *justos.* Un método de decisión justo refuerza la igualdad de oportunidades. Una evaluación justa refuerza la igualdad de elección. La equidad es un principio comparativo (confrontando las oportunidades de uno con las de otros), que se aplica a los mismos procedimientos de decisión.

La equidad de procedimiento exige que se cumplan de hecho las formalidades prescritas. En un concurso de méritos equitativo no suelen darse tales formalidades (aunque puedan existir reglas de discusión), salvo la inexistencia de coerción o de incentivos. De no ser así, el procedimiento degenera en regateo. En una competición equitativa deben seguirse las reglas, cualesquiera sean, y la determinación de autoridad debe seguir los procedimientos establecidos para la misma. La equidad de origen constituye un perfeccionamiento de la equidad de procedimiento y exige que las partes involucradas partan de una posición inicial correcta. En la determinación de autoridad está también la consideración de que el árbitro aplique de forma correcta las reglas pertinentes. La aplicación correcta de una regla conduce a la consistencia, requisito básico de la justicia.

Todas estas consideraciones sobre la equidad llevan al resultado "justo" del procedimiento de decisión. En conjunto, constituyen la igualdad de oportunidades. En general, la equidad del procedimiento de decisión es crítica en una sociedad liberal porque las personas se ponen de acuerdo sobre los procedimientos y no sobre los resultados. Es más, cuando alguien está de acuerdo en tal procedimiento y acepta los beneficios que se deriven de él, sólo cabe en justicia seguir aceptando dicho método aun en circunstancias que no resulten personalmente beneficiosas (Rawls, 1971).

La evaluación como procedimiento de decisión moral

La evaluación puede estructurarse como procedimiento de decisión social, aunque me parece más exacto decir que forma parte de una combinación compleja de procedimientos de decisión. Esta combinación varía de un contexto social a otro. La evaluación no suele decidir, en realidad, cuestiones sociales, aunque pueda hacerlo. Con mayor frecuencia suministra datos a

© Ediciones Morata, S. L.

Scanned by CamScanner

otro procedimiento de decisión en el que se lleva a cabo la asignación efectiva de bienes. Es importante cómo ve el evaluador este procedimiento final de decisión.

Como el descubrimiento de hechos y el "descubrimiento de valores" forman parte de una cadena de procedimientos de decisión, la evaluación prevé y adopta características de estos procedimientos. En la medida en que se estructure como un concurso de méritos, una competición o una determinación de autoridad, está sometida a consideraciones semejantes de equidad. La preocupación extrema por la metodología, por el "proceso debido", refleja el carácter crítico de la creencia en los adecuados procedimientos de decisión.

Gran parte de la definición concreta de la metodología "adecuada" se deriva de las comunidades profesionales y técnicas. Mediante sus metodologías, los técnicos tratan de eliminar los "sesgos", es decir, de garantizar la capacidad de reproducción de los resultados, la cual, sin embargo, tiene poco que ver con la garantía de que la evaluación sea democrática o moralmente aceptable.

Los resultados reproducibles pueden ser discutibles desde el punto de vista democrático o moral. La metodología positivista, basada en la ciencia física reestructurada, confunde aquí al evaluador. Es posible que el funcionario gubernativo, previendo una severa oposición a sus decisiones de autoridad, exija al evaluador que emplee sus métodos más "duros" y más científicos, para respaldar la autoridad del gobierno. Pero el evaluador comprometido descubrirá, quizá tarde, que la reproducibilidad es inadecuada en este ámbito.

La teoría democrática sostiene que, para ser aceptable, una política debe reflejar y responder a los intereses de los miembros de la comunidad (CARE, 1978). Los procedimientos democráticos de decisión están diseñados para recoger y organizar esos intereses. Como método de decisión en una sociedad democrática, así debe ser la evaluación. Pero aún esto puede ser inadecuado. Un procedimiento que asignara a cada persona un reparto igual o proporcionara, de forma automática, muchos más recursos a las clases inferiores, ¿podría considerarse equitativo? Creo que no; al menos, no en una sociedad liberal; aunque podemos imaginar una sociedad en la que tales procedimientos fuesen justos.

No todos los procedimientos o políticas que reflejen los intereses de los miembros tienen por qué ser morales (CARE, 1978). Tales procedimientos o políticas pueden ser el resultado de regateos o compromisos en los que se utilizaran incentivos o amenazas. Esto no es moralmente aceptable en una evaluación que pretenda proporcionar el fundamento de un concurso de méritos. Así, pues, ¿cómo habrá que estructurar un procedimiento de evaluación para que sea justo o moralmente aceptable?

De nuevo, la respuesta se orienta a la elección, en que las partes tengan algo que decir en la misma evaluación. La participación en el diseño del procedimiento de evaluación ofrece un modo de establecer la aceptabilidad moral (CARE, 1978). Creo que ésta se relaciona íntimamente con la

© Ediciones Morata, S. L.

autonomía moral y el consentimiento. No podemos imponer la propia voluntad sobre nadie. Entiendo que el acuerdo voluntario sobre un procedimiento de decisión vincula a quien lo acepta, aunque quizá no le gusten los resultados del procedimiento. No obstante, no todo acuerdo es moralmente vinculante.

CARE (1978) presenta la idea de "aceptabilidad moral de los procedimientos", de manera que la participación en un acuerdo haga moralmente aceptables los resultados del mismo. Cuando las personas que participan en el acuerdo alcanzan ciertos niveles y cumplen determinadas condiciones, los resultados son moralmente aceptables, como cuando se siguen las reglas de un juego. De acuerdo con CARE, he planteado las condiciones necesarias para que un acuerdo de evaluación sea equitativo (véase el Capítulo VIII).

En otras palabras, las partes implicadas en una evaluación han de llegar de antemano a un acuerdo sobre lo que constituirá la evaluación, o bien a la comprensión de la misma. Este acuerdo sirve de fundamento para juzgar la justicia de la evaluación, del mismo modo que la aceptación de las reglas de un juego hacen que su desarrollo sea equitativo. Pero no sirve cualquier tipo de acuerdo, pues puede tratarse sólo de un regateo. Con el fin de conferirle aceptabilidad moral, el acuerdo debe alcanzarse en determinadas condiciones que garanticen que los participantes sean capaces de identificar sus intereses reales en la cuestión. Para que un acuerdo de evaluación sea equitativo es preciso que se cumplan doce condiciones.

Una interesante cuestión práctica consiste en si siempre será posible cumplir estas condiciones. En todo caso, constituyen un ideal moral en relación con el cual puede valorarse una evaluación. En la medida en que el acuerdo de evaluación no cumpla estas condiciones, no podrá decirse que sea equitativo ni moralmente vinculante. Por tanto, se hace sospechoso como modelo de equidad de la evaluación misma.

En consecuencia, un acuerdo equitativo de evaluación constituye una forma posible de garantizar la igualdad de elección. El acuerdo puede ser totalmente informal y no escrito, aunque participa de la idea de "contrato". En una sociedad concebida como conjunto de individuos independientes y autónomos, que sólo cooperan en beneficio de sus propios fines —la esencia del liberalismo—, el contrato constituye un medio por el que los individuos se obliguen voluntariamente. El contrato social es la base moral más firme que puede ofrecer el liberalismo.

© Ediciones Morata, S. L.

Scanned by CamScanner

CAPÍTULO VIII

Un contrato de evaluación equitativo

Hace unos años, Daniel STUFFLEBEAM, Wendell RIVERS y yo contratamos con la *Michigan Education Association* y con la *National Education Association* la realización de una evaluación del *Michigan Department of Education Accountability System* (HOUSE y cols., 1974). Como grupo de evaluadores, sabíamos que nuestra tarea sería políticamente muy difícil, con independencia de cuáles fueran nuestras conclusiones. Era fácil que una u otra parte quedase insatisfecha y quizá ambas.

En dos días de negociación con representantes de las asociaciones, tratamos de protegernos lo mejor que pudimos llegando a un acuerdo explícito con ellos. Nos preocupaba especialmente el grado de libertad que tendríamos para desarrollar nuestra evaluación, y también que nuestro trabajo no fuera sometido a la censura de las asociaciones, en el caso de que sus resultados parecieran oponerse a sus intereses.

Cuando discutimos entre nosotros la tarea, STUFFLEBEAM se mostró a favor de establecer un contrato por escrito en el que figurasen nuestras preocupaciones, en vez de dejarlas inespecificadas y ambiguas. A mí no me estusiasmaba redactar un documento en ese momento, dado que pensaba que haría lo que creyera adecuado, con o sin documento. STUFFLEBEAM demostró que, a este respecto, tenía toda la razón. Él redactó el contrato que más tarde firmamos y nos demostró su valor no sólo para definir nuestras relaciones con nuestros patrocinadores, sino también para ponerlas de manifiesto ante terceros. (El Apéndice C es copia del mismo.)

He llegado a la conclusión de que este tipo de contratos es fundamental en una evaluación y que influyen en grado importante en lo que suceda

NOTA DEL AUTOR: Este capítulo se publicó como: Ernest R. HOUSE y Norman S. CARE (1979): "Fair Evaluation Agreement", *Educational Theory*, Vol. 29, n.º 3, verano 1979. Reimpreso con autorización.

© Ediciones Morata, S. L.

durante la evaluación. STAKE (1976) ha ordenado las opiniones de diversos expertos acerca de lo que hay que tener presente en la negociación de estos contratos de evaluación. La cuestión que pretendo tratar en este capítulo es: ¿en qué consiste un contrato de evaluación *equitativo?*, o sea, ¿cuándo un contrato de evaluación es equitativo en relación con las partes afectadas por la evaluación y con los evaluadores mismos?

La caracterización de la equidad

Es obvio que, en sí, la equidad constituye un tema de gran interés para todos los que tienen que ver con prácticas, sistemas e instituciones. En el contexto de esta exposición, hemos de considerar la equidad en relación con las obligaciones que surgen de ciertas expectativas, como las que se derivan de reglas o reglamentos, sean explícitos o implícitos. El *principio de equidad* sostiene que una persona está obligada a cumplir su parte, según queda definida por determinadas reglas institucionales, cuando ha aceptado voluntariamente los beneficios que se deriven de un acuerdo, o ha sacado algún beneficio de las oportunidades que se le ofrecen de promover sus intereses (RAWLS, 1971). La idea básica consiste en que, cuando una serie de personas se comprometen en una tarea que les depara mutuos beneficios de acuerdo con unas reglas y ellas mismas se limitan con el fin de que todas se beneficien, quienes se someten a esas limitaciones tienen derecho a un trato semejante por parte de quienes se han beneficiado. Nadie debe beneficiarse de los esfuerzos de otros sin hacer nada a cambio. Las transgresiones de este principio se consideran *no equitativas.*

RAWLS (1971) plantea dos restricciones respecto a las obligaciones que surgen de la equidad así entendida. En primer lugar, las instituciones de base también tienen que ser *justas.* Aquí no me ocuparé de esta restricción, pero podemos ilustrar su fuerza de este modo: si un esclavo no entrega los bienes apalabrados a su amo, cuando (digámoslo así) han acordado que así se hiciese, no tenemos por qué considerar no equitativo al esclavo respecto al amo, dado que la institución general a cuyo amparo se estableció el acuerdo es claramente *injusta.* Desde este punto de vista, no podemos estar vinculados a instituciones injustas. La segunda restricción, según RAWLS, consiste en que las obligaciones surgen de actos *voluntarios*, como promesas y acuerdos, aunque tales actos voluntarios sean implícitos o "tácitos", como en el caso de la aceptación de los beneficios que se derivan de un contrato.

Las obligaciones en cuestión pueden distinguirse de las "obligaciones naturales" que son independientes de las instituciones y de las reglas sociales. Por ejemplo, tenemos la obligación natural de ayudar a otros que se encuentren en peligro, con independencia de cualesquiera contratos o instituciones antecedentes. Si no proporcionamos esa ayuda, la gente no dirá que no somos equitativos. Quizá "antinaturales" sea un término más adecuado. Las obligaciones naturales tienen preferencia sobre el tipo de obli-

© Ediciones Morata, S. L.

Scanned by CamScanner

gaciones a las que me refiero. A veces, los evaluadores tienen que reorganizar sus conciencias más que sus contratos.

Aunque las obligaciones pueden surgir a partir de actos voluntarios, como promesas y acuerdos, estos últimos pueden ser tácitos. Por ejemplo, un tipo de acuerdo tácito implicaría ciertas expectativas materiales a cambio de la realización del trabajo. Si desarrollamos la tarea teniendo presentes determinadas expectativas y no recibimos la remuneración adecuada, nos sentiremos tratados de manera poco equitativa. Los contenidos de las obligaciones están definidos por las reglas o promesas que especifican lo que tenemos que hacer. Las obligaciones se refieren a individuos concretos que cooperan en el trabajo en cuestión.

En la evaluación de programas educativos hay muchas posibilidades de malentendidos, por lo que conviene concretar al máximo todas las cuestiones en los contratos de evaluación, aunque no sea posible la completa especificación de todos los aspectos. STAKE (1976) señala cinco áreas fundamentales de interés para la negociación: los fines de la evaluación; sus destinatarios; los métodos de investigación, en los que se incluyen los procedimientos de tratamiento de la información; la confidencialidad, y las opciones para continuar las negociaciones. Presenta una lista de comprobación de trece puntos respecto a los temas que se deben tratar en la negociación de un contrato.

No obstante, sigue siendo posible que el contrato de evaluación celebrado no sea equitativo para todas las partes interesadas, con independencia de la mayor o menor organización o especificación del mismo. El contrato puede resultar no equitativo para algún grupo participante en la negociación o afectado por el proceso mismo de evaluación. El hecho de que se llegue a un acuerdo no es, de por sí, condición suficiente de su equidad. Por tanto, podemos preguntarnos: ¿en qué condiciones se garantiza o, al menos, se contribuye a la equidad en los contratos de evaluación?

Algunas condiciones para un contrato equitativo

Norman S. CARE (1978) ha estudiado las condiciones en las que, por regla general, son moralmente aceptables los contratos normativos celebrados entre personas; y su investigación, sin pretender que sea completa, nos presenta un conjunto de doce condiciones que deben cumplirse para que un contrato sea equitativo. Éstas se hallan clasificadas de forma un tanto arbitraria, de manera que algunas se aplican a las personas que llevan a cabo el contrato, otras al contrato mismo y otras a los procedimientos empleados para alcanzar el acuerdo. A continuación explicaré brevemente estas condiciones, haciendo algunas observaciones relativas a su cumplimiento en nuestro contrato de evaluación de Michigan antes mencionado, y haré algunos comentarios en general sobre el cumplimiento de las mismas en el contexto de los contratos de evaluación para la revisión y valoración de programas educativos. De nuevo, pretendo facilitar la mejor comprensión

© Ediciones Morata, S. L.

de lo que significa que un contrato de evaluación posea la dimensión moral de la equidad. Es obvio que en estos contratos están en juego otras dimensiones (o valores), en especial la eficiencia y la eficacia, pero en esta exposición nos ocuparemos sólo del valor de la equidad.

1. Ausencia de coerción. En primer lugar, por tanto, la equidad exige que los participantes interesados en llegar al acuerdo no estén sujetos a presiones coercitivas. O sea, ninguno de los participantes debe estar sometido al control o influencia especial de ningún otro. Hay que señalar que esta característica contrasta claramente con los acuerdos de negociación colectiva en los que la coerción de cada parte sobre la otra constituye un elemento integrante de las negociaciones mismas. Pero, en general, desde el punto de vista de CARE, los resultados de las negociaciones basadas en el regateo no tienen por qué ser acuerdos equitativos, aunque resulten aceptables para las partes implicadas.

En el informe general de CARE se reconocen tres tipos de coerción: *primero,* los participantes no deben sentirse coartados por otros individuos que tengan algún poder especial sobre ellos; *segundo,* no deben formarse entre los participantes grupos y facciones con el fin de ejercer presiones; *por último,* los participantes no deben sentirse coartados por las circunstancias en las que ellos mismos se encuentran. Es obvio que cualquiera puede verse tentado a todo tipo de "pactos" cuando se encuentra en circunstancias desesperadas o de tensión, incluso. Pero, en general, la existencia de tales circunstancias nos lleva a dudar de la equidad de los acuerdos alcanzados en ellas.

Creo que se puede afirmar, sin temor a equivocarnos, que el contrato de Michigan no estuvo sometido a ninguna presión de este tipo. Ni los portavoces de las asociaciones ni los evaluadores sufrieron presiones para llegar a acuerdos, aunque todas las partes se beneficiaban de ello. No hubo coerción de ningún individuo sobre otro; ni por parte de grupos sobre individuos, ni las circunstancias de ninguna de las partes eran desesperadas. Ambos grupos podían haber abandonado el proyecto sin grave lesión de sus intereses. Las asociaciones no necesitaban que se hiciera el trabajo ni, desde luego, que lo realizara este grupo concreto y los evaluadores no tenían necesidad imperiosa de trabajo.

Y, ¿qué decir de la situación más habitual, en la que un organismo gubernamental patrocinador pide que se realice una evaluación? ¿Es esto coercitivo? En la mayoría de los casos, creo que no. Aunque el organismo gubernamental pida una evaluación, esa petición suele formar parte de un contrato previo de financiación entre el organismo y el beneficiario de los fondos. En realidad, la misma equidad exige que se lleve a cabo la evaluación si se pide en virtud del contrato de financiación original.

La situación se hace menos obvia cuando el organismo gubernamental especifica el contenido del contrato de evaluación y designa quién se hará cargo de la misma. Es posible que el beneficiario tenga poco margen para negociar en relación con sus propios intereses. No obstante, si la petición

© Ediciones Morata, S. L.

Scanned by CamScanner

de la evaluación se hace al amparo del contrato original de financiación y el beneficiario lo entiende de forma adecuada, la petición será equitativa, salvo que el mismo contrato de financiación no sea equitativo. En este último caso, las circunstancias crean un marco apto para la coerción y, por tanto, para contratos no equitativos.

Supongamos, por ejemplo, que un laboratorio educativo regional necesita fondos desesperadamente. Quizá su misma existencia dependa de la consecución de un contrato importante con el gobierno. En tales circunstancias, es mucho más fácil que se llegue a un acuerdo que tanto sus signatarios como terceros considerarían no equitativo. En el marco de la evaluación, esta situación se refleja de un modo característico en cientos de pequeñas firmas privadas de consultoría que llevan a cabo gran cantidad de evaluaciones para organismos gubernamentales. Muchas de estas empresas tienen necesidades acuciantes de dinero, por lo que son mucho más susceptibles a la coerción en los contratos que las instituciones más estables.

De hecho, a veces las evaluaciones que realizan estas empresas son escandalosas y su desenmascaramiento sólo depende de que algún periodista investigador disponga de tiempo libre. La política deliberada del gobierno durante la administración Nixon, que gastó varias decenas de millones de dólares en financiar pequeñas firmas privadas de consultoría, se basaba en la idea de que estas firmas "responderían mejor" a los organismos gubernamentales. En el mejor de los casos, esta capacidad de respuesta significaba la disposición para negociar cuestiones relevantes de la evaluación, una idea importante y válida; en el peor, significaba la vulnerabilidad a la coerción gubernamental, es decir, la disposición para preparar el tipo de estudio y producir la clase de logros pretendidos por un organismo concreto. Me temo que los resultados hablan por sí mismos. Los contratos no equitativos y firmados bajo coerción producen resultados no equitativos y sesgados.

2. Racionalidad. La segunda condición de la equidad consiste en que los participantes sean racionales. No quiere decir esto que deban llevar a cabo raras proezas mentales, sino que sean capaces de discurrir de una proposición a otra, presentar pruebas que avalen sus puntos de vista y estimar las consecuencias de sus acciones (y que las tengan en cuenta). Parece obvio que las personas son racionales en distintos grados, pero, como especula CARE, no es preciso que los participantes sean igualmente racionales para que un acuerdo sea equitativo.

Creo que, en general, la racionalidad de los participantes no supone un problema para llegar a acuerdos de evaluación. Puede que las personas sean racionales de modos algo diferentes. Por ejemplo, los evaluadores suelen estar más a favor del razonamiento de medios a fines y de los argumentos pragmáticos. Quizá a otros participantes en el contrato este tipo de razonamiento no les parezca tan persuasivo como a los evaluadores, pero creo que, en la mayoría de los casos, no habrá diferencias significativas

© Ediciones Morata, S. L.

en esta área. (Véase un análisis de los tipos de razonamiento utilizados por los evaluadores en el Apéndice A.)

3. Aceptación de los términos. La tercera condición de la equidad es que las partes signatarias del contrato acepten las reglas operativas que el procedimiento impone a su pensamiento. Esas "reglas operativas" estructuran los esfuerzos para alcanzar el contrato, p. ej., en contextos legislativos conocidos, las *Robert's Rules of Order* o, en otros contextos, quizá las mismas limitaciones mutuamente aceptadas del debate informal. En general, cuanto mayor sea la dificultad que planteen los términos en los que discurre el esfuerzo para alcanzar el acuerdo —por ejemplo, cuanto más nos sintamos artificialmente limitados o comprometidos por ellos—, menos equitativo parecerá el acuerdo al que se llegue. Podemos, por ejemplo, imaginar casos en los que los evaluadores sientan que en absoluto tendrían que negociar el contenido de la evaluación con los clientes. En otros casos, las partes pueden estar dispuestas a negociar determinados objetivos de la evaluación, pero no los métodos. En la evaluación de Michigan, creo que sus términos y procedimientos resultaron aceptables para ambas partes.

4. Acuerdo conjunto. Otra condición de la equidad es que el contrato sea "conjunto". Es difícil formular con claridad esta cuestión pero intuitivamente, podemos decir que la equidad requiere que el acuerdo se considere como algo más que una coincidencia entre opciones individuales. Los participantes han de ser conscientes de que la consecución del resultado requiere el acuerdo de los demás y que el resultado ha de constituir un objeto de compromiso respecto al cual se espera la aquiescencia de todos los miembros del grupo que establece el contrato. En otras palabras, no podemos fijar sin más nuestra postura y esperar que los demás se comprometan. Ha de quedar claro que el contrato constituye el objeto del esfuerzo de grupo, alcanzado por los miembros del grupo que lo establece, y no como algo que se acepta dependiendo del estado de ánimo del momento. La estructuración de un contrato verdaderamente "conjunto" de este modo convierte las negociaciones en un proceso serio y difícil en vez de quedar como un asunto casual y sencillo. De nuevo, el cumplimiento de esta condición es cuestión de grado.

Creo que, en general, en el caso de Michigan esta condición se satisfizo, aunque quizá en menor grado que algunas otras. Algunos evaluadores (en especial yo mismo) dudaban a veces entre actuar de acuerdo con sus propios principios y comprometerse en un contrato conjunto. En realidad, una de las razones por las que, al principio, no estaba especialmente entusiasmado por establecer un contrato escrito era que yo pensaba que actuaría de acuerdo con mis principios personales y que los demás harían lo mismo, por lo que el contrato escrito estaría de más. Pero ahora creo que un acuerdo conjunto explícito es tan fundamental para definir las expectativas y obligaciones, que determina el que todas las partes sientan que han sido tratadas con equidad.

© Ediciones Morata, S. L.

Scanned by CamScanner

Por supuesto, aunque se llegue a un acuerdo conjunto, conviene añadir que aún podemos imaginar situaciones en las que el cumplimiento de las condiciones de equidad sea menos importante que la aplicación de algún otro principio. Por ejemplo, puede que sea menos importante conseguir una evaluación equitativa (basada en las previsiones contractuales) que llevar a cabo una que sea justa, p. ej., una evaluación que no lesione a los miembros menos favorecidos de la sociedad o que no sea incoherente con el valor fundamental de la igualdad. Se plantea aquí una importante cuestión filosófica respecto a si un contrato que satisfaga estas condiciones "de procedimiento" relativas a la equidad, como las expuestas en este capítulo, podría incumplir algunos principios independientes y esenciales de la justicia, de aplicación más general y profunda, como los que acabamos de mencionar. Planteamos aquí la cuestión y volveremos sobre ella más adelante, aunque no se resolverá en este capítulo.

5. Desinterés. Esta condición es una de las primeras que surgen cuando reflexionamos sobre la equidad, aunque quizá sea la más difícil de cumplir en la práctica. Para llegar a un acuerdo, no tenemos que prestar excesiva atención a nuestros propios intereses. Podemos manifestarlos y hacer un trato o llegar a un compromiso con otros, pero el trato no tendrá por qué ser un contrato *equitativo.* No quiere decir esto que debamos dejar completamente de lado nuestros intereses. Debemos procurar, más bien, lo que solemos llamar "imparcialidad" cuando tratamos de alcanzar un acuerdo. Existen diversas formas de organización que pretenden ayudar a conseguir la imparcialidad en el juicio, no hace mucho expuestas por SCRIVEN (1976a, 1976b). En todo caso, esta condición de desinterés admite diversos grados de cumplimiento.

6. Universalidad. Íntimamente relacionada con la condición de desinterés está la de universalidad. Pretendemos establecer una normativa que afecte a todos por igual, incluidos nosotros mismos. Podemos entender esto en el sentido de que, si la normativa exige que se trate a las personas de distintos modos, estaríamos dispuestos a ocupar cualquiera de las diferentes situaciones en las que se aplique esta política. De nuevo, la capacidad y disposición para discernir y procurar dar soluciones, en este sentido "universales", parece cuestión de grado.

7. Interés comunitario. Aunque debemos evitar la preocupación excesiva por los intereses individuales, quienes procuran llegar a un acuerdo deben optar por una política que sea mejor para todos ellos, en cuanto miembros del grupo a quienes se va a aplicar la normativa. La equidad permite que las personas se interesen por ellas mismas como miembros del grupo (y, quizá, en distintas situaciones dentro del mismo), sobre quienes tendrá alguna influencia lo acordado. En la evaluación, esto debe incluir a las personas evaluadas, aunque por cualquier causa no estén de hecho

© Ediciones Morata, S. L.

representadas en la negociación entre los patrocinadores de la evaluación y los evaluadores. También han de tenerse en cuenta sus intereses.

Estas tres últimas condiciones de equidad están muy relacionadas entre sí. Quizá en su conjunto —desinterés individual, universalidad e interés comunitario— constituyan una representación del valor moral de la imparcialidad. CARE (1978) afirma que deliberar de un modo imparcial supone deliberar sin prestar excesiva atención a los propios intereses, considerar sólo soluciones universales y optar de acuerdo con los propios intereses en cuanto miembro del grupo afectado por el contrato.

He dicho antes que la imparcialidad es una condición necesaria para una evaluación equitativa y que la objetividad es una condición insuficiente (Capítulo IV). Si la objetividad supone un procedimiento exteriorizado, público y repetible, es muy posible que tal procedimiento objetivo se incline a favor de determinados programas. Por ejemplo, un test basado en un *curriculum* puede resultar sesgado cuando se aplica a otro *curriculum* diferente, aunque siga siendo objetivo en el sentido mencionado (véase WALKER y SCHAFFARZICK, 1974).

Aunque la objetividad indica una carencia de actitud o una actitud neutral, la imparcialidad se refiere a la capacidad de equilibrar (de forma justificada) distintas actitudes en relación con el objeto que se juzga. Las personas a quienes se juzga no esperan que sus evaluadores se despreocupen, sino que sean capaces de reconocer sus preocupaciones y de equilibrarlas entre sí. Por ejemplo, ante un tribunal de justicia, no querríamos que nos juzgase un juez perteneciente a otra cultura que fuera perfectamente objetivo aplicando criterios sin importancia, sino otro de nuestra misma cultura que, siendo imparcial, utilizase criterios pertinentes. Las tres condiciones de CARE, sobre todo la de interés comunitario, reflejan esta característica de los procedimientos para establecer contratos equitativos.

¿Cómo era el contrato de Michigan a este respecto? ¿Cómo se cumplieron estas condiciones que CARE asocia con la imparcialidad? Una de las más sonadas protestas del *Michigan Department of Education* se refería a que la evaluación era "sesgada" (KEARNEY y cols., 1974). Respecto a la condición de desinterés, creo que el contrato era equitativo. Ninguna de las partes se excedió promoviendo sus propios intereses. Los representantes de las asociaciones no controlaron en absoluto el proceso de evaluación ni los posibles resultados negativos. En la medida en que yo puedo detectarlos, los intereses de los evaluadores no aparecían reflejados en exceso. La remuneración financiera era bastante modesta y, de hecho, los evaluadores sabían que podían imponerse determinadas sanciones si los resultados fuesen negativos para el organismo estatal. La evaluación supuso una publicidad personal de considerable magnitud, pero que habría sido la misma si los resultados hubieran sido diferentes.

Me parece, también, que se cumplió en una medida razonable la condición de universalidad. Habríamos evaluado del mismo modo cualquier otro programa semejante y nos hubiera gustado que evaluasen nuestros propios programas de igual manera. Buena prueba de ello es que pudimos

© Ediciones Morata, S. L.

Scanned by CamScanner

justificar ante nuestros colegas del campo de la evaluación nuestros procedimientos y resultados con notable éxito.

Más problemática es la tercera condición de la equidad (interés comunitario), en la medida en que se relaciona con la imparcialidad. Es cierto que los representantes de las asociaciones escogieron a los evaluadores que quisieron y que, a juzgar por sus trabajos precedentes, mejor podían representar sus intereses de grupo, de la misma forma que los organismos gubernamentales suelen tener libertad para escoger a los evaluadores de sus programas. No obstante, tras reflexionar sobre la experiencia en general, no estoy seguro de que los intereses del *Michigan Department of Education* estuvieran suficientemente reflejados en el contrato, aunque de ningún modo se hallaran ausentes en nuestras deliberaciones. Por ejemplo, no permitíamos que un miembro de una asociación estuviera presente cuando recibíamos testimonios a menos que también se invitara a un funcionario del departamento[1].

La situación se hizo más compleja por el hecho de que nuestra evaluación se centrara sobre el sistema de rendición de cuentas de Michigan que, ya de por sí, es un esquema de evaluación en el que (a nuestro modo de ver) no están bien representados los intereses más importantes de los profesores y alumnos de Michigan. Una de nuestras críticas principales consistió en que ese esquema de evaluación no era equitativo en relación con algunas de las condiciones que expongo aquí. No obstante, si se critica nuestro contrato por su parcialidad, creo que es más vulnerable a la acusación de la falta de representación de los intereses de una de las partes más afectadas por la evaluación, es decir, el *Michigan Department of Education*. El fallo consistente en la insuficiente representación de los intereses de quienes se ven afectados por la evaluación es uno de los más corrientes en estos procesos, sobre todo cuando los afectados no participan en el contrato.

8. Información igual y completa. Otra condición de la equidad consiste en que los participantes reciban información sobre los hechos pertinentes y que ésta se distribuya por igual. Se pretende que unos participantes no puedan disponer de información pertinente de la que carezcan los demás. Creo que es característico que esta condición no se cumpla del todo. Si el evaluador sabe que pocos programas son capaces de provocar la elevación de las puntuaciones de los tests normalizados en condiciones de control

[1] STUFFLEBEAM manifiesta que, antes de que firmáramos el contrato de evaluación, él acudió al *Michigan Department of Education* en compañía de su decano, para informar al departamento de lo que se estaba preparando, pedir su cooperación y solicitar su punto de vista. Los funcionarios de Michigan se incomodaron a causa de la evaluación que se iba a hacer, pero dijeron que cooperarían y se alegraron de que STUFFLEBEAM participara en ella. Philip KEARNEY, a la sazón Inspector General Adjunto de Michigan, advirtió que podían verse afectadas las relaciones de STUFFLEBEAM con el Departamento de Estado, dependiendo de cómo se desarrollase la evaluación. Así fue. Al menos en este sentido, los intereses del *Michigan Department of Education* estuvieron representados en la evaluación (correspondencia privada).

© Ediciones Morata, S. L.

minucioso, se considera equitativo que, en las negociaciones, se informe al cliente de tal extremo, suponiendo, claro está, que constituya una información pertinente. Del mismo modo, el cliente o patrocinador no debe ocultar al evaluador hechos importantes. La negociación respecto a un diseño multivariado con un grupo no versado en el tema lo sitúa en una posición incómoda, pues sus miembros sólo tendrán una idea vaga de lo que suponen los procedimientos multivariados. Cuanto más técnico sea el diseño de la evaluación, más difícil será para quienes no sean evaluadores llegar a un contrato equitativo, desde su punto de vista, pues los introduce en el ámbito de la mistificación.

9. Sin riesgos. CARE dice también que el contrato debe plantearse de manera que los participantes no "teman" las consecuencias de la implementación del programa, porque si ellos, que prestan su aquiescencia al programa en cuestión, no creen que merezca la pena implementarlo, no es fácil que adquieran una conciencia clara de su papel, como participantes en un esfuerzo conjunto para resolver un problema *previo acuerdo*. Esta condición plantea especiales dificultades para establecer un contrato en relación con una evaluación porque, en general, ésta siempre encierra ciertos riesgos, sobre todo para las personas sometidas a la misma. Pero, por otra parte, uno de los propósitos de un contrato de evaluación consiste en *reducir* los riesgos que pudiera suponer para quienes se someten a ella, dándoles voz en la evaluación y fijando ciertos límites a la misma. Por ejemplo, SCRIVEN (1976a) propone que los evaluados tengan derecho a examinar un primer borrador del informe de evaluación antes de que se entregue al cliente, así como el derecho a incluir en dicho informe puntos de vista contrarios a los expuestos en el cuerpo del mismo.

Algunos de los trabajos más interesantes de teoría de la evaluación están dedicados a proteger al evaluado de evaluaciones incompetentes y arbitrarias. Aunque el riesgo que supone para el evaluado pueda reducirse, no parece posible eliminarlo por completo. En la evaluación del sistema de rendición de cuentas de Michigan, por ejemplo, todas las partes implicadas corrieron riesgos definidos; no conozco forma alguna de *equipararlos*.

10. Posibilidad. Otra condición del contrato consiste en que sea posible llevarla a cabo. Por desconcertante que parezca, tratándose de una condición trivial, los evaluadores suelen pasarla por alto. También es demasiado habitual que los evaluadores prometan actividades y oportunidades fantásticas por poco dinero que luego no pueden proporcionar. No sé qué porcentaje de evaluaciones entran en esta categoría, pero es muy elevado. En parte, esta incapacidad está relacionada con la inexperiencia en la evaluación y con un desmedido interés por conseguir el contrato. Tampoco es raro que la otra parte no cumpla sus promesas, sobre todo la de facilitar el acceso a las fuentes de datos. Cuando es imposible cumplir el contrato, éste no es equitativo. En estos casos, la informalidad de una parte hace que la otra vea incumplidas sus expectativas, a pesar del compromiso.

© Ediciones Morata, S. L.

Scanned by CamScanner

En la evaluación de Michigan, ambos grupos cumplieron lo que habían prometido.

11. Contar con todas las opiniones. Para que el contrato sea equitativo, las partes deben tener la oportunidad de hacer constar lo que crean conveniente en la etapa final del proceso de preparación del acuerdo. No obstante, teniendo presentes aquellos contextos en los que prima una justicia de compensación, CARE señala que, para que el contrato sea equitativo, no hay por qué *equiparar* la opinión de todo el mundo.

12. Participación. La condición final del contrato es que todas las partes puedan participar. Todo el mundo debe "tener voz", de acuerdo con lo estipulado en los procedimientos. En la evaluación de Michigan, se cumplieron estos dos últimos criterios en relación con los procedimientos. Todas las partes signatarias del contrato tuvieron oportunidad de participar y de hacer oír su "voz".

Reflexiones sobre las condiciones de equidad

Como dice CARE, es interesante señalar que sólo algunas de las doce condiciones necesarias para un contrato equitativo lo sean, a su vez, de los procedimientos para llegar al acuerdo. La inmensa mayoría de las condiciones aquí expuestas corresponden a los mismos *participantes*. Esto resulta un tanto sorprendente porque la mayoría de nosotros tiende a pensar que lo que garantiza o justifica la equidad de una evaluación es la adecuada selección de los *procedimientos*. Solemos creer que, si los utilizados son correctos, el contrato será satisfactorio. Pero si es cierto lo que comentamos, es patente que muchas de las condiciones que influyen en la equidad escapan del control de los procedimientos, en sentido estricto, del mismo modo que el gobierno de la mayoría no garantiza la justicia en una democracia. Cuestiones tales como el grado de coerción, el desinterés, etc., de los participantes constituyen importantes influencias en la justicia de los contratos a los que se llegue y de la consiguiente evaluación. Es obvio que esto no significa que los procedimientos "no tengan nada que ver". Si se emplean los adecuados *y* se cumplen las demás condiciones, el contrato será equitativo. Es más, hay distintos procedimientos diseñados para reducir la parcialidad, la coerción, etc., y satisfacer las demás condiciones. La cuestión importante es que los procedimientos solos no garantizan el cumplimiento de aquellas condiciones ni que el contrato de evaluación sea equitativo.

Una segunda característica de estas condiciones es que, en la mayoría de los casos, su cumplimiento parece una cuestión de grado. Un contrato puede ser más o menos equitativo, dependiendo del grado en el que se cumplan diversas condiciones. Es probable que los contratos no sean del todo equitativos ni parciales, aunque cuanto mayor sea el grado de cumplimiento de las condiciones, más fácil será que el contrato se aproxime al ideal de equidad.

© Ediciones Morata, S. L.

Es más, siempre existirá una incertidumbre básica respecto a la equidad de un contrato. Como indica CARE, es imposible asegurar que los demás, e incluso uno mismo, cumplan plenamente muchas de estas condiciones (p. ej., desinterés, ausencia de coerción). SCRIVEN dice que, aunque nunca podamos eliminar por completo los sesgos, el objetivo de la evaluación debe ser reducirlos a proporciones manejables. De igual manera, aunque nunca podamos garantizar que seamos completamente equitativos, el objetivo consiste en serlo lo más posible.

¿Qué podemos observar en el contrato mismo? En la monografía de STAKE (1976), diversos evaluadores y administradores que han realizado evaluaciones asesoran sobre el *contenido* de un contrato de evaluación. Algunos aspectos de este contenido (p. ej., los propósitos de la evaluación, la posibilidad de llevarla a cabo, la participación, la comunicación de la información, la confidencialidad, la información proporcionada por una parte a otra) corresponden a las condiciones necesarias para la equidad. Otros consejos no se refieren a la equidad del contrato. En general, el contrato de evaluación sólo proporciona indicios limitados acerca de su equidad.

Por ejemplo, en el contrato de Michigan (que aparece en el Apéndice C), es evidente que hubo una negociación y que se llegó a un acuerdo mutuo respecto a los destinatarios y las cuestiones sometidas a investigación. Los evaluadores hicieron valer su derecho a emitir el informe y los representantes de las asociaciones el suyo de aceptarlo o no. Hubo acuerdo respecto a los recursos, al sistema de distribución de los mismos y al acceso a los datos. Por último, los procedimientos de evaluación propiamente dichos quedaron bajo el control de los evaluadores, aunque los procedimientos esenciales se definieron de mutuo acuerdo.

Por tanto, a partir del contrato, podemos suponer razonablemente, si no inferir, en sentido estricto, que las condiciones de racionalidad, aceptación de los términos, acuerdo conjunto, posibilidad y participación se han cumplido. Sin embargo, basándonos sólo en el documento, poco puede decirse sobre las condiciones de falta de coerción, desinterés, universalidad, interés comunitario, información igual y completa, ausencia de riesgos y voz de todos los participantes. El hecho de que los evaluadores tengan derecho a emitir su informe sin censura es importante, porque ayuda a eliminar el problema del interés propio de los representantes de las asociaciones si tuvieran la posibilidad de censurarlo. En este caso, ese procedimiento reduce, aunque no elimina, la posibilidad de inclinaciones tendenciosas. Los sesgos pueden introducirse por otro lado, como el interés propio de los evaluadores. Por supuesto, también podemos buscar la confirmación o no de las tendencias en el informe final y en los mismos procedimientos de evaluación.

© Ediciones Morata, S. L.

Scanned by CamScanner

¿Hasta qué punto es vinculante el contrato?

Se trata de una cuestión difícil y nos recuerda el problema, en el ámbito de la jurisprudencia, de si los ciudadanos tienen o no obligación (moral) absoluta, *prima facie*, de obedecer las leyes de su comunidad establecidas de forma regular. Pero la mayoría de los evaluadores reconoce que hay un momento en el que incluso un contrato equitativo debe rescindirse. Por ejemplo, en situaciones difíciles, como las que indican hechos delictivos, cuando se plantea el dilema entre rescindir un contrato o seguir la propia conciencia, STUFFLEBEAM dice:

> Creo que el evaluador no debe ser el único responsable de las decisiones si ha aceptado de antemano limitar su informe a ciertos destinatarios definidos. En cambio, pienso que debe reemprender la negociación del contrato y tratar de llegar a acuerdos con las otras partes respecto a la publicidad de la información y cumplirlos en consecuencia. Por supuesto, hay ocasiones en las que se suscitan cuestiones de moralidad fundamental, como las surgidas en el conflicto del Watergate. En estos casos, creo que el evaluador debe hacer lo que crea en conciencia que se deba en beneficio de la sociedad y teniendo en cuenta los intereses de las personas a quienes afecte la publicidad que se dé a la información.
>
> (STAKE, 1976).

Creo que la postura de STUFFLEBEAM refleja el consenso mayoritario de los expertos en evaluación, aunque no todos estén de acuerdo. Ralph TYLER dice: "No debe considerarse válida conciencia individual alguna. La manera de tratar la información ha de figurar en el contrato" (STAKE, 1976).

El punto de vista filosófico, representado por CARE (1978), consiste en que, si se han cumplido las condiciones de equidad "en la mayor medida posible", tendremos buenas razones para aceptar los resultados del acuerdo. No obstante, podemos concebir el caso de que el rechazo de los mismos sea "racional y correcto", cuando se cuestione el cumplimiento de las condiciones antes expuestas. En la medida en que nos fiemos menos del total cumplimiento de las condiciones, más justificada estará la reserva del derecho a rechazar los resultados del acuerdo. En todo caso, la oposición a los mismos constituye una cuestión muy seria, que no puede tomarse a la ligera ni pasarse por alto sin causa suficiente. Está en juego la equidad.

Mi propia interpretación del derecho a rescindir un contrato equitativo concuerda con las opiniones hasta aquí expuestas. Es característico que surja la cuestión de la equidad en el contexto de contratos voluntarios establecidos en beneficio mutuo. Pero cuando entran en conflicto las peticiones a favor de la equidad y las que se refieren a la justicia, esta última tiene prioridad. Por ejemplo, si el cumplimiento de un contrato equitativo significa que tengamos que cometer un acto injusto, hemos de romper el contrato. Del mismo modo, las reclamaciones a favor de la equidad tienen

© Ediciones Morata, S. L.

un carácter secundario ante las peticiones que defienden una obligación natural, como la de ayudar a una persona que se encuentre en peligro.

Es evidente que no todo el mundo está de acuerdo sobre determinados principios específicos de justicia. Como he indicado en otra parte, en el mismo campo de la evaluación hay diversas opiniones al respecto. Según la concepción rawlsiana, el principio de equidad se sitúa dentro de los límites marcados por dos principios de justicia. Como estos principios están definidos de manera muy amplia, yo no esperaría que los detalles concretos de un contrato de evaluación estuviesen determinados por ellos. No obstante, existen posibles áreas de conflicto. Por ejemplo, si el contrato exigiera que un evaluador violara la autoestima de alguien sometido a evaluación, estaría obligado a actuar de forma no equitativa respecto al cumplimiento de los términos contractuales antes de actuar de manera injusta, pues la autoestima constituye un bien primario en la concepción rawlsiana de justicia. En la concepción utilitaria de la justicia, no podríamos transgredir el principio de elevar al máximo las satisfacciones netas de la sociedad.

En resumen, la equidad es una idea que aparece por todas partes en la evaluación, casi siempre en un nivel intuitivo de conciencia. Aquí nos hemos ocupado de conocer las condiciones que ha de cumplir un contrato de evaluación para que pueda considerarse equitativo. Entre esas condiciones están: la ausencia de coerción, la racionalidad, la aceptación de los términos, el acuerdo conjunto, el desinterés, la universalidad, el interés comunitario, la información igual y completa, la falta de riesgos, la posibilidad, la voz de todos los participantes y la participación. El contrato equitativo es vinculante salvo que se oponga a nuestros deberes naturales o a principios morales más importantes como los de justicia.

© Ediciones Morata, S. L.

Scanned by CamScanner

CAPÍTULO IX

Poder y deliberación

Las debilidades del liberalismo

La cuestión sobre cómo quedan representados en una evaluación el interés público y los de todas las partes interesadas es fundamental. Los enfoques utilitaristas asumen que los índices de bienestar reflejan el interés público y que los gobiernos deben emprender acciones claras sobre la base de estas decisiones. Los liberales modernos del bienestar creen en la concentración del poder en el gobierno, poder ejercido en nombre del interés público o de los intereses de grupos especiales a quienes el gobierno trate de ayudar, como los más desfavorecidos.

Como el utilitarismo permite el beneficio de las clases superiores a costa de las pretensiones de las inferiores, al constituir el criterio la cantidad total de satisfacción de aspiraciones, este enfoque puede resultar antidemocrático, sobre todo por no representar los intereses de quienes se encuentran en peor posición. No tiene por qué ser así, pero al buscar la máxima satisfacción bruta de aspiraciones, la política federal no se ha ocupado bastante de la distribución de aquélla. Se ha preferido la eficiencia a la justicia.

Los enfoques más democráticos de la evaluación, los participativos, son pluralistas en dos sentidos. En el primero, se basan en una combinación de diversos principios o valores y no en un solo criterio, como ocurre en un enfoque utilitarista estricto. En el segundo, representan diversos intereses distintos; teóricamente, los intereses de todos los afectados por el programa o política. En la medida en que respondan a los intereses de los miembros de la comunidad, podemos considerarlos democráticos (CARE, 1978).

Sin embargo, incluso los enfoques pluralistas pueden criticarse en la misma medida en que puede serlo el pluralismo político. El pluralismo ha recibido ataques por discriminar ciertas clases de problemas. Sólo se tienen en cuenta los planteados por grupos legitimados para ello. Si un problema

© Ediciones Morata, S. L.

no reviste especial interés para un grupo social legitimado, suele dejarse de lado. Es difícil plantear problemas relativos al bien común, referidos al bienestar de toda la sociedad, porque no hay ningún grupo concreto al que le interese plantear tales problemas. En realidad, la percepción de la sociedad como la suma de individuos o grupos dificulta la identificación del bien común.

En la versión "arbitral" del pluralismo, el gobierno actúa como árbitro, dejando que los individuos resuelvan sus propios problemas, aunque asegurándose de que se cumplan las reglas. Es un papel semejante al del evaluador "democratico" de MACDONALD, que hace de intermediario entre individuos, facilitando el intercambio de información entre ellos, aunque no se interfiere de forma directa ni hace recomendaciones de su propia cosecha.

Se ha criticado la versión arbitral del pluralismo por favorecer a los individuos más fuertes en perjuicio de los débiles y por consolidar el poder de aquéllos en las posiciones de mayor importancia (WOLFF, 1968). Por ejemplo, cuando desempeñan este papel organismos reguladores del gobierno, los neutrales sólo se ocupan de los líderes de los distintos grupos de intereses y no de los miembros de base de tales organizaciones. En consecuencia, a menudo los intereses representados corresponden a los ejecutivos, los dirigentes sindicales, los agricultores más influyentes, etc.

Las acciones arbitrales emprendidas para responder a los más fuertes contribuyen a la consolidación de las posiciones ocupadas por los líderes, desempeñando, por tanto, un papel más conservador que neutral. Aunque el fundamento del enfoque consiste en que queden representados los intereses de cada uno de los individuos, la interacción libre entre fuerzas opuestas conduce a la dominación de los más fuertes (al menos, eso dicen las críticas).

La versión del "vector suma" del pluralismo recibe una crítica semejante. Concibe los grupos, y no los individuos, como unidades básicas. La representación de los intereses del individuo se hace a través de su grupo. Diversos grupos tratan de conseguir que el gobierno actúe en su nombre. El papel del gobierno consiste en responder a esos intereses de grupo. Las posteriores acciones y orientaciones del gobierno vendrán determinadas por el vector suma de tales presiones de los grupos.

Ese papel es semejante al que desempeña el evaluador "responsivo" de STAKE. El evaluador actúa en respuesta a los grupos que participan en el programa y tienen relación con el mismo, configurándose la evaluación de acuerdo con esas presiones cambiantes. En el desarrollo de la evaluación, el evaluador responde a los intereses, opiniones, preocupaciones, etc., del grupo. Como en la versión arbitral del pluralismo, la del vector suma considera legítimas las peticiones rivales, pero no se define con respecto a ellas.

La crítica que se hace contra la versión del vector suma del pluralismo consiste en que no se tienen en cuenta los intereses de todos los grupos (WOLFF, 1968). La distinción entre grupos legítimos e ilegítimos es muy clara y no hay obligación de responder a los "ilegítimos". Por regla general, se

© Ediciones Morata, S. L.

Scanned by CamScanner

admite que los grupos organizados son legítimos, a diferencia de los no organizados. Están mejor representados los que disponen de recursos y técnicas de organización, lo que también favorece a los grupos más pequeños, que pueden reunirse con mayor facilidad. De este modo, en la práctica salen beneficiados los grupos establecidos, en perjuicio de los de reciente formación. Los perjudicados, que tienen intereses pero poca fuerza, no son admitidos. En la acción de gobierno, esta versión actúa a menudo como un freno para el cambio social (al menos, eso dicen las críticas).

En resumen, los enfoques participativos de la evaluación son significativamente más democráticos que otros muchos enfoques liberales. Ofrecen más opciones sobre una cantidad mayor de cuestiones a un público más amplio. Reflejan los intereses de muchos grupos. No obstante, también son deficientes. No se libran del todo de algunas dificultades del liberalismo en general y del pluralismo en particular.

Tampoco excluyo de esta crítica mi propia propuesta de contrato para una evaluación equitativa (HOUSE y CARE, 1979). Aunque, en teoría, resulta más aceptable a los participantes, desde el punto de vista moral, es muy probable que ese contrato se celebre, en la práctica, entre el evaluador y el patrocinador. Una de las condiciones de un contrato equitativo es que las partes que establezcan un acuerdo tengan en cuenta los intereses de todos los afectados por el mismo, pero ésta constituye sólo una cautela, no una garantía.

Poder e intereses

La mayor parte de las críticas que se dirigen contra el liberalismo se refieren a que no están representados por igual todos los intereses porque, en general, un cierto desequilibrio a favor del poder impide que ello se lleve a cabo. LUKES (1974) codifica estas objeciones. El liberalismo —dice— toma a los hombres como son y aplica a lo que ellos prefieren unos principios relativos a sus pretensiones. Sus preferencias políticas se manifiestan mediante la participación política (como en los enfoques participativos de la evaluación).

Los pluralistas asumen que los "intereses" pueden entenderse como preferencias políticas concretas. Un conflicto de intereses equivale a uno de preferencias. Los liberales se oponen a la idea de que los intereses de grupo puedan no estar articulados, ser inobservables o que las personas puedan no ser conscientes de sus propios intereses. Tal como los conciben los liberales, los intereses son intereses "subjetivos", tal como los ven las partes.

En correspondencia con esta concepción de los intereses está el punto de vista liberal sobre el poder, que se centra en la conducta individual al tomar decisiones acerca de cuestiones en las que existe un conflicto observable entre intereses subjetivos. Éstos deben expresarse como preferencias políticas que se ponen de manifiesto en la participación política

© Ediciones Morata, S. L.

(LUKES, 1974). Es obvio que en los enfoques utilitaristas de la evaluación, la participación se limita a la "valoración de las necesidades", diseñada para recoger las pretensiones. En los enfoques pluralistas, la entrevista es capaz de detectar preferencias políticas complejas.

Otra concepción es la de los reformistas, que dicen que nunca llegan a plantearse muchos conflictos y problemas a causa de desequilibrios de poder. Hay muchos problemas que nunca alcanzan el nivel de las decisiones porque se oponen a los intereses y valores de los responsables de las mismas. Por tanto, se suprimen. Existe un conflicto entre los intereses de los responsables de las decisiones y aquellos cuyos intereses se excluyen del proceso de decisión.

Desde el punto de vista reformista, el poder supone la capacidad de definir los problemas y enterrar conflictos, y no sólo de decidir los conflictos cuando surgen. Aún se conciben los intereses como subjetivos, que pueden articularse y observarse. Sencillamente, algunos intereses nunca tienen la oportunidad de llegar a la superficie. En consecuencia, los reformistas tienen un concepto más amplio de los "intereses" que los pluralistas. Aquéllos no sólo consideran intereses legítimos las preferencias políticas de quienes participan en el sistema político, sino también las de los parcial o totalmente excluidos del mismo.

El tercer punto de vista sobre el poder es el de los radicales. Dicen que la perspectiva reformista también es individualista, desde la perspectiva metodológica. El poder consiste en fuerzas colectivas de fórmulas sociales, así como en los individuos que llevan a cabo sus pretensiones frente a la oposición de los otros. El conflicto que en realidad se observa no es el único indicio del poder, puesto que éste puede ejercerse de manera que determine lo que desee una persona. El conflicto puede no ser observable. En realidad, el sistema social configura las pretensiones de la persona, y esas pretensiones pueden oponerse a sus mejores intereses. Los que percibe de forma subjetiva la persona pueden no ser los "verdaderamente" suyos.

En consecuencia, pueden existir conflictos "latentes" entre los intereses de quienes ejercen el poder y los reales de quienes están excluidos de él. *Existiría* un conflicto entre pretensiones o preferencias *si* los sometidos al poder fueran conscientes de sus propios intereses. Desde el punto de vista de los radicales, suponer que la ausencia de quejas refleja un auténtico consenso es dar carta de naturaleza a la posibilidad de un consenso falso o manipulado.

En general, liberales, reformistas y radicales están de acuerdo en que el poder no debe impedir que salgan a la luz los intereses, conflictos y problemas. Se muestran en desacuerdo respecto a la medida en que la estructura social actual permite el surgimiento de los mismos. Tanto reformistas como radicales señalan que hay intereses que nunca suelen tenerse en cuenta en las evaluaciones liberales.

En contra de la idea liberal, sostienen que muchos intereses se excluyen de las evaluaciones, sobre todo los de quienes carecen de poder. Éstos no son capaces de articular sus intereses porque carecen de igualdad de

© Ediciones Morata, S. L.

Scanned by CamScanner

oportunidades. Se dejan de lado las quejas encubiertas. Además, las decisiones reales nunca se someten a evaluación pública. Sólo se permite que los evaluadores valoren lo que quieren los responsables de las decisiones, siempre concordantes con los propios intereses de esos mismos responsables. Si el ideal consiste en que se reflejen los intereses de todos los afectados por el programa, no es habitual que se alcance, ni de lejos, ese ideal.

Desde el punto de vista empírico, hay considerables pruebas que respaldan estas afirmaciones. Pocas evaluaciones reflejan los intereses de nadie que no sean los patrocinadores ni los grupos más organizados y articulados. Con frecuencia, los problemas se definen de antemano, impidiendo así que surjan otras cuestiones en el estudio, de manera que los problemas que aparecen en el desarrollo del mismo suelen surgir de conflictos manifiestos. Incluso cuando los evaluadores tratan conscientemente de hacer valer los intereses de grupos menos favorecidos, a menudo no son capaces de llevar esto a cabo. Por ejemplo, en una evaluación que incluía un programa bilingüe, nunca conseguimos que algunos padres de origen latino se reunieran con nosotros para expresar sus puntos de vista. No tuvimos problema alguno al pedir a los padres de clase media y habla inglesa que nos manifestaran sus perspectivas (AMAREL y cols., 1979). En el supuesto de que el evaluador conozca sus intereses, los trata de forma paternalista.

La adopción e implementación de la postura radical es problemática. Requiere que el evaluador reconozca que los intereses y pretensiones que las personas perciben de manera subjetiva no concuerdan con sus intereses "reales". La definición de éstos constituye una dificultad muy importante. Una definición es que los intereses reales consisten en lo que las personas querrían y preferirían, si fueran capaces de elegir. Dicho con mayor precisión, la política X favorece más los intereses de A que la política Y, si A, habiendo experimentado los resultados de ambas políticas X e Y, escogiera X en vez de Y (LUKES, 1974, pág. 34). Con frecuencia, los intereses reales se relacionan con la elección en condiciones de autonomía.

Ahora bien, sucede que una persona puede pretender, erróneamente, una política que lesione sus propios intereses. En realidad, la evaluación constituye un método para decidir si un programa o política tiene determinados resultados. Por regla general, queremos los *resultados* de un programa o política más que la política misma en cuestión (BARRY, 1965). Por tanto, es posible que otros juzguen los intereses de alguien mejor de lo que lo haría el interesado mismo. Por supuesto, esto también puede constituir un derrotero peligroso.

El hecho de que otros juzguen los intereses de alguien puede promover una política que el interesado rechazaría desde otro punto de vista. Además, la persona puede escoger deliberadamente una política que sabe no favorece sus intereses por una simple cuestión de principio. Sin embargo, incluso el juzgar los intereses de alguien sobre la base de lo que querría si supiera los resultados no neutraliza la crítica radical. Los mismos deseos pueden ser sospechosos. En consecuencia, es difícil descubrir siquiera una definición

© Ediciones Morata, S. L.

radical que identifique cuáles son los intereses "reales", aunque podamos apreciar que los intereses aparentes sean equivocados.

Ante estas dificultades, la solución liberal clásica consiste en "aceptar" los intereses y las pretensiones percibidos tal como aparecen. Mi postura consiste en que los intereses "reales" pueden descubrirse mediante una determinada forma de participación, que permita la elección del individuo en condiciones de autonomía (CARE, 1978; HOUSE y CARE, 1979). Esta postura se acerca a la definición de los intereses "reales" o a lo que el individuo escogería de forma autónoma —postura radical—, aunque manteniendo la opción individual. El individuo puede descubrir sus intereses auténticos a través de su participación con otros para decidir la evaluación. Los intereses auténticos son aquellos a los que se llega en este proceso. Esto no supone una identificación infalible de las pretensiones reales, pero, al menos, proporciona un método de cuestionar las percibidas. Es de esperar que ambos converjan.

Poder y autoridad

Hay diversas definiciones de poder. Esencialmente, A tiene poder sobre B cuando A afecta a B de alguna forma significativa (LUKES, 1974). La definición de poder utilizada en el apartado anterior consistía en que A tenía poder sobre B cuando A afectaba a B en contra de los propios intereses de B. Otras definiciones vincularían por completo el poder con la autoridad y el consenso, disociándolo, por tanto, de los conflictos de intereses. En tal caso, el poder se ejercería con autoridad legítima.

No obstante, la definición de poder que empleamos aquí lo distingue de la autoridad. El poder puede suponer la imposición de la propia voluntad sobre alguien sin que se acepte como legítimo. La autoridad supone que B acata a A porque tanto el contenido del mandato de A como el procedimiento utilizado por A para emitirlo son legítimos. La Figura 3 ilustra estas diferencias. En consecuencia, el poder constituye o no una forma de influencia, dependiendo de que se utilicen o no sanciones, mientras que la autoridad puede constituir o no una forma de poder, dependiendo de la presencia o no de conflictos de intereses (LUKES, 1974).

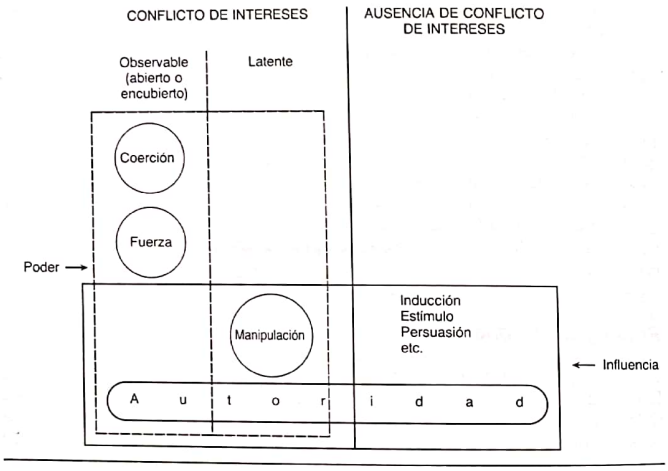
Según esta concepción, el poder se opone a la influencia. El poder surge cuando existe un conflicto de intereses. Si se utilizan sanciones, el poder se convierte en coerción (amenaza de privación) y en fuerza (ausencia de elección) y no se funda en la autoridad. Uno se impone contra la voluntad del otro. Si no existen sanciones, el poder puede ser una forma de influencia, es decir, se ejerce dentro de los límites reconocidos de la autoridad. B acata a A porque A es razonable, aunque exista un conflicto de intereses. Se trata del ejercicio legítimo del poder.

Sin embargo, desde el punto de vista radical, existe un problema. El conflicto de intereses puede estar latente. B puede apoyar una política y considerarla legítima por creer —erróneamente— que le beneficia. En

© Ediciones Morata, S. L.

Scanned by CamScanner

Figura 3: *Mapa conceptual del poder*



Nota: Tomado de Steven LUKES (1974): *Power: A Radical View.* Londres: Macmillan. Reimpreso con autorización de A. D. Peters & Co. Ltd.

esto consiste el poder ejercido en forma de "manipulación". B no se da cuenta de lo que está haciendo. Para los radicales, esto constituye un punto de partida importante. B está confundido respecto a sus pretensiones e intereses reales. Los liberales toman la aceptación de B tal como se manifiesta. Aquí se clasifica como influencia no basada en la autoridad.

Por último, cuando no existe conflicto de intereses, A se limita a ejercer influencia sobre B. Se trata de una autoridad legítima por consenso ejercida en los términos de los propios valores de B. La influencia adopta la forma de inducción, estimulación o persuasión. Sin embargo, el evaluador sólo puede utilizar la persuasión.

La única forma de influencia o poder que utiliza legítimamente la evaluación es la persuasión (Capítulo IV). A causaría que B actuara o pensara de cierto modo porque B acepta las razones de A. B consideraría autorizadas las proposiciones de A porque su contenido sería legítimo y razonable en relación con los valores de B o porque se ha llegado a ellas mediante un procedimiento legítimo. La fuerza, la coerción, la manipulación y las inducciones quedan proscritas.

En el sentido de que A causa que B haga o piense algo que, en otro caso no haría ni pensaría, la persuasión racional es una forma de poder.

© Ediciones Morata, S. L.

Pero en el sentido de que B acepta de forma autónoma las razones de A para actuar de ese modo, cambiando su forma de pensar, no va en contra de la voluntad de B. Por tanto, no se trata de poder. La persuasión racional se encuentra entre la definición de poder basada en la causalidad, por una parte, y la que se basa en la razón y la autonomía, por otra (LUKES, 1974).

Podemos suponer que los radicales estarían dispuestos a ejercer el poder sobre alguien a favor de su propio interés real, asumiendo el conflicto entre las preferencias de A y de B. En el otro extremo, podríamos suponer que cualquiera o la mayoría de las formas de control de A sobre B son ilegítimas, dando una prioridad absoluta a la autonomía de B. En el primer caso, la evaluación se desarrollaría a partir de la identificación de los intereses reales de B y los programas y políticas que puedan favorecerlos, sin la cooperación de B. Esta forma de actuar se acerca al paternalismo, pero así se desarrollan muchas evaluaciones gubernamentales. De nuevo, podemos suponer que los radicales seleccionarían unas medidas de los intereses reales diferentes de las escogidas por los evaluadores gubernamentales, pero sus enfoques y técnicas, p. ej., el análisis de sistemas, serían semejantes. En el otro extremo, los evaluadores no harían recomendaciones y procurarían que su influencia fuera mínima. Esto supone la amenaza de la anarquía y hay muy pocos enfoques de evaluación que defiendan esta línea de actuación. La mayor parte de las evaluaciones se sitúan entre ambos extremos de influencia y poder, aunque la mayoría tiende al polo paternalista.

La perspectiva radical sugiere que los evaluadores deberían preocuparse más por la estructura institucional global en la que opera el programa o política de que se trate, y del contexto social general en el que se desarrolla la evaluación. El evaluador no puede basarse sólo en la definición del problema y de la situación que le presenta el patrocinador o el responsable de las decisiones. Si lo hace así, refuerza el *status quo*, una postura de valor tan tangible y arbitraria como cualquier otra y, quizá, prime los intereses del patrocinador o responsable de las decisiones sobre las de otros. Tal postura en el evaluador no podría considerarse democrática ni moral, pues convierte la evaluación en una forma de fuerza, coerción o manipulación.

Deliberación

La esencia de las críticas reformistas y radicales contra los enfoques liberales consiste en que las relaciones de poder determinan, en realidad, las decisiones que se adopten. Una forma de comprobar estas proposiciones es comparar la deliberación que se lleva a cabo en la evaluación concreta con la forma en que debería producirse. La mayoría de las personas estaría de acuerdo en que quienes tienen el poder máximo no deberían determinar las cuestiones de la evaluación. Si éstas se deciden con arreglo a las relaciones de poder, poca falta haría una evaluación salvo como medio de justificar decisiones tomadas desde el poder. Algunos dirían que la única

© Ediciones Morata, S. L.

Scanned by CamScanner

función que desempeña la evaluación es ésa, la legitimadora, pero pocos sostendrían que ése fuera el papel que debería desempeñar.

Todos los enfoques de la evaluación dan por supuesto el significativo papel que la evaluación desempeña en el proceso de decisión. Sin embargo, cada uno de ellos parte de una situación de decisión algo distinta y ese punto de partida diverso explica algunas de las diferencias que se registran entre los enfoques. Por ejemplo, los enfoques de gestión suponen la existencia de unos pocos responsables clave de las decisiones que toman los resultados de la evaluación y, de acuerdo con sus partidarios, adoptan decisiones que afectan a muchas personas en un gran radio de acción, como sucede con los programas federales. Las técnicas de evaluación de gestión están pensadas para que proporcionen una agregación y resumen de datos que faciliten la decisión de los responsables. Aunque las relaciones de poder no carezcan completamente de interés, no se prevé que dominen la evaluación.

En el extremo opuesto del espectro de la concentración de poder, los enfoques participativos consideran una situación de decisión en la que los participantes en el programa reciben los resultados de la evaluación y se reúnen para dialogar sobre ellos y dilucidar, quizá frente a frente, qué decisiones han de adoptarse. La información de la evaluación debe ser comprensible para un grupo de este estilo, demanda muy diferente de la que plantea el estilo de gestión.

Desde el punto de vista de gestión, el problema fundamental consiste en conseguir que los resultados de la evaluación sean creíbles para destinatarios muy distantes, tanto física como psicológicamente, lo que se consigue mediante la "objetivación" de los datos, convirtiéndolos en números reproducibles gracias a la autoridad de la que goza la metodología de las ciencias sociales. En los enfoques relacionados con la difusión del poder, la credibilidad constituye un problema muy distinto. Para aquellos que conocen de cerca el programa, los resultados de la evaluación serán creíbles si no se oponen a su experiencia del programa. El problema estriba en hacer que la evaluación sea lo bastante comprensible para que los participantes la entiendan y tan pertinente que les permita actuar sobre la base de la misma. De nuevo, se da por hecho que las relaciones de poder no descartarán los resultados de la evaluación ni los deformarán para otros fines.

Otros enfoques de la evaluación imaginan deliberaciones cuasijudiciales en las que los elementos contradictorios estén igualados en poder, deliberaciones en donde los profesionales discuten entre sí o los usuarios individuales actúan en situaciones de mercado. Cada una de estas situaciones previstas de antemano impone distintas exigencias a la evaluación, pero todas ellas dan por supuesto que el poder no la invalidará.

¿Qué clase de situación de decisión *debe* prever el evaluador? Primero, en la evaluación de programas públicos, debe prever una decisión *colectiva* y no una meramente individual, privada. Aunque haya un responsable máximo de las decisiones, sólo emprenderá acciones tras un proceso de deliberación con otros y sobre la base de consideraciones públicamente justificables.

© Ediciones Morata, S. L.

...IDAD DE PLAYA ANCHA

Puede llevarse a cabo una evaluación en relación con intereses individuales, privados, como ante la adquisición de un coche. Hasta cierto punto, el proceso de evaluación es igual pero, cuando la evaluación pública se estructura como la meramente individual, se comete un error fundamental. La evaluación pública supone criterios y valores públicos. Debe prever procesos de decisión colectiva. Un proceso de decisión colectiva no es igual al de sopesar y equilibrar criterios de manera individual en beneficio propio. La construcción de los procedimientos de decisión pública como si se tratase de procedimientos individuales ampliados constituye un defecto fundamental del utilitarismo. La selección de los mejores medios para conseguir determinado conjunto de fines para un individuo no puede coincidir con el proceso encaminado hacia fines plurales. Este último necesita un proceso de decisión colectiva y, para la evaluación, esto significa un procedimiento de discusión sobre méritos.

En torno a esas deliberaciones colectivas, se polarizan dos escuelas de pensamiento (BARRY, 1973). Una de ellas sostiene que, en esas deliberaciones colectivas, hay que participar de una manera que nos permita defender nuestros propios intereses. Después de todo, cada uno es el que los conoce mejor y debe plantearlos. En consecuencia, la deliberación consiste en la suma y consideración de todos los intereses del grupo.

La otra escuela sostiene que una deliberación tal con el fin de tomar decisiones sociales es catastrófica. Si cada persona trata de conseguir sus propios intereses, es imposible llegar a una decisión justa. La discusión de los méritos degenerará, convirtiéndose en un vulgar regateo respecto a lo que conseguirá cada uno. Los compromisos se basarían en la coerción y el poder y no sobre lo que debe hacerse. En esas condiciones, la deliberación degenera hasta convertirse en un compromiso entre intereses (RAWLS, 1971).

Esta escuela sostiene que la base de la discusión debe consistir en un diálogo sobre los méritos, en el que participe cada persona, pues todas tienen contribuciones importantes que hacer. La discusión se perfecciona al presentarse puntos de vista alternativos. En esa deliberación, la discusión se funda en principios o premisas comunes a todos. Cada participante debe adoptar un punto de vista amplio y ser completamente desinteresado. El intercambio de opiniones conjura la parcialidad. Para que puedan tomarse decisiones justas, los deliberantes deben asumir un papel totalmente desinteresado. Tal es la visión de las deliberaciones en la acción.

No cabe duda de que los intereses egoístas deben limitarse para poder llegar a una deliberación racional en la que desempeñen algún papel los resultados de la evaluación. Pero este procedimiento de deliberación pura dista tanto de las situaciones en las que las personas emprenden acciones que parece imposible alcanzarlo. En relación con el contrato equitativo de evaluación, previo a la evaluación misma, hemos sugerido ciertas restricciones que deben imponerse a los intereses, pero no su eliminación. En una deliberación sobre los resultados de la evaluación con vistas a la acción, parece poco práctico y quizá imprudente excluir la consideración de los intereses de las personas.

© Ediciones Morata, S. L.

Scanned by CamScanner

Una postura intermedia entre ambos extremos consiste en que quienes comparten el poder sean capaces de desarrollar argumentos y deliberaciones basados en principios que ellos mismos consideren como objetos de decisiones (BARRY, 1973). Pero, si se excluye del poder a otros grupos con intereses característicos, también es probable que se excluyan sus intereses del ámbito de los principios utilizados por los grupos con poder. Para que las relaciones de poder no determinen los resultados, es preciso equiparar la participación en el poder.

En otras palabras, los participantes pueden representar intereses sin que ello impida discutir una decisión racional sobre la base de principios, habida cuenta de la existencia de cierto equilibrio de poder entre los grupos o individuos participantes. Al mismo tiempo, los intereses de los excluidos no estarán bien representados. Cuanto mayor sea la igualdad de poderes a la que se llegue, mayor será la igualdad de consideración otorgada a los intereses (BARRY, 1973).

Propongo que esta postura modificada se utilice como norma para la deliberación sobre los resultados de la evaluación, en vez de una postura pura en la que se dé por supuesto que los deliberantes sean iguales en poder y carezcan de intereses, es decir, sean "desinteresados". Propongo que el evaluador prevea una posterior deliberación sobre los resultados de la evaluación en la que las cuestiones se decidan mediante la discusión racional sobre principios comunes, aunque no determinados por las relaciones de poder.

Será entonces obligación del evaluador reflejar *todos* los intereses pertinentes en la evaluación. En general, los intereses peor representados serán los de los menos poderosos. Para que el evaluador no favorezca los intereses dominantes del poder de manera parcial, aunque sea inconsciente, debe asegurarse de que todos los intereses estén reflejados en la evaluación. La postura modificada, de imparcialidad y no de desinterés, permite también que las minorías especialmente afectadas cuenten más en un programa o política, con mayor influencia en la decisión, dado que dichas minorías podrán representar sus intereses tanto en la evaluación como en la deliberación. Esto reviste particular importancia porque muchos programas y políticas afectan a las personas de formas muy diferentes. El proceso de deliberación pura no permite estas especiales consideraciones, dado que no da lugar a la representación de intereses: todos los participantes son desinteresados.

Por tanto, el evaluador debe defender una evaluación en la que se reflejen todos los intereses y prever una deliberación sobre los resultados de la evaluación en la que todos los intereses se hallen representados y no esté determinada por las relaciones de poder. Una evaluación desinteresada, no comprometida, tiene la triste cualidad de carecer de relieve al no representar los intereses de los más afectados. Lo que en realidad desean las personas es una evaluación imparcial, en la que se reflejen los intereses de grupo, en la que los grupos reciban un tratamiento equitativo y en la que el evaluador no favorezca en exceso sus propios intereses (véase el Capítulo VIII). En este sentido, como fundamento de la evaluación, la objetividad es insuficiente.

© Ediciones Morata, S. L.

Ahora bien, es obvio que las relaciones de poder existen y desempeñan un papel en todas las decisiones. El evaluador no puede esperar que su evaluación determine completamente las decisiones concretas. Hay otros muchos factores legítimos que surgen en la posterior situación de deliberación. Quienes han de emprender las acciones deben sopesar y equilibrar las consecuencias a la luz de sus propios intereses. El papel de la evaluación se limita a informar y persuadir a los participantes del estado de la cuestión.

El evaluador debe prever una situación de deliberación que no esté determinada por el poder. ¿Qué ocurre si la evaluación suministra informes imparciales que reflejan todos los intereses, y algunos de éstos se pasan por alto reiteradamente? ¿Y si las decisiones se toman, por regla general, sobre la base del poder y no sobre la deliberación de consenso? En tal caso, las críticas radicales del liberalismo están en lo cierto: en realidad, el poder decide las cuestiones y la evaluación es puro ornamento.

Determinación colectiva de las opciones

En la idea de deliberación, quizá esté implícita la de determinación colectiva de las opciones. La citada determinación es posible o al menos imaginable sin recurrir a la deliberación. Podemos imaginar a cada individuo haciéndose su propia idea y votando sobre una cuestión sin entablar diálogo alguno con sus compañeros. La simple suma de decisiones individuales decide la cuestión. No tiene por qué producirse una deliberación. Sin embargo, este procedimiento encierra un defecto fundamental: en realidad, es antinatural. Para justificar algo, si no para persuadir de algo, es necesario el diálogo.

En la teoría del liberalismo no hay nada que impida el desarrollo de un procedimiento sin que medie la deliberación. La suma de preferencias individuales y de intereses privados no exige el diálogo. Los intereses privados pueden expresarse a través de mecanismos de mercado. Sin embargo, llega un momento en que parece necesario descubrir el interés común mediante la determinación colectiva. Quizá el defecto fundamental del liberalismo sea que los individuos no son, en realidad, átomos independientes que sólo, o de forma primordial, cooperan entre sí por el propio interés de cada uno. Son miembros y productos de la sociedad en la que viven. Siempre están incluidos en una comunidad e implicados en relaciones recíprocas.

Desde el punto de vista metodológico, el liberalismo es individualista en la medida en que reduce todos los valores a estados de conciencia de uno o más individuos independientes. Los valores son o bien simples, formando parte del estado de conciencia de una persona, o bien compuestos, implicando estados independientes de conciencia de diversas personas. No hay lugar para valores recíprocos, que supongan una reciprocidad de conciencias entre dos o más personas. Los valores recíprocos están implicados en la

© Ediciones Morata, S. L.

Scanned by CamScanner

opción colectiva, en el mantenimiento de una comunidad de personas y en la participación en la misma.

WOLFF (1968) denomina "comunidad racional" a esta deliberación colectiva sobre objetivos. Es una actividad en la que diversos agentes racionales, iguales desde el punto de vista político, se reúnen para deliberar con el fin de ponerse de acuerdo sobre objetivos y acciones colectivas. Esto se logra mediante el diálogo. La comunidad constituye la metáfora orientadora subyacente.

Todo diálogo supone la reciprocidad de conciencias entre los participantes, pero la reciprocidad puede darse entre el amo y su esclavo o entre un jefe y sus súbditos. Estas relaciones son recíprocas, pero desiguales. La consecución de la comunidad racional no sólo requiere el diálogo, sino que éste sea entre iguales. El diálogo libre exige la réplica sin cortapisas. Las relaciones de poder deforman el diálogo. Cada participante en una comunidad racional debe reconocer a sus compañeros como agentes morales y sus derechos a la igualdad recíproca en el diálogo (WOLFF, 1968).

La comunidad racional requiere la reciprocidad de conciencia de participación. Cuando las personas se reúnen y experimentan una conciencia recíproca, tienen la sensación de comunidad, lo que contrasta con la clásica visión liberal de los individuos que sólo cooperan de forma instrumental en beneficio de sus propios fines. La comunidad y la reciprocidad no son valores fundamentales del liberalismo. Éste trata de establecer obligaciones recíprocas por medio de la idea del contrato social.

Esta concepción de deliberación para la acción es similar a otras, como la de HABERMAS (1973), que piensa en una deliberación para la acción en la que quienes emprenden la acción participan en un diálogo práctico, conscientes de sus intereses comunes y conocedores de las circunstancias y posibles consecuencias de su acción. El objetivo de ese diálogo es ponerse de acuerdo sobre lo que deben hacer.

La teoría o el conocimiento científico entra en ese diálogo práctico, pero nunca es suficiente para determinar lo que deban hacer los participantes. Sólo ellos pueden sopesar los riesgos y consecuencias de su acción. Con palabras de HABERMAS, ni la "teoría" ni la "ilustración" deben justificar la acción. Esto sólo puede hacerse mediante el diálogo práctico.

Del mismo modo, la concepción de la deliberación que presentamos aquí contempla a los destinatarios de la evaluación informados por los resultados de la misma, pero no obligados por ella ni por las recomendaciones del evaluador, en caso de que las haya. Quienes han de actuar sobre la base de los resultados deben deliberar juntos sobre lo que haya de hacerse. Las relaciones de poder están excluidas de esta deliberación, pero los participantes son conscientes de sus intereses. El evaluador debe prever estas circunstancias. No debe prever que los participantes actúen sobre la base exclusiva de sus resultados sin deliberación ni la intervención de una figura de poder que utilice los resultados en su beneficio. Los resultados han de utilizarse en el diálogo práctico y el evaluador desempeña su labor lo mejor posible para lograrlo.

© Ediciones Morata, S. L.

El fundamento moral de la evaluación

En resumen, propongo cuatro valores como fundamento moral de la evaluación: igualdad moral, autonomía moral, imparcialidad y reciprocidad. La noción fundamental de igualdad consiste en que hay que considerar a todas las personas como miembros del mismo grupo de referencia y, en consecuencia, han de ser tratadas por igual.

Primero, la igualdad moral indica que cualquier persona tiene el mismo derecho para procurar la satisfacción de sus propios intereses.

Segundo, la autonomía moral supone que nadie debe imponer su voluntad a los demás mediante la fuerza, la coerción u otros medios ilegítimos. Y a nadie debe imponérsele nada en contra de su voluntad. Este valor está íntimamente ligado al concepto de elección, un principio liberal básico. Podemos suponer que no se ha violado la autonomía moral en la medida en que el individuo decide lo que ha de hacer (aunque algunos consideran insuficiente este criterio). La autonomía moral que se manifiesta en la opción se fundamenta en el individualismo metodológico del liberalismo, en la primacía del individuo y en la tradición del contrato social.

Tercero, los conflictos entre pretensiones e intereses han de zanjarse con *imparcialidad,* o sea, estando representados todos los intereses y sin que ningún procedimiento de decisión que se emplee favorezca a ninguno. La imparcialidad, propiamente dicha, es un valor moral. A veces, sobre todo en la evaluación, se confunde imparcialidad con objetividad, siendo responsables de esta confusión las presiones a favor de la justificación de la evaluación. La objetividad es más fácil de demostrar y puede ser una forma de indicar imparcialidad, pero un procedimiento objetivo puede ser muy parcial respecto a un interés particular. Es decir, podemos tener un procedimiento objetivo (reproducible) que muestre un pronunciado sesgo. Un resultado reproducible no basta para demostrar la imparcialidad.

El *cuarto* valor que propongo para constituir el fundamento moral de la evaluación en una democracia liberal no se relaciona tan íntima ni necesariamente con el liberalismo como los tres primeros. Quizá en ello resida la mayor deficiencia moral del liberalismo. Se trata del valor de la reciprocidad, del que depende el sentido de comunidad. Podemos afirmar que tanto la reciprocidad como la comunidad están y deben estar implícitas en las cosas humanas.

El liberalismo considera la sociedad como una colección de individuos interesados, en ellos mismos, cada uno de los cuales busca cumplir sus propios fines. Entre ellos no existen relaciones necesarias ni obligatorias, sino aquellas voluntarias que se establezcan con fines instrumentales. Las disposiciones contractuales sirven para constituir el fundamento de las obligaciones mutuas. Sobre la base de la igualdad, la autonomía y la imparcialidad, cada persona puede plantear sus propios intereses, sin imposiciones sobre los demás y pudiendo unirse a otros para zanjar de forma imparcial los conflictos de intereses. Pero, ¿qué les ocurre a los perdedores en tales procedimientos? Nos sentimos intuitivamente incómodos. Los tres primeros

© Ediciones Morata, S. L.

Scanned by CamScanner

valores permiten que cada uno mantenga una actitud completamente desinteresada y despreocupada. Cada cual puede plantear sus propios intereses para conseguir un reparto imparcial sin otra responsabilidad que la derivada del cumplimiento de los procedimientos acordados. Un trato es un trato. Los ganadores no tienen responsabilidad alguna respecto a los perdedores.

Nuestra intuición moral se rebela contra esta concepción. De hecho, con frecuencia en las democracias liberales las personas tienen una fuerte sensación de incomodidad respecto a los perdedores sociales. Su presencia resulta molesta. Una forma de sobrellevar tal sentimiento de culpabilidad consiste en el altruismo, la tendencia a ir más allá de la obligación. Esto se refleja en las importantes donaciones voluntarias a obras de caridad que tienen lugar en los países más liberales, como Estados Unidos. Aunque el liberalismo dispone ciertas prestaciones mínimas y timoratas para los perdedores, las personas sienten a menudo de manera intuitiva que son de algún modo responsables. Instintivamente, la mayoría percibe que es miembro de una comunidad. Del correcto cumplimiento de las reglas sólo puede surgir un sentido muy débil de comunidad.

La idea de reciprocidad es antigua y se basa en la ética humanitaria: "Haz a los otros..." La ética humanista exige que no haya posiciones privilegiadas entre las personas. La reciprocidad indica que las relaciones entre las personas son simétricas, que tratamos a los otros como querríamos ser tratados. En parte, esto refleja la negativa a aprovecharse de los demás, aunque se tenga la oportunidad de hacerlo. Mientras que la igualdad sólo nos permite participar en la carrera hacia la satisfacción de nuestros propios intereses en condiciones equitativas, aun suponiendo que la vida consista en esa competición, la reciprocidad hace a los ganadores parcialmente responsables de los perdedores. Hay un equilibrio recíproco entre nuestras peticiones y las de los demás, entre el interés egoísta y el altruismo.

A menudo, los argumentos basados en el principio de reciprocidad se derivan de trasposiciones de puntos de vista. A veces, esos argumentos adoptan la perspectiva de un tercero imparcial (p. ej., un evaluador), con el fin de eliminar consideraciones como la categoría social y la riqueza, que podrían ocultar las relaciones simétricas (PERELMAN y OLBRECHTS-TYTECA, 1969). Hacia esa relación deben orientar sus esfuerzos los evaluadores.

> Por tanto, respetar a otro como persona moral es tratar de comprender sus aspiraciones e intereses desde su punto de vista y proponerle consideraciones que le permitan aceptar las limitaciones de su conducta.
>
> (RAWLS, 1971, pág. 338).

La concepción subyacente de justicia

En su forma más abstracta, la regla de justicia "requiere otorgar idéntico tratamiento a los seres o situaciones de la misma clase" (PERELMAN y OLBRECHTS-TYTECA, 1969, pág. 218). Esto impone un requisito de justicia para la coherencia de la acción, pero no especifica cuáles son las categorías de las personas, tratamientos y situaciones. El concepto de justicia debe manifestarse en la concepción de estas relaciones.

© Ediciones Morata, S. L.

Entre los antiguos, PLATÓN pensaba que la justicia consistía en "dar a cada uno lo debido" (DIGGS, 1974). La sociedad estaba formada por clases de personas (campesinos, artesanos, guerreros, gobernantes, etc.) y todos estos elementos debían estar en armonía. Cada uno debía recibir "lo debido", del mismo modo que los elementos estaban en armonía dentro de los individuos. Por su parte, ARISTÓTELES pensaba que una constitución era justa cuando favorecía el interés común y no los intereses de los gobernantes.

Sin embargo, las concepciones modernas de la justicia se iniciaron con el liberalismo de HOBBES. A diferencia de los filósofos antiguos y cristianos, HOBBES concebía a los seres humanos como animales egocéntricos que utilizaban su inteligencia para satisfacer sus propios fines egoístas. Influido por la ciencia natural de GALILEO, HOBBES creía que la conducta humana podía explicarse mediante principios mecánicos, de manera que cada persona se orientaría de forma mecánica hacia la búsqueda de su propia satisfacción (DIGGS, 1974). HOBBES elaboró una epistemología mecanicista de los sentidos para explicar esa conducta.

Dado que la sociedad era una colección de individuos interesados, era necesario un soberano fuerte que pusiera coto a tan potentes tendencias. Cada individuo vería como una "ley natural" que todos se unieran para proteger sus vidas. Esto se debería a los intereses individuales de cada cual. El interés común sería la suma de los individuales. El Estado era un instrumento para la satisfacción egoísta de cada uno. La justicia consistía en obedecer las leyes del soberano. La perspectiva de HOBBES constituye una completa expresión del individualismo metodológico. El individuo era anterior a la sociedad.

Siguiendo a HOBBES, LOCKE reinterpretó la "ley natural" de manera que incluyera la libertad de cada persona para hacer lo que quisiera, en la medida en que no estuviera específicamente prohibido por la ley y en tanto no afectara a la vida o la voluntad de otra persona. Llegaron a conocerse estas libertades como "libertades naturales". Cada persona tenía un derecho moral igual a la libertad, derecho que se estructuraba de forma negativa como ausencia de sometimiento a la voluntad de otros. Las únicas obligaciones morales se referían a la ley natural y a la razón. Incluso las leyes civiles tenían que ser morales. El gobierno se derivaba del consentimiento de los gobernados. El Estado mismo existía para proteger las libertades. La injusticia consistía en la infracción de las libertades naturales. LOCKE transformó el individualismo basado en el interés de HOBBES, en la autonomía individual.

HUME trató de derivar la moralidad de la observación y la experiencia, pero concluyó que era imposible. Estableció la distinción primordial entre hechos y valores. Como la moralidad era práctica, debía incluir la motivación y derivarse de los valores, mientras que la razón sólo puede determinar

© Ediciones Morata, S. L.

Scanned by CamScanner

hechos a partir de observaciones. Según HUME, la justicia consistía en vivir de acuerdo con convenciones sociales que sólo eran convenciones, aunque promovían los intereses privados y comunes.

Por otra parte ROUSSEAU amplió en gran medida la idea de autonomía individual de LOCKE (DIGGS, 1974). La legitimidad moral no puede derivarse de la fuerza ni de la coerción. Las personas son lo que su gobierno las hace y la libertad moral consiste en seguir las leyes que uno se prescribe a sí mismo. La libertad consiste en la participación, en hacer y cumplir realmente las leyes y no sólo en ser libre respecto de la voluntad de otro. La participación de todos los ciudadanos se traduce en la formación de la "voluntad general", que expresa el interés común. Cuando los intereses especiales entran en conflicto, el común prima sobre ellos. La peor injusticia es la violación de la propia autonomía. El punto de vista de ROUSSEAU constituye la máxima expresión de la autonomía moral.

Todas estas teorías de la justicia exigían que el bien común primara frente a los intereses especiales (DIGGS, 1974). En general, el bien común significaba el conjunto de beneficios compartidos, frente a los que sólo favorecen a un individuo. El objetivo del gobierno consistía en el bien común y el gobierno que sólo favoreciera intereses especiales sería injusto (DIGGS, 1974).

En la idea de servicio al bien común estaba implícita la de imparcialidad. En concreto, la justicia exigía la administración imparcial del derecho. Nadie debía resultar beneficiado. Todas las teorías requerían la imparcialidad (acatar las leyes e instituciones justas existentes). Todas las personas debían recibir el mismo trato cuando no existieran diferencias de relieve. Así, los valores de igualdad, autonomía e imparcialidad subyacían a estas teorías liberales de la justicia.

Sin embargo, DIGGS (1974) pone en evidencia la tendencia conservadora de estas teorías. ARISTÓTELES había hecho la distinción entre justicia distributiva y justicia conmutativa. La primera se refería a la distribución de la riqueza y de otros bienes entre las personas. La justicia conmutativa se refería a las transacciones entre las personas, en especial a la rectificación del mal trato dispensado por una persona a *otra*, como en el caso de una compraventa o de un robo. HOBBES rechazó por completo la justicia distributiva, de modo que se identificó con el cumplimiento de los compromisos, con la justicia conmutativa. Los filósofos que continuaron esta tradición siguieron haciendo caso omiso de la justicia distributiva.

En el siglo XIX, ante la creciente acumulación de riquezas industriales, las consideraciones relativas a la igualdad de derechos parecían inadecuadas. La justicia distributiva pasó a primer plano. En cuanto teoría de la justicia, el cumplimiento de la ley o de las convenciones resultaba conservadora. Se hizo más apremiante prestar atención a la promoción de los intereses de todos.

Estas teorías se desarrollaron en dos direcciones; una según la cual el Estado estaba llamado a poner coto a los intereses dominantes y, otra, en la que todos los intereses tenían que ser escuchados (DIGGS, 1974). En contraste con las liberales, que asignaban al Estado un papel mínimo en la

© Ediciones Morata, S. L.

sociedad, las teorías organicistas, sobre todo la de HEGEL, seguido por MARX, consideraban la sociedad como un organismo natural y las instituciones como encarnaciones objetivadas de la razón humana. Era imposible definir la naturaleza humana con independencia de la sociedad en la que vivía el individuo, como había hecho HOBBES. El orden social configuraba su vida. Se concebía la sociedad como un organismo del que formaba parte el individuo. Debía juzgarse el orden social según su contribución a la libertad, es decir, a la autodeterminación natural. A MARX, por su parte, no le entusiasmaban las consideraciones de la justicia distributiva (DIGGS, 1974), sino que tendía a equiparar la justicia con el bien común. Pretendía un orden social en el que las personas fueran libres para desarrollar sus potencialidades.

Estos teóricos reafirmaron los lazos entre los humanos y su comunidad, así como el papel del Estado en la configuración de los individuos. Podemos descubrir aquí las ideas de comunidad y reciprocidad. Más adelante, estas ideas sirvieron de fundamento a las críticas radicales de las ideas liberales. El orden social configuraba las pretensiones e intereses de los individuos. En realidad, el Estado tenía la responsabilidad de realizarlo, de hacer mejores a los hombres.

Por otra parte, los utilitaristas, como BENTHAM y MILL, continuaron la tradición liberal del individualismo metodológico. Siguiendo a HUME, pensaban que las instituciones eran meras costumbres y la encarnación de la sinrazón. Debían someterse a la razón. Como la conducta racional del individuo consistía en seleccionar la acción que promoviera la consecución de los objetivos individuales de modo más eficiente, la conducta racional de la sociedad consistía en seleccionar la acción que mejor promoviera la consecución de los objetivos sociales (DIGGS, 1974). El individualismo metodológico de los utilitaristas les llevó a equiparar la decisión colectiva con la individual.

Basándose en la presunta motivación del interés egoísta para conseguir el placer y evitar el dolor, los utilitaristas concluyeron que el objetivo social consistía en proporcionar el máximo placer o felicidad para el mayor número de personas. Toda acción que cumpliera este objetivo era justa. La justicia consistía en elevar al máximo la felicidad o la satisfacción de las pretensiones. Esta forma de pensar se convirtió en la filosofía social liberal predominante en el siglo XX.

La concepción liberal más reciente de la justicia es la de RAWLS (1971). Utilizó el instrumento del contrato social para preguntarse en qué principios de justicia estarían de acuerdo personas libres, iguales y racionales, en condiciones de un contrato equitativo respecto a sus propios intereses. Desde el punto de vista teorético, toda persona tendría libertad para presentar sus peticiones y para aceptar o rechazar los principios, o sea, tendría autonomía. Es más, como en la "posición original" de RAWLS nadie conoce con exactitud cuál es su situación social, nadie puede inclinar los resultados a favor de sus intereses. Todas las personas ocupan una posición negociadora semejante.

© Ediciones Morata, S. L.

Scanned by CamScanner

A partir de esta situación hipotética, RAWLS deriva lógicamente dos principios de justicia en los que todos estamos de acuerdo. El primer principio garantiza unos derechos y libertades iguales para todos, en el sentido lockeano. El segundo garantiza la justa igualdad de oportunidades. En un sentido más original, el segundo principio establece que las desigualdades sólo pueden existir si benefician a las personas menos favorecidas de la sociedad. Me parece que el "principio de diferencia" constituye un intento de introducción del valor de la reciprocidad en la estructura social, una idea que no aparece en el utilitarismo.

En efecto, el segundo principio hace, en parte, responsables de los miembros menos favorecidos de la sociedad al resto de los ciudadanos, algo que el utilitarismo sólo hace en forma, un tanto gratuita, de altruismo. La postura de RAWLS es individualista, desde el punto de vista metodológico, en la medida en que deriva sus principios de justicia de deducciones *a priori* respecto de lo que desearían los individuos. A partir de consideraciones individuales, se extraen conclusiones sobre el bien colectivo. Como en el caso de HOBBES, los individuos interesados constituyen el punto de partida.

La teoría de RAWLS ha recibido muchas críticas, pero una de las más interesantes consiste en que el bien colectivo no puede derivarse de ese modo a partir de las pretensiones individuales. Con el fin de decidir qué principios deben aplicarse a la sociedad, debemos considerarla en su totalidad. El bien individual no equivale al bien colectivo (BARRY, 1974). Por ejemplo, todos pueden querer un coche, pero eso no significa que una sociedad con coches sea la mejor. Tenemos que considerarla en su totalidad y no sólo las pretensiones individuales.

La teoría de RAWLS constituye un perfeccionamiento del utilitarismo en la medida en que es característicamente más igualitaria. Presenta el valor de la reciprocidad, aunque de forma un tanto indirecta. No obstante, plantea muchos problemas lógicos y tiende al paternalismo. Ninguna teoría de la justicia parece totalmente adecuada como fundamento de la evaluación.

Una teoría adecuada de la justicia tendría en cuenta los valores de igualdad moral, autonomía moral, imparcialidad y reciprocidad, aunque sea difícil decir cómo habrían de equilibrarse entre sí. RAWLS (1971) denomina teoría "pluralista/intuicionista" de la justicia a la que cuenta con diversos principios en competencia, sin establecer un orden de prioridad entre ellos. Hay una familia irreductible de primeros principios que han de sopesarse entre sí sin reglas explícitas para hacerlo. La acción de sopesar es intuitiva, aunque podamos describir cómo la hayamos efectuado en una situación concreta.

En la concepción de justicia que propongo como fundamento moral de la evaluación, los cuatro valores, o principios basados en los valores (como la elección), sirven como consideraciones primarias que deben sopesarse y equilibrarse entre sí en una situación determinada. Sólo cuando entran en conflicto en una situación concreta es posible decidir que uno tenga mayor importancia que otro (BARRY, 1965). Ninguno goza de prioridad absoluta

© Ediciones Morata, S. L.

sobre los demás. Pueden primarse unos en detrimento de los otros. No existe entre ellos ningún principio indiscutible. El más importante sólo puede determinarse mediante opción en los casos en los que no entran en conflicto mutuo (BARRY, 1965).

He de admitir que el presente resumen de lo que debería incluir una teoría subyacente de la justicia es incompleto e impreciso. En una teoría completa deberían aparecer otros valores reconocidos con mayor generalidad, como la eficiencia, pero los valores morales son los que con mayor frecuencia se pasan por alto. Para que la evaluación de los programas públicos sea válida, debe contribuir a la justicia en la sociedad en general.

© Ediciones Morata, S. L.

Scanned by CamScanner

CUARTA PARTE

Metaevaluación

Pero también debemos reconocer que los seres humanos son capaces de tomar conciencia de las interpretaciones, evaluaciones y normas que aceptan tácitamente y pueden someterlas a la crítica racional.

Richard J. BERNSTEIN.
The Restructuring of Social and Political Theory, 1978, pág. 236.

Scanned by CamScanner

CAPÍTULO X

La objetividad, equidad y justicia de la política federal de evaluación

La política federal de evaluación

En 1965, el senador Robert Kennedy hizo hincapié en que se añadiese al Título I (educación compensatoria) de la *Elementary and Secondary Education Act* una cláusula adicional relativa a la evaluación. Pensaba que las escuelas eran, en parte, culpables del bajo nivel de rendimiento de los alumnos, y que la nueva financiación federal no se utilizaba con provecho. El propósito que perseguía la obligación de elaborar informes era dar noticia a los padres, sobre todo a los padres pobres, de cómo trabajaban sus escuelas. Los educadores serían responsables ante sus usuarios. Más adelante, la evaluación obligatoria se extendió a casi todos los programas sociales federales.

En el mismo año, el presidente Johnson introdujo en el gobierno federal el *Planning, Programming and Budgeting System* (PPBS), ya entonces utilizado de forma generalizada en el Pentágono. El secretario del departamento de Salud, Educación y Bienestar, John Gardner, creó un nuevo organismo denominado *Assistant Secretary for Program Evaluation* (ASPE). Al frente de dicho organismo puso a William Gorham, que contaba con gran experiencia en el departamento de Defensa. Sus subsecretarios eran Robert Grosse y Alice Rivlin. El objetivo consistía en "elaborar programas que puedan enunciarse, medirse y evaluarse en términos de coste/beneficio". La excelente historia de la evaluación del Título I de McLAUGHLIN (1975) consigna estos hechos.

NOTA DEL AUTOR: Este capítulo fue publicado como: "The Objectivity, Fairness, and Justice of Federal Evaluation Policy as Reflected in the Follow Through Evaluation", *Educational Evaluation and Policy Analysis*, vol. 1, n.º, 1979, págs. 28-42. Copyright 1979: *American Educational Research Association*, Washington, D.C. Reimpreso con autorización. Gene CLASS, Elisabeth HUTCHINS y Rochelle MAYER me hicieron valiosos comentarios al respecto. Este trabajo fue patrocinado por la *Ford Foundation*.

© Ediciones Morata, S. L.

Scanned by CamScanner

Mientras Kennedy estaba preocupado por la responsividad de las escuelas hacia los pobres, el grupo ASPE pretendía descubrir los enfoques más eficientes para educar a los niños menos favorecidos. Operando desde la perspectiva del "análisis de sistemas", el ASPE trataría de identificar los programas más eficaces de manera que la asignación de recursos resultara más eficiente. El enfoque del ASPE suponía que podía definirse una "función de producción" de la educación, o sea, definir de manera precisa cómo un conjunto de entradas, como dinero o determinados programas, estaría relacionado con un conjunto de resultados, como las puntuaciones de tests. Definida la función, podría decidirse que sólo se utilizaran los enfoques educativos más eficientes.

Los primeros años de la evaluación según el Título I no satisficieron al ASPE. Las evaluaciones se llevaron a cabo a nivel local y los resultados fueron positivos casi de un modo uniforme. Es más, a falta de una medida común de resultados para todos los programas, era imposible efectuar un análisis coste/beneficio para determinar qué programas eran más eficientes. En parte con este fin y en parte para contrarrestar los efectos negativos del informe Coleman, RIVLIN y WHOLEY, del ASPE, prepararon el estudio TEMPO, llevado a cabo por una división de *General Electric*. El objetivo era averiguar la cantidad de recursos que producían un determinado incremento de las puntuaciones en un test concreto. Los resultados fueron desalentadores. El estudio no pudo descubrir incremento alguno en las puntuaciones. Aún peor, Moynihan, a la sazón en el gabinete de Nixon, utilizó los resultados del estudio para oponerse a mayores incrementos en la ayuda a las escuelas.

¿Dónde estaba el error? El personal del ASPE sostenía que su enfoque era absolutamente correcto, pero los datos facilitados por los distritos escolares no eran fiables. Optaron por centralizar el control de las actividades de evaluación. Así recogerían sus propios datos para asegurar la fiabilidad. A consecuencia de ello, se iniciaron las encuestas nacionales con arreglo al Título I. Además, decidieron transformar el programa *Follow Through* que, en principio, era de servicios y continuaba la educación de los niños acogidos al programa *Head Start*, convirtiéndolo en un experimento con variaciones planificadas.

Las encuestas nacionales se llevaron a cabo en 1968 y 1969. Los resultados de la correspondiente a 1968 fueron tan decepcionantes como el estudio TEMPO: no se observaron incrementos en las puntuaciones de los tests. Los resultados de la encuesta de 1969 no llegaron a publicarse. Todo ello le llevó a decir al secretario Finch, del departamento de Salud, Educación y Bienestar, que los educadores no sabían cómo educar. El presidente Nixon utilizó los resultados como pretexto para impedir el incremento de la financiación correspondiente al Título I.

El grupo del ASPE estaba decepcionado, pero no dispuesto a abandonar. Alice RIVLIN sostenía que el problema consistía en que los servicios sociales no estaban organizados de forma adecuada para responder a las cuestiones planteadas por el ASPE relativas a la eficacia y la eficiencia. Los programas

© Ediciones Morata, S. L.

de servicios sociales debían organizarse de manera que pudieran responder a las cuestiones del evaluador. La solución de RIVLIN consistía en que el gobierno controlara más las variables de iniciales del programa y su diseño, así como la evaluación. Los programas sociales tenían que plantearse como experimentos con variaciones planificadas.

El programa *Follow Through* se organizó de manera que pudiese responder a las cuestiones del ASPE, a modo de experimento masivo. Las clases del *Follow Through* tenían grupos de control con los que se compararían. Se introdujo la idea de las variaciones planificadas. Si los distritos locales no podían diseñar programas lo bastante innovadores como para obtener puntuaciones más elevadas en los tests, se utilizarían programas especiales para introducir las variaciones planificadas en la educación. No se consideraba que los cambios naturales fuesen lo bastante adecuados. En el programa *Follow Through* se acudió a "colaboradores" con modelos especiales de práctica claramente distintos de los de las escuelas públicas para implementar nuevos programas. Las variaciones naturales fueron sustituidas por las planificadas.

Aunque, según RIVLIN, a los efectos del ASPE, el programa *Follow Through* no era perfecto, sí era el único que se acercaba a los requisitos propuestos por el ASPE. "A pesar de sus defectos de diseño experimental, el *Follow Through* es significativo como la primera tentativa importante del gobierno federal para ensayar distintos enfoques educativos de manera razonablemente sistemática" (RIVLIN, 1971).

La política federal de evaluación, tal como aparece en *Systematic Thinking for Social Action*, de RIVLIN, adoptó la siguiente forma: en los servicios sociales, las decisiones clave deben tomarse en los niveles superiores del gobierno federal. Los servicios sociales no estaban organizados de modo eficiente y no se conocía bien lo que producían dichos servicios. El único conocimiento auténtico que pueda tenerse de ellos es una función de producción que especifique las relaciones estables entre las variables iniciales y los resultados, y el único medio para adquirir ese conocimiento consiste en los métodos experimentales y las técnicas estadísticas. "Es imposible obtener de otro modo la información necesaria para mejorar la eficacia de los servicios sociales" (RIVLIN, 1971, pág. 108). Utilizando esas técnicas, es posible ponerse de acuerdo sobre los objetivos y sobre algunas medidas de resultados, como las puntuaciones de los tests. El fin de la evaluación consiste en la eficiencia de la producción de servicios sociales.

Según la filosofía de RIVLIN, existe un paralelismo entre la producción de servicios sociales y la de productos manufacturados. Se aplican técnicas semejantes de análisis y los programas sociales deben organizarse para proporcionar datos para ese análisis. Los analistas federales descubrirían qué era lo más correcto. No era necesario tener experiencia en los programas ni contacto con los mismos para poder evaluarlos. Es más, poner en marcha este enfoque de evaluación era cuestión de dirección estricta y de dominio del proceso. Sólo un gobierno muy centralizado podría llevar a cabo esta tarea.

© Ediciones Morata, S. L.

Scanned by CamScanner

Esta filosofía de la evaluación predominó en el gobierno federal de los Estados Unidos durante la pasada década. Para descubrir cómo operó en la práctica, podemos examinar la evaluación *Follow Through.*

Los comienzos de "Follow Through"

Follow Through comenzó como un programa para proporcionar servicios educativos a los alumnos menos favorecidos en los primeros años de escolarización. En 1967, a causa de dificultades financieras, el programa sólo recibió quince millones de dólares, en vez de los ciento cincuenta millones previstos. En el departamento de Salud, Educación y Bienestar, el grupo del ASPE vió ahí la oportunidad para convertir el *Follow Through* en un experimento de variaciones planificadas para descubrir "qué funcionaba mejor". HANEY (1977) y ELMORE (1976) han documentado estos hechos.

En 1967, William Gorham, a la sazón subsecretario de Planificación y Evaluación (ASPE) del Departamento de Salud, Educación y Bienestar, organizó la *White House Task Force on Child Development* (Grupo de trabajo sobre desarrollo infantil de la Casa Blanca). El grupo de trabajo estaba constituido por miembros del grupo de Gorham, del que formaba parte Alice RIVLIN (ELMORE, 1976, pág. 159). Este grupo llevó a cabo su tarea en un ambiente de rígido control presupuestario impuesto por la Administración. Por ello, el máximo interés se centró en la investigación y el desarrollo, más baratos que los costosos programas de servicios.

El grupo de trabajo hizo veinte recomendaciones. Una abogaba por la utilización de fondos federales para descubrir cómo hacer más eficaz la educación. Otra sugería que el *Follow Through* se utilizase para evaluar las variaciones más destacadas que se registraran en la educación compensatoria. El informe del grupo de trabajo concluía que "el único hecho patente... es nuestra falta de conocimiento de formas eficaces de elevar el aprovechamiento educativo de los estudiantes" (ELMORE, 1976). El *Follow Through* serviría para solucionar ese problema de desconocimiento.

Con la plena cooperación de la administración del *Follow Through,* la Oficina de Presupuestos y el ASPE, el *Follow Through* se convirtió en un experimento de variaciones planificadas. Sin embargo, ante el Congreso y los lugares en los que ya estaba funcionando, el *Follow Through* siguió definiéndose como acción comunitaria y programa de servicios sociales, así como ante los padres de los alumnos.

Para quienes elaboraban los modelos de buena práctica durante la primera infancia, los "colaboradores", el *Follow Through* constituía un programa de desarrollo. Uno de los colaboradores originales se quejaba de que la idea general de las variaciones planificadas "mistificaba" a los mismos patrocinadores. "La reunión de los futuros colaboradores, celebrada en Washington para presentar sus diversas orientaciones sobre el *curriculum,* la idea de que la evaluación podría estar a cargo de un tercer organismo... y los supuestos de que cada patrocinador dispusiera de un conjunto completo

© Ediciones Morata, S. L.

que presentar, provocaron una sensación de desconcierto y hasta de locura" (WEIKART y BANET, 1975). Pocos eran, si es que existía alguno, los modelos relativos a la primera infancia completamente elaborados y dispuestos para su implementación, premisa de la idea de las variaciones planificadas. Por necesidad, para quienes fueron seleccionados como colaboradores, el programa se transformó en uno de desarrollo.

Durante los años siguientes, fueron añadiéndose, de forma un tanto oportunista, nuevos colaboradores y nuevos lugares para proporcionar respaldo político al frágil programa. Los alcaldes de las grandes ciudades querían que se sumaran nuevos centros de acción en sus municipios, consiguiéndolo, en consecuencia. El estado de California quería convertirse en patrocinador, por lo que fue designado. Esta estrategia tuvo un éxito admirable para garantizar el respaldo político. A mediados de los setenta, el presupuesto anual del *Follow Through* había ascendido hasta los sesenta millones de dólares. Pero esto llevó consigo importantes desajustes de la correspondencia entre los grupos de comparación que en el diseño experimental se habían escogido para la evaluación. No obstante, esto no preocupaba en exceso a los administradores de la Oficina de Educación, pues creían que la elevación de las puntuaciones de los tests derivada de los modelos de actuación en la primera infancia sería tan grande que el desajuste de los grupos de control no provocaría grandes diferencias (EGBERT, 1977).

La evaluación tuvo un comienzo vacilante con el contrato firmado con la *University of Pittsburgh,* en 1967. Por alguna razón el informe final del grupo de Pittsburgh sólo presentaba las puntuaciones del pretest. La evaluación se reanudó en 1968 con un importante contrato con el *Stanford Research Institute* (SRI). Podemos dividir el transcurso de la evaluación en tres fases, que denominaremos con el apellido del funcionario responsable de la *Office of Education* y el grupo contratante que más caracterizara la evaluación en el período correspondiente. Las tres fases mencionadas son: la EGBERT/SRI, entre 1968 y 1971; la MCDANIELS/Huron Institute, entre 1972 y 1973, y la EVANS/Abt Associates, entre 1974 y 1977.

La fase EGBERT/Stanford Research Institute: 1968-1971

Esta fase de la evaluación fue configurada sobre todo por Robert EGBERT, director del *Follow Through,* Richard SNYDER, director de la evaluación, y el SRI. Se caracterizó por una idea relativamente expansiva de lo que debería ser la evaluación, marcado por intentos, por regla general fracasados, de responder a las críticas sobre diversos aspectos.

En febrero de 1968, en Kansas City, se asignaron noventa y una plazas a los colaboradores, en gran medida de forma voluntaria. En junio, el SRI fue recompensado de manera un tanto apresurada con un gran contrato de evaluación posteriormente criticado por la *General Accounting Office.* En esa época, al menos en teoría, la evaluación abarcaba gran cantidad de aspectos. La relación oficial de criterios de evaluación era muy amplia *(U.S.*

© Ediciones Morata, S. L.

Scanned by CamScanner

Office of Education, 1969), incluyendo comparaciones relativas al desarrollo cognitivo y afectivo, comparaciones entre niños que habían estado en el *Head Start;* grado de participación de los padres; efectos sobre sus actitudes; grado de atención médica, odontológica, psicológica y social; pruebas de cambios en los programas escolares y responsividad ante padres y niños con ingresos reducidos; pruebas de cambios de actitud de los participantes y de su comunidad, y pruebas de la coordinación de los servicios comunitarios.

Los planes del SRI reconocían con toda claridad esta amplia responsabilidad. "El objetivo general de esta evaluación consiste en valorar el impacto del *Follow Through* sobre los alumnos, padres, personal de las escuelas, miembros de la comunidad y en la escuela, como institución social, en los ambientes en los que se han inaugurado los programas *Follow Through*" (*Stanford Research Institute,* 1972, pág. 2). El SRI también consideró como objetivo primordial la descripción de los programas que se iban a evaluar. Sin embargo, al final, el SRI acabó administrando sólo tests cognitivos en cuarenta y siete lugares.

En octubre de 1968, en Atlanta, el congreso anual de *Follow Through* estalló en protestas, sobre todo de los grupos minoritarios. Acusaban a los evaluadores del SRI de tener una visión sesgada del programa y de "ser insensibles a los aspectos relativos a la participación de los padres y comunidades" (HANEY, 1977, pág. 27). La violencia de las protestas sorprendió a los administradores del *Follow Through.* Se financiaron nuevos colaboradores de los grupos minoritarios, convirtiéndose el cambio institucional en un objetivo primordial de la evaluación del SRI.

En consecuencia, el SRI emprendió diversos estudios de casos de comunidades, pero no marcharon bien (KRULEE, 1973, págs. 246-251). El personal dedicado a estos menesteres quedó aislado del resto del SRI, cuya formación básica era de psicología y no entendía bien el enfoque del estudio de casos. Poco a poco, estos estudios se tradujeron a "un nivel más realista" (KRULEE, 1973, pág. 248). Más tarde, se publicaron algunos estudios de casos, pero no influyeron en absoluto en la evaluación global. Según el SRI, cada comunidad parecía diferente, por lo que era difícil efectuar generalizaciones con fines normativos. Hacia 1971 (tras manifestaciones tan fuertes como las del congreso de Atlanta), se abandonó sin grandes alharacas la evaluación del cambio institucional.

A partir de los estudios de comunidades, surgieron, como estudio independiente, las entrevistas con los padres, subcontratadas como actividad autónoma con el *National Opinion Research Center.* Se llevaron a cabo sesenta mil entrevistas. Esta actividad se prolongó durante varios años, pero no tuvo efecto alguno en los resultados finales de la evaluación. Mientras tanto, el SRI continuó con los tests cognitivos.

En abril de 1969, en una reunión celebrada en Pajaro Dunes (California), los mismos colaboradores atacaron los trabajos de evaluación del SRI. Pensaban que la evaluación estaba "gravemente sesgada" por su excesiva confianza en las medidas cognitivas, dado que muchos colaboradores no

© Ediciones Morata, S. L.

consideraban como objetivo primordial la elevación de las puntuaciones en tales medidas. "En concreto, existía la impresión de que las circunstancias estaban, por la confianza depositada en las medidas de cambio cognitivo, a favor de los colaboradores de orientación conductual (o sea, cognitiva) y en contra de todos los demás" (KRULEE, 1973, pág. 219). El resentimiento había ido creciendo con el tiempo e incluso los conductistas estaban de acuerdo. De nuevo, los colaboradores insistieron en los objetivos de cambio institucional. El SRI reconoció públicamente la validez de las críticas y prometió actuar en consecuencia.

Tres meses más tarde, en Palo Alto, los colaboradores volvieron a acusar con virulencia al SRI por su inexplicable lentitud para desarrollar medidas no cognitivas, estudios de comunidades, entrevistas con padres y con profesores, y observaciones en las aulas. El personal del SRI estaba dividido al respecto. Unos pensaban que los estudios de comunidades no eran lo bastante "neutrales" para que se utilizaran como información. No obstante, los colaboradores presionaron al SRI para que actuase en esta dirección.

Como es habitual, la cuestión que suscitaba más pasión era que la evaluación se centraba demasiado en medidas cognitivas convencionales, de las que, según los colaboradores, se desprenderían juicios no muy equitativos respecto a ellos mismos. Las intenciones puestas de manifiesto por el SRI acerca de la ampliación del ámbito de la evaluación, restablecieron hasta cierto punto la calma, aunque ellos todavía mostraron su preocupación por los juicios que pudieran derivarse de las medidas que en ese momento se estaban realizando. En parte como compensación, los colaboradores recibieron algunos fondos para que evaluaran sus propios programas. Esta pauta, consistente en fuertes críticas seguidas de promesas, aunque no de cambios importantes, fue característica en las tres fases de la evaluación.

El SRI trató de elaborar medidas no normalizadas que reflejaran mejor algunos objetivos de los colaboradores. El SRI solicitó de los colaboradores ítems de tests cognitivos, recibiendo 1.500 ítems. La estrategia consistía en elaborar un banco de ítems que reflejara los objetivos comunes a todos los colaboradores. Estos ítems se redujeron a 600, disponiéndose, junto con una batería de tests normalizados, en 22 formas de 8 tests (HANEY, 1977, pág. 182). La administración de los tests planteó muy pronto problemas de logística y costes al SRI. De manera semejante, el SRI trató de elaborar instrumentos no cognitivos pero, efectuadas pruebas de campo, las rechazó por ser demasiado poco fiables. El personal del SRI admitió que no se había percatado de las limitaciones de los tests existentes ni de las dificultades para elaborar otros nuevos (HANEY, 1977, pág. 195). En consecuencia, se abandonaron los esfuerzos para elaborar instrumentos sensibles a los objetivos propios de los colaboradores.

En 1970, los resultados de la primera recogida de datos no mostraron efectos positivos del *Follow Through* y no se publicaron. En mayo, un tribunal de la *Office of Education* revisó la evaluación y criticó al SRI por su fracaso en la comparación de los modelos y en su evaluación respecto a

© Ediciones Morata, S. L.

Scanned by CamScanner

objetivos específicos. Se llevó a cabo una serie de estudios complementarios, ninguno de los cuales produjo efecto alguno en los resultados finales.

En febrero de 1971, la *Office of Education* reafirmó de nuevo sus prioridades para la evaluación. Richard FAIRLEY, Director de Educación Compensatoria, visitó el SRI y se quejó de que los resultados no normalizados del SRI, relativos al aprovechamiento de los alumnos, no pudieran representarse en equivalentes graduados. El plan de FAIRLEY consistía en descubrir a los mejores colaboradores y obligar a utilizar sus modelos en las acciones derivadas del Título I. Este plan requería una medida normalizada.

Durante 1971, la *General Accounting Office,* el departamento de Salud, Educación y Bienestar e, incluso, los *Nader's Raiders** efectuaron investigaciones sobre la evaluación del *Follow Through.* En el verano, tanto EGBERT como SNYDER dimitieron de sus cargos. En noviembre, Garry MCDANIELS fue nombrado director de evaluación del *Follow Through.* Durante los cuatro años de contrato, la *Office of Education* gastó doce millones de dólares en la evaluación.

La fase McDaniels/Huron Institute: 1972-1973

Bajo la dirección de MCDANIELS, la *Office of Education* mantuvo un control más directo de la evaluación. Cuatro meses antes, Laurence LYNN, director del ASPE, había recomendado al Secretario de Salud, Educación y Bienestar, Richardson, anonadado por el coste de la evaluación, que el programa debería volver al rigor experimental. Michael TIMPANE y Joan BISSELL, del ASPE (ELMORE, 1976, págs. 291-292), redactaron la recomendación. MCDANIELS concluyó que había que frenar el desarrollo de nuevos instrumentos. Firmó un contrato en exclusiva con el *Huron Institute,* dirigido por David COHEN y Marshall SMITH. En un informe, SMITH recomienda que MCDANIELS consiga en la *Office of Education* un acuerdo sobre los objetivos a largo plazo de la evaluación y formule un diseño que se ajuste al máximo a los objetivos (HANEY, 1977, pág. 45).

Jane DAVID, del *Huron Institute,* diseñó un plan de muestreo para "elevar al máximo el potencial de comparaciones válidas" entre los colaboradores y entre clases integradas en *Follow Through* y las no integradas en el programa, así como para determinar la eficacia del *Follow Through.* El fundamento del muestreo de DAVID exigía separar los efectos del tratamiento de los colaboradores de los efectos propios de los lugares de aplicación. Esto podría conseguirse incluyendo varios lugares por cada patrocinador. "Si se incluye una serie amplia de proyectos diferentes correspondientes a un mismo patrocinador, tendremos fundamento suficiente para confiar en

* Ralph NADER se hizo famoso por sus acciones en beneficio de los consumidores y usuarios norteamericanos durante los años sesenta y setenta. Sus agentes son los *"Nader's Raiders". (N. del T.)*

© Ediciones Morata, S. L.

que los sesgos restantes no sean muy fuertes". La confusión entre los efectos derivados del modelo y del lugar de aplicación se disiparía. Al final, prevalecería este diseño.

La nueva petición de propuestas de MCDANIELS se refería a una evaluación del impacto sobre una porción restringida de los objetivos originales. Al tener que hacer frente a costes y escándalos cada vez mayores, limitó drásticamente el ámbito de la evaluación, que HANEY (1977) denominó "efecto embudo". El nuevo contrato para el análisis de datos se concedió a *Abt Associates,* quienes estaban seguros de poder emplear nuevas técnicas estadísticas para manejar los desordenados y complejos datos recogidos en condiciones cuestionables. El objetivo básico consistía en determinar "de entre un conjunto de enfoques, cuál y en qué condiciones es más eficaz". La recogida de datos continuó efectuándola el SRI.

El trabajo con los colaboradores para especificar sus objetivos y modelos dejó de formar parte de la evaluación. En la reorganización, MCDANIELS repartió tareas entre distintos grupos. El *Huron Institute* preparó la muestra, el SRI se encargó de recoger los datos, *Abt Associates* los analizó y un tribunal nombrado por la *Office of Education* seleccionó los instrumentos. Aunque MCDANIELS ofreció dinero a los colaboradores para que trabajaran sobre sus propios instrumentos, se le dió la mínima prioridad al desarrollo de los mismos. Esto suavizó considerablemente el golpe para los colaboradores.

MCDANIELS convocó un tribunal de la *Office of Education* para determinar el conjunto de medidas comunes respecto a las cuales se compararían los modelos. Se escogió el *Metropolitan Achievement Test* porque "a efectos de comparación, parecía deseable contar con un único instrumento que pudiera utilizarse en varios niveles" (EPPS, s/f). El MAT presentaba también propiedades psicométricas adecuadas. Se seleccionó el test de matrices progresivas en color de RAVEN porque el MAT era demasiado limitado y porque no resulta fácil preparar un entrenamiento previo para realizar el de RAVEN, por lo que los colaboradores no podrían enseñárselo a sus alumnos. El tribunal expresó sus reservas en relación con ambos tests.

Se manifestaron reservas sobre los dos tests no cognitivos, la *Intellectual Achievement Responsibility Scale* y el *Coopersmith Self-Esteem Inventory.* El tribunal creía que "en realidad, no existen medidas adecuadas en estas áreas" (EPPS, s/f).

De nuevo, pronto volvió a surgir el descontento. En el congreso anual del *Follow Through,* celebrado en Denver, en octubre de 1972, los padres encabezaron las protestas; decían que no habían desempeñado papel real alguno en las decisiones, en la planificación ni en la evaluación, a pesar de las apariencias (HANEY, 1977, págs. 60-61). Insistían en que se encargara a los padres la evaluación anual de los colaboradores. En el documento que presentaron se dice: "Estamos cansados de que otros decidan cuándo un programa es o no 'bueno' para nosotros, basándose en su concepto de 'datos' y en su concepto de lo que es 'erróneo' para nuestros hijos y qué hay que hacer para corregir esos 'errores'". Y continúa: "Los expertos

© Ediciones Morata, S. L.

Scanned by CamScanner

deberían dedicarse a medir el tipo de programa que queremos, en vez de diseñar un programa que saben cómo medir". Pedían que los padres participasen en la interpretación y recogida de datos (*Steering Committee*, 1972).

Los colaboradores redactaron un documento apoyando las reivindicaciones de los padres y mostrando sus simpatías hacia ellas. Aunque el estallido fue semejante al que se produjo en el anterior congreso de Atlanta, éste ocasionó cambios inmediatos, aunque pasajeros. En esta ocasión ni siquiera se trató de estudiar las críticas. HANEY señala que la evaluación había quedado, en realidad, aislada de la política dominante de la época. De hecho, éste fue el último congreso general anual del *Follow Through*, de manera que no pudieron producirse alborotos de este tipo. No obstante, los colaboradores siguieron reuniéndose periódicamente.

En 1973, se publicaron dos estudios auxiliares iniciados con anterioridad. Estudiando las clases, SOAR (1973) descubrió que los modelos no tenían una gran coherencia interna. El estudio de STALLING (1973) fue el único que trató de describir, desde el punto de vista empírico, cómo eran algunos modelos, uno de los objetivos iniciales de la evaluación. Aunque halló algunas diferencias sistemáticas entre clases incluidas en *Follow Through* y otras no pertenecientes al programa, también descubrió que los modelos no constituían tratamientos unitarios. Estos estudios auxiliares no influyeron en absoluto en la evaluación final.

Aunque la fase MCDANIELS/*Huron Institute* sólo duró un par de años, supuso modificaciones espectaculares en la evaluación. La *Office of Education* adoptó un control firme, reduciendo el ámbito de la evaluación a unas pocas medidas normalizadas de resultados y transformando la evaluación de operación política en actuación técnica. El *Huron Institute* prestó su dominio técnico de la cuestión y la justificación racional de esos cambios.

La fase EVANS/Abt Associates: 1974-1977

La última fase supuso la ejecución técnica del diseño de evaluación a cargo de *Abt Associates* y el completo dominio de la *Office of Education* en la evaluación. Eugene TUCKER sustituyó a Garry MCDANIELS, transfiriéndose toda la evaluación del *Bureau of Elementary and Secondary Education* a la *Office of Planning, Budgeting, and Evaluation* (OPBE) en enero de 1974. John EVANS, director de la OPBE, fue el responsable efectivo de las decisiones.

Cuando estaba en la *Office of Economic Opportunity*, EVANS fue el responsable principal de la evaluación Westinghouse/Ohio del programa *Head Start*. En muchos aspectos, la primitiva evaluación de *Head Start* era semejante a la del *Follow Through*. El trabajo *Head Start* fue un estudio de "impacto", concentrándose en la elevación de las puntuaciones de los niños en los tests, aunque no era éste el aspecto planeado de la variación. El *Head Start Research Council* se manifestó en contra del diseño de ese estudio, realizando las mismas críticas que más tarde se le harían, cuando

© Ediciones Morata, S. L.

se terminó (DATTA, 1976). Las quejas en contra se referían a la insensibilidad del diseño, la dificultad de encontrar grupos de control y de supervisar, desde el punto de vista estadístico, los sesgos, la falta de medidas adecuadas de desarrollo personal y social, la falta de conocimiento de los programas concretos, etcétera. El grupo de EVANS decía que la información era necesaria para poder tomar decisiones y que los sesgos se anularían mutuamente. El estudio se llevó a cabo y no mostró diferencias entre las clases pertenecientes al programa *Head Start* y las no incluidas en él. Los resultados fueron muy controvertidos, utilizándolos Nixon para impedir la concesión de mayor cantidad de fondos para el programa.

Para centralizar las operaciones de evaluación en la *Office of Education*, James Albert, vicesubsecretario del departamento de Salud, Educación y Bienestar para la Evaluación y Supervisión, llevó a EVANS a la *Office of Education*. Tanto Albert como EVANS creían en la evaluación del impacto, es decir que los efectos de un programa se juzgarían por su repercusión en las puntuaciones que los niños obtuvieran en los tests (MCLAUGHLIN, 1975, página 110).

EVANS no perdió el tiempo enfrentándose a los colaboradores. En una reunión celebrada en octubre de 1974 en Clinton (Michigan), respondiendo a un patrocinador partidario de medir el cambio institucional, dijo:

> La evaluación es una carrera de caballos; una evaluación comparativa de la eficacia relativa de diversos enfoques... Decir que todos los modelos han realizado avances importantes, es una política sentimentaloide, así que "elijan ustedes"... Estamos de acuerdo en que los efectos últimos del *Follow Through* deben consistir en que, al cabo de diez años, sus graduados tengan un estilo de vida diferente, siendo menor entre ellos el número de drogadictos o presidiarios. Pero el problema es: ¿cómo pueden ocuparse de esto los investigadores? Utilizamos las pruebas que tenemos como aproximación a aquellos efectos a largo plazo.
>
> Las decisiones, incluso las de mayor peso político, pueden verse influidas por las pruebas. Con frecuencia, el Congreso actúa de forma irresponsable al no hacer frente a los problemas. Responde a los padres desde las gradas del Capitolio. Pero las pruebas aún tienen fuerza. Sin embargo, si padres y niños se muestran entusiastas, pero las pruebas son anecdóticas y no persuasivas, sería una irresponsabilidad de nuestra parte decir a otras comunidades que tuvieran en cuenta el modelo. Los juicios que se hagan respecto a la cualidad irresistible de las pruebas tendrán que ser arbitrarios. El problema consiste en el reparto de recursos escasos.
>
> (Informe de la *Fourth Annual Follow Through Working Conference*, 1974.)

En la reunión de Clinton, algunos colaboradores sugirieron que la *Office of Education* considerase algunos enfoques alternativos de la evaluación, como los antropológicos. Sobre esto, EVANS se mostró firme: "No estoy de acuerdo. La otra metodología es de baja categoría en relación con la calidad de sus pruebas. Las ciencias sociales tienden hacia un mayor rigor. No adoptaremos voluntariamente enfoques inferiores". Una vez más, los cola-

© Ediciones Morata, S. L.

Scanned by CamScanner

boradores alegaron que las medidas no servían por su carácter restringido y su invalidez.

Mientras tanto, *Abt Associates* llevaba a cabo el análisis de los datos. Los dos primeros informes de *Abt Associates* se publicaron en 1974 y 1975. Los resultados no complacieron a los funcionarios de la *Office of Education*. Los informes mostraban que los efectos de cada modelo diferían de modo espectacular de un sitio a otro. Esto hacía imposible decir qué modelo era mejor. Es más, CLINE y sus colegas de *Abt Associates* informaron de que, al diferir entre sí las muestras de los colaboradores de todas las formas posibles, "no podemos comparar los efectos entre los distintos colaboradores" (HANEY, 1977, pág. 212).

La respuesta de la *Office of Education* ante estas noticias fue inequívoca. Se preparó una nueva propuesta de análisis de datos y la primera cuestión de la misma consistió en comparar directamente entre sí a los colaboradores. *Abt Associates* reemplazó a CLINE como director del proyecto, diciendo que ellos podían comparar a los colaboradores. Participaron en la correspondiente licitación y consiguieron por segunda vez el contrato para hacer lo que dijeron que no podían hacer en la primera ocasión. Las principales decisiones del análisis sólo pudieron realizarse de acuerdo con la *Office of Education*. Del nuevo plan de análisis se eliminó la comparación global entre el *Follow Through* y el sistema no perteneciente al mismo. La *Office of Education* decidió que no quería que se hiciera tal comparación que, en realidad, no mostraría diferencias, centrándose únicamente en las comparaciones entre modelos.

En mayo de 1975, se concedió un contrato de análisis de costes a *Resource Management Corporation* para valorar los correspondientes a cada modelo. Cuando la evaluación principal determinara los incrementos comparados de las puntuaciones obtenidos por cada modelo, los costes servirían para designar el modelo más barato, de mejor relación coste/producto. Este análisis se frustró porque cada patrocinador recibió la misma cantidad de dinero por niño, mientras la financiación local a cada centro de aplicación variaba en gran medida.

Cuando se mostró a los colaboradores los resultados del tercer informe de *Abt Associates* (aunque no el capítulo en el que se comparaban los modelos entre sí), una vez más se molestaron por el carácter restringido de la evaluación, así como por las descripciones que *Abt Associates* hizo de sus modelos. Organizaron una reunión de protesta en Washington, en abril de 1976, y presentaron a *Abt Associates* y a la *Office of Education* una larga lista de críticas en contra de la validez de la evaluación. Como era habitual, *Abt Associates* prometió efectuar más análisis estadísticos.

El cuarto, último y principal informe de *Abt Associates* fue aceptado por la *Office of Education* en abril de 1977. En él se comparaban de forma explícita cada modelo de atención a la primera infancia con los demás. Los colaboradores lo condenaron rotundamente. Las enormes variaciones entre distintos lugares de aplicación obligaron a los analistas de *Abt Associates* a ser más cautos en relación con sus descubrimientos. Los resultados de

© Ediciones Morata, S. L.

todos los modelos eran buenos en unos sitios y malos en otros. En realidad, la variación dentro de cada modelo era tan grande que los analistas se negaron a decir que un modelo fuese mejor que otro, aunque podían apreciarse las diferencias en los diagramas. Los analistas dijeron que los llamados modelos de "destrezas básicas" eran mejores.

Cuando un periodista le preguntó acerca del importante hallazgo de la impresionante variación que mostraba cada modelo según su lugar de aplicación, John EVANS dijo que la variación era "frustrante", pero indicó que "no se trataba de que el mundo esté loco o que los lugares fuesen totalmente diferentes", sino que los modelos "no se habían implementado de manera uniforme". "Cuando se paga a un patrocinador para que ponga en práctica un modelo —dijo— los lugares de aplicación deben ser relativamente homogéneos". La *Office of Education* se ocuparía de hacer un esfuerzo para comprender "las amplias variaciones entre los diversos lugares" (*Education Daily*, 10 de agosto de 1977). La idea de que el mismo modelo obtuviera resultados diferentes en distintos lugares no parecía una conclusión aceptable.

La lógica de la evaluación

Toda evaluación tiene una lógica propia cuyas posibilidades he señalado en otra parte. La lógica de la evaluación del *Follow Through* estaba configurada, en definitiva, por el enfoque del análisis de sistemas, que ha dominado la política federal durante una década. Este paradigma busca el programa más eficaz y requiere que se comparen de manera bastante precisa tanto los resultados como los costes. Esto supone utilizar unas cuantas medidas cuantitativas con respecto a las cuales puedan compararse los modelos entre sí.

El *Follow Through* comenzó con un amplio conjunto de objetivos sociales que debía cumplir. Los colaboradores y los padres trataron de mantener estos objetivos de la evaluación, pero los evaluadores y la *Office of Education* consiguieron reducir los resultados a unas pocas medidas comunes sobre las que efectuar comparaciones de grado. Por ejemplo, aún después de que el programa se redefiniese como experimento, en teoría, la evaluación podría haber examinado las consecuencias de cada modelo en una amplia escala y en cada una de las nueve áreas de impacto, incluido el cambio institucional. Esto podría haberse llevado a cabo mediante un razonamiento pragmático, juzgando un modelo por sus muchos efectos.

Sin embargo, no fue ése el razonamiento que orientó la evaluación. De nuevo, la cuestión dominante fue: "¿qué modelo funciona mejor?", procurando determinar qué modelo produjo el máximo efecto con un coste determinado, y no los efectos globales. La determinación del modelo "mejor" obligó a que la evaluación se desarrollase según el razonamiento de medios a fines. Dados los fines previstos, ¿qué modelo los cumple mejor o al menor coste? Ese modelo sería el mejor.

© Ediciones Morata, S. L.

Scanned by CamScanner

Esto suele denominarse racionalidad técnica y es la predominante en economía. En el ámbito del gobierno federal, fue defendido por el ASPE, la OPBE y la *Office of Management and Budget*, todos involucrados en la evaluación del *Follow Through*. Parte de la base de que la racionalidad consiste en comparar alternativas claramente definidas y escoger entre ellas según sus efectos sobre el objetivo concreto en el que se piensa. La idea de las variaciones planificadas constituye de por sí una manifestación de esta idea. Hay que dar por supuesta la existencia tanto de unos medios y fines definidos con claridad como una relación de causa a efecto entre ellos. La evaluación obligó a que el programa *Follow Through* siguiera el paradigma técnico.

Era necesario especificar cuáles eran los medios y los fines. La variación planificada suponía la existencia de medios claramente definibles, es decir, los modelos. Los patrocionadores, así como los estudios de SOAR y de STALLING, mostraron que no era así, pero la evaluación partía de la base de que era de ese modo, pasando por alto esta evidencia conflictiva. Los modelos se establecieron por definición, nombrándolos.

La descripción de los fines era aún más problemática. El programa se justificaba sobre la base de que trataba de cumplir gran número de fines sociales. Pero la comparación directa sería imposible o, al menos, confusa, si se utilizaba gran número de indicadores. El paradigma parte de la simplicidad. Las comparaciones sólo podrían hacerse a partir de medidas comunes y cuantas menos, mejor. El SRI comenzó con tests cognitivos normalizados y, como reacción a las protestas de los colaboradores, trató de elaborar un conjunto de ítems de test comunes a todos y sensibles a los modelos de los colaboradores. Estos tests especiales no proporcionaban equivalentes de grados normalizados, resultando impracticables, tanto desde el punto de vista logístico como financiero. En realidad, los colaboradores no tenían en común unos pocos objetivos sencillos.

Con la llegada de MCDANIELS, la *Office of Education* resolvió el problema por imposición. Prescribió un conjunto de fines comunes y cuatro medidas respecto a los resultados. El fundamento de la utilización del *Metropolitan Achievement Test* consistía en su posibilidad de utilización a efectos comparativos. En su reunión con los colaboradores, EVANS reconoció que, en último término, se trataba de mejorar las oportunidades de los niños menos favorecidos, pero afirmaba que los tests constituían la mejor aproximación a la medida del éxito. La elevación de las puntuaciones en los tests cognitivos constituyó el fin principal que los evaluadores tenían presente al principio y siguió siéndolo hasta el final.

Habida cuenta de la necesidad de comparar los modelos entre sí, también era preciso utilizar medidas cuantitativas, como son las puntuaciones en los tests. La cuantificación permite escalar los efectos de los modelos, agudizando las diferencias entre ellos y facilitando las comparaciones o, al menos, haciéndolas más justificables.

Las clases participantes en el programa *Follow Through* se compararon con clases locales de control mediante análisis de covarianza. El hecho de

© Ediciones Morata, S. L.

que este análisis no se adaptase demasiado bien a la realidad del programa planteó problemas, sobre todo para quienes conocían bien los modelos, que se dieron cuenta de que el ajuste entre las técnicas cuantitativas y las clases reales no era bueno. No obstante, se emplearon para facilitar las comparaciones previstas. La preocupación por agudizar al máximo las diferencias mediante las técnicas cuantitativas era patente en el fundamento en el que DAVID basaba la muestra y en el aportado por EVANS para las técnicas cuantitativas.

A lo largo de la evaluación, tanto los fines como los medios se restringieron drásticamente para poder efectuar las comparaciones. Los mismos colaboradores elaboraron más de sesenta instrumentos para medir sus modelos. Esos instrumentos aparecen reseñados en un volumen de la *Office of Education*. Pero, en el informe final de *Abt Associates* se comparaban trece modelos respecto a cuatro tests que daban once puntuaciones. Aún esto resultaba demasiado complejo para efectuar comparaciones sencillas. Los modelos se redujeron entonces a tres tipos y las medidas de resultados a tres índices.

La aplicación de técnicas estadísticas deja a los analistas un conjunto de resultados de difícil interpretación. Además, los resultados no se obtenían de forma clara. Aparecían lugares de aplicación que arrojaban malos resultados, aun con modelos "buenos". Sin embargo, en virtud del contrato, los analistas tenían que comparar modelos. Su solución consistió en definir un conjunto de tipos de modelo y prorratear las medidas de resultados entre estos tipos. Mediante la denominación de los modelos como "de destrezas básicas", "cognitivo/conceptual" y "afectivo/cognitivo", crearon un conjunto de tipos que, de forma correcta o incorrecta, la gente podía interpretar con facilidad.

Pero las medidas de resultados se inclinaban con toda claridad a favor de los llamados modelos "de destrezas básicas". ¿Qué se podía hacer? Los analistas denominaron también las medidas de resultados como de destrezas básicas, cognitivas y afectivas, dando a entender que las medidas se ajustaban a los tipos pertinentes. En realidad, algunas de ellas se ajustaban bien a los modelos "de destrezas básicas", pero las demás no se adaptaban a los otros tipos equivalentes. Esta clasificación *ex post facto* constituía un modo de interpretar los resultados de acuerdo con categorías fáciles de reconocer, aunque fuesen erróneas, y disimulaban el sesgo de las medidas a favor de determinados modelos.

La clasificación de éstos se efectuó por definición, con poco fundamento empírico. La primitiva clasificación había definido los modelos de forma muy distinta (por el grado de estructuración, por la importancia que concedían a las tareas para casa, por la orientación del programa, etc.). Se obligó a los colaboradores a que utilizaran las nuevas categorías. La denominación otorgada a los modelos que hacían hincapié en la mecánica de la lectura y de las matemáticas como "de destrezas básicas" proporcionó a esa categoría una aceptación pública. Pero las denominaciones eran desorientadoras. Por ejemplo, la lectura no se incluía como medida de "destreza básica".

© Ediciones Morata, S. L.

Scanned by CamScanner

Oponerse al razonamiento del estilo medios-fines atacando la estrechez de miras, los instrumentos de medida, las técnicas estadísticas, la definición y clasificación de modelos y medidas y la interpretación de los resultados era oponerse a la misma lógica de la evaluación. Tanto los colaboradores como otros lo hicieron de manera reiterada. Tras el informe final de *Abt Associates,* un tribunal de evaluadores independientes criticó la evaluación, encontrando en ella tan graves defectos, en relación con los aspectos mencionados, que hacían inválidas muchas de sus conclusiones (HOUSE y cols., 1978).

Para algunos defensores de la evaluación, la variación hallada entre distintos lugares de aplicación del mismo modelo también resultaba incómoda. No tenía por qué. La *Office of Education* decidió borrarla del mapa ante el primer informe de *Abt Associates,* al financiar un nuevo análisis de datos. Hay que decir, a favor de los analistas de *Abt Associates,* que ellos la reconocieron como el descubrimiento primordial en su informe final. En otras palabras, los modelos no producían efectos uniformes en todas partes.

Esto molestó a EVANS. Si cada lugar fuese tan diferente de los demás que los mismos modelos produjeran un impacto distinto en cada sitio, estaríamos en un "mundo loco". Él atribuyó esta variación al error de implementación del patrocinador. En caso contrario, a la *Office of Education* le habría sido más difícil prescribir determinados programas o decidir la distribución de recursos escasos.

Objetividad

La evaluación se defiende, en último extremo, por su objetividad y por la necesidad de tomar iniciativas. Como es necesario decidir, es mejor hacerlo sobre la base de datos objetivos que sobre información subjetiva, aunque la evaluación no sea buena. Este razonamiento se basa en la presunta objetividad de los instrumentos, la muestra y los procedimientos estadísticos. Es una definición de objetividad muy especial.

Con frecuencia, se equipara la objetividad con el acuerdo entre observadores. Se llega a ese acuerdo mediante procedimientos de observación externos y especificados de antemano. Según esta definición, la objetividad se consigue cuando se dispone de observadores que estén de acuerdo en lo que ven (repetición de observaciones). Se acepta lo que acuerda un conjunto de personas. Todo lo que no se observe a menudo, se presume que es meramente subjetivo. Ésta es la idea cuantitativa de objetividad, que se identifica con la epistemología objetivista del análisis de sistemas, y constituía un concepto orientador de la evaluación del programa *Follow Through.*

En diversas fases de la evaluación, distintas medidas de resultados fueron rechazadas como "poco fiables". La fiabilidad es casi equivalente a la idea cuantitativa de objetividad, basándose en repeticiones de las observaciones. Y esto se acerca mucho a lo que EVANS llamaba "rigor" en la

© Ediciones Morata, S. L.

discusión con los patrocinadores. La fiabilidad puede representarse también en forma matemática.

El SRI abandonó las medidas no cognitivas y, hasta cierto punto, el estudio de casos de cambio institucional a causa de su poca fiabilidad. *Abt Associates* eliminó su escala de implementación por la misma razón. El *Huron Institute* extrajo una muestra con varios lugares de aplicación por cada patrocinador basándose en que los sesgos se eliminarían si se repetían las observaciones. La *Office of Education* seleccionó la batería normalizada de tests dando por supuesto que cumplía los requisitos psicométricos de fiabilidad. Un requisito del enfoque consistía en que los resultados se redujeran con el fin de efectuar comparaciones; la fiabilidad constituía una forma fundamental de tomar decisiones.

(Es interesante señalar que *Abt Associates* abandonó también el "número de días de ausencia", un indicador muy fiable, sobre la base de que factores ambientales de la comunidad podrían haber determinado la ausencia. Por la misma razón, refrendada por los propios resultados obtenidos, podrían haberse eliminado los tests. Éste es el criterio de validez, no de fiabilidad. Es decir, decían que la ausencia no era un buen indicador de los efectos de los modelos, del mismo modo que los colaboradores afirmaban que los tests tampoco lo eran.)

En su forma extrema, la idea cuantitativa de objetividad, denominada intersubjetivismo, se convierte en operacionalismo: para que algo se considere "cierto" ha de ser visto y especificado de manera conjunta. SCRIVEN (1972) hace hincapié en que esta noción de objetividad constituye un error filosófico fundamental. Confunde el método de verificación con la "verdad". Sostiene SCRIVEN que hay otra noción de objetividad que depende más de la calidad de la idea que de la cantidad de personas que la defienden. Llama a esto "sentido cualitativo de la objetividad". En este sentido, ser objetivo significa ausencia de sesgos o deformaciones.

Menciona SCRIVEN el incidente del evaluador de receptores de televisión que acopla un instrumento mecánico al televisor para medir la ganancia en decibelios, aunque el aumento de éstos no correlacione demasiado con la calidad de imagen. A simple vista, se podría juzgar mejor la calidad de imagen. El técnico hace esto porque se llega a un acuerdo mayor entre observadores sobre la lectura de un instrumento, aunque también se consiga la fiabilidad a simple vista, en un nivel más bajo. En otras palabras, la fiabilidad suplanta la validez. Se utilizan instrumentos más fiables aunque sean menos válidos.

Lo mismo podemos decir de la evaluación del *Follow Through.* Se defiende la objetividad de la evaluación porque utiliza instrumentos fiables, aunque quizá no fuesen válidos para los fines planteados. En otras palabras, la evaluación era objetiva en el sentido cuantitativo de que sus hallazgos podían repetirse (eran "estables", utilizando el término empleado por los analistas de *Abt Associates*), pero la evaluación se inclinaba a favor de los modelos beneficiados por los tests. Por ejemplo, el MAT mide de manera fiable el empleo de los signos de puntuación y pone de manifiesto que el modelo de

© Ediciones Morata, S. L.

Scanned by CamScanner

"instrucción directa", cuyos materiales están pensados en relación con la puntuación, resuelve mejor esta parte del test. Este resultado es repetible, pero ello no implica necesariamente que el modelo de "instrucción directa" sea el mejor.

El segundo sentido en el que se dice que la evaluación es objetiva consiste en que emplea un procedimiento estadístico que puede desarrollarse con mayor o menor corrección. La validez de las matemáticas se presta a la inspección pública. Esto hace que la evaluación parezca objetiva aunque se suscite la cuestión de si la técnica matemática adecuada que se debe emplear ha de ser ésta. Se cuestionan tanto la formulación como la interpretación de los resultados, aunque la técnica matemática sea impecable. En esta evaluación son discutibles la formulación, la interpretación y las matemáticas (HOUSE y cols., 1978).

Por tanto, la última defensa de la evaluación es su objetividad, pero ésta desaparece cuando se somete a examen. Las medidas son fiables, pero su validez es discutible (algunas tampoco son fiables). La estadística es más o menos correcta pero, a veces, resulta inadecuada y la interpretación es discutible. La evaluación es replicable, fiable y no válida.

La identificación de la objetividad con procedimientos especificados de forma externa tiene otro efecto. Inclina al evaluador a desentenderse de su responsabilidad respecto de los hallazgos. Éstos son el resultado de sus instrumentos y procedimientos objetivos. ¿Cómo puede responsabilizársele de ellos? En esto consiste el llamado "objetivismo" (POLANYI, 1958). No puede sorprender a nadie que los analistas de *Abt Associates,* al enfrentarse con las paradojas de su tarea, se viesen tentados a veces a negar su responsabilidad. De hecho, la parcelación de tareas específicas impuesta por la *Office of Education* redujo la responsabilidad que cualquiera de los grupos pudiera sentir en relación con la evaluación.

Nada de esto pretende que se deje de lado el concepto de fiabilidad. En efecto, se trata de un concepto útil, pero no debe emplearse para excluir la validez. Por ejemplo, la evaluación podría haber contado con psicólogos que entrevistaran a los niños sobre su salud mental general, obteniendo información más válida en el terreno afectivo. La objetividad (en sentido cuantitativo) es una justificación insuficiente. Una evaluación puede ser objetiva y, sin embargo, sesgada.

Equidad

La evaluación del *Follow Through* no era equitativa en dos sentidos. En primer lugar, por la forma en que las medidas de resultados primaban ciertos modelos sobre otros. La clasificación de modelos y medidas de resultados en las categorías de "destrezas básicas", "cognitivos/conceptuales" y "afectivos", y la correspondencia entre medidas y modelos sobre esta base, da sólo una apariencia de equidad. En realidad, era mucho más fácil para los modelos de "destrezas básicas" obtener buenas puntuaciones

© Ediciones Morata, S. L.

en "sus" medidas que para los demás. Por ejemplo, el tribunal de la *Office of Education* seleccionó el test de RAVEN porque era imposible enseñar a hacerlo bien, pero ningún modelo puntuó mejor en él que los grupos de comparación. Ahora bien, como estaba incluido entre las medidas cognitivas, parecía que los modelos cognitivos tendrían que favorecer específicamente una buena puntuación en dicho test.

El segundo sentido en el que la evaluación no era equitativa se refiere a la idea que tenían los colaboradores de los evaluadores y de la *Office of Education.* Se hizo creer reiteradamente a los colaboradores que la evaluación respondería a determinados objetivos que pretendían conseguir. La lista original de criterios de evaluación era lo bastante amplia como para que tuviesen cabida las pretensiones de los colaboradores y sobre esa base se prestaron a cooperar. Pero ya en abril de 1969, en Pajaro Dunes, comenzaron a protestar por la excesiva confianza depositada en las medidas cognitivas normalizadas. El SRI prometió elaborar nuevas medidas cognitivas y no cognitivas. Trató de hacerlo, pero fracasó.

Si alguien llega a un acuerdo con otro, sea tácito o explícito, y cumple sus promesas, pero la otra parte no, se sentirá tratado de manera injusta. Los colaboradores protestaron de forma reiterada y se les prometieron los cambios pertinentes, pero, al final, la evaluación se basó sobre todo en medidas cognitivas convencionales. Esta sensación de trato injusto y desigual explica las protestas contra la evaluación expresadas por tantos colaboradores. Se sintieron traicionados.

Para que se considere equitativo, un contrato o acuerdo de evaluación debe cumplir ciertas condiciones. En el Capítulo VIII, señalé trece condiciones de este tipo, como la ausencia de coerción, información completa y un interés particular no excesivo. Una de esas condiciones es la de "posibilidad": que el contrato no se base en algo que una parte no pueda llevar a cabo. Es obvio que el SRI aceptó un contrato sobre algo que no podía hacer. En este sentido, el contrato original no era equitativo y así lo consideraron los colaboradores.

Cuando se produjo la redefinición de la evaluación, en la época de MCDANIELS, quedó claro lo que ésta haría. Las acciones de este autor, y muchas otras posteriores de la *Office of Education,* se tomaron de modo unilateral, sin mediar una negociación seria. La *Office of Education* rechazó el acuerdo anterior y ni siquiera consideró deseable establecer otro. Podían hacer lo que quisieran, sólo sometidos a las restricciones políticas.

Los colaboradores no están del todo libres de culpa en esta cuestión. Desde 1973, estuvo claro lo que llevaría consigo la evaluación final, aunque no las conclusiones a las que llegara. Ellos siguieron cooperando y aceptando grandes cantidades de dinero del gobierno. Por otra parte, en esa época habían invertido ya grandes sumas en el desarrollo de sus modelos. Algunos iniciaron un trabajo de desarrollo de cinco años, por lo que no cabía abandonarlo. Siguieron protestando y esperando que mejorase la situación.

En realidad, las circunstancias a las que tenían que hacer frente los colaboradores se aproximaban mucho a la coerción. Si el gobierno hubiera

© Ediciones Morata, S. L.

Scanned by CamScanner

negociado el tipo de evaluación que llevaría a cabo en relación con el contrato original, y esta evaluación hubiera sido conocida y aceptada en todos sus términos por los colaboradores, el contrato de evaluación hubiese sido equitativo. Sin embargo, los cambios de orientación habidos en pleno proceso, cuando los patrocinadores habían realizado ya inversiones, produjeron una situación muy diferente. Como mínimo, debería haberse renegociado la idea de la evaluación.

La equidad también se puso en peligro de otra forma: por la falta de imparcialidad. Deliberar con imparcialidad es deliberar sin prestar excesiva atención a los propios intereses, considerar sólo soluciones universales y escoger de acuerdo con el propio interés como miembro del grupo afectado por el contrato. Aunque la *Office of Education* cumplió las dos primeras condiciones, no así la tercera (actuar de acuerdo con los intereses de los afectados). No atendió de forma adecuada las quejas de los colaboradores. Por tanto, la evaluación careció de imparcialidad.

No basta con que una evaluación sea objetiva, sino que debe ser imparcial también. Los procedimientos externos, repetibles, pueden estar sesgados. Aunque la objetividad sugiere una actitud desprendida o neutral, característica que sí presentaba la evaluación, las personas sometidas a evaluación sí esperan que se les juzgue con imparcialidad, una preocupación expresada en juicios equilibrados sobre los intereses del programa que se juzga. Los patrocinadores percibieron correctamente ambas ausencias: de interés y de imparcialidad. Si la primera parte de la evaluación adoleció de falta de posibilidad, en la última no se produjo la imparcialidad. Y no sólo la evaluación fue poco equitativa y desigual, sino que el mismo *contrato* de evaluación no era equitativo.

Justicia

La evaluación del *Follow Through* concuerda bastante bien con la filosofía utilitarista de la justicia, de la que se derivaba en parte. La justicia utilitarista pretende conseguir el mayor equilibrio neto de satisfacción en la sociedad, mediante la suma de la satisfacción de todos los individuos. El utilitarismo exige una medida común de satisfacción que permita efectuar el cálculo de la utilidad. En educación, esta medida se estructura en torno a las puntuaciones obtenidas en tests normalizados, que sirven como pronósticos respecto a otros fines, como indicaron RIVLIN y EVANS.

En el utilitarismo, todo lo que procure el máximo "bien", es decir, las puntuaciones de los tests, en este caso, está "bien" hecho. Así, el modelo correcto de atención a la primera infancia es el que consiga las puntuaciones más elevadas en los tests. No procurar la máxima cantidad de bien es inmoral. La misma racionalidad se estructura como la selección de la mejor alternativa entre un conjunto de ellas. El diseño experimental, la instrumentación y la estadística son métodos para determinar la "mejor" alternativa.

© Ediciones Morata, S. L.

Por la intención, el experimento de variaciones planificadas del *Follow Through* constituía una espléndida manifestación de la ética utilitarista.

No obstante, el utilitarismo plantea ciertos problemas. Para lograr la máxima satisfacción total, es permisible reducir las satisfacciones de unos para aumentar las de otros. En la práctica, los así privados suelen ser los menos poderosos, las clases inferiores. Asimismo, alguien debe encargarse de hacer estas valoraciones de la satisfacción. A menudo, se ocupa de ello el gobierno o un administrador. En la práctica, el utilitarismo suele ser paternalista.

Estos elementos estaban presentes en la evaluación del *Follow Through*. En último término, la *Office of Education* decidía las cuestiones que tratar y las indicaciones que atender. Los grupos menos favorecidos protestaban cuando veían que la evaluación iba en contra de sus propios intereses. La primera protesta procedió de los grupos minoritarios en Atlanta. Los miembros de las minorías se quejaron de que los evaluadores se mostraban insensibles respecto a la participación de los padres y de la comunidad. Como respuesta a esta protesta, recibieron ayudas financieras varios colaboradores pertenecientes a las minorías y el SRI comenzó los estudios de casos comunitarios de cambio institucional. Al cabo de unos pocos años, se abandonaron los estudios de casos. No hubo investigaciones que se ocuparan de la comunidad en general. En la reducción de datos de la segunda fase, no se incluyó a la mayoría de los colaboradores de grupos minoritarios, aun siendo pocos.

Las entrevistas con los padres surgieron como un estudio independiente, siendo subcontratado con el *National Opinion Research Center*. Se llevaron a cabo sesenta mil entrevistas en el transcurso de varios años, pero nadie parecía conocer el objetivo exacto. Es posible que consistiera en recoger actitudes de los padres en relación con el programa y en proporcionar información demográfica. Nadie parecía capaz de decir qué hacer con los datos ni cómo interpretarlos. Cuando los analistas de *Abt Associates* trataron de incluir parte de la información en su tercer informe, uno de los colaboradores dijo que era demasiado "personalista" (HANEY, 1977, pág. 269). En el informe final de *Abt Associates* no se mencionan los datos recabados de los padres. HANEY (1977, pág. 270) dice que estos datos eran mucho más elogiosos del *Follow Through*.

En el diseño original del *Follow Through*, se pensaba que los padres desempeñaran un papel relevante en la toma de decisiones y en otros importantes aspectos del programa. Excepto en los pocos modelos que hacían hincapié en su participación, en realidad, fueron excluidos del programa, salvo de forma muy simbólica, lo que también es cierto respecto a la evaluación. Su última protesta importante se produjo en el congreso de Denver, en 1972, en donde se quejaron de haber sido excluidos y pidieron evaluar a los colaboradores. Decían que estaban cansados de que otros decidieran lo que era malo para sus hijos y cuáles eran los modelos buenos o malos. Estaban hartos de que los expertos sólo midieran lo que sabían cómo medir. Su protesta no consiguió nada, a pesar de que eran los únicos que podían representar con propiedad los intereses de sus hijos.

© Ediciones Morata, S. L.

Scanned by CamScanner

Los demás grupos carentes de poder excluidos de la evaluación fueron los profesores. En unos pocos estudios auxiliares, como el de SOAR y el de STALLINGS, se observó a los profesores, pero a la mayoría se le envió un cuestionario por correo. Una vez más, como no se les aceptó como participantes de pleno derecho, ninguno sabía con exactitud para qué era el cuestionario. En una ocasión, se propuso que los datos de los profesores se utilizaran como datos demográficos que mostraran las condiciones de implementación. De nuevo, los colaboradores se quejaron de la utilización de la implementación como medida. Después, nada se hizo con la información.

Visto en retrospectiva, los participantes menos poderosos del programa —minorías, padres y profesores— fueron, en realidad, excluidos de la evaluación. En la primera fase, las cuestiones se decidieron mediante la cooperación entre la *Office of Education,* los evaluadores y los colaboradores. En la segunda fase, la *Office of Education* y los evaluadores decidieron las cuestiones. En la tercera fase, sólo la *Office of Education* decidió las cuestiones críticas.

En el marco utilitarista, esta situación es justa porque, una vez determinado e implementado el modelo que produce el incremento máximo de las puntuaciones en los tests, aumentará la satisfacción total en la sociedad. Si se sacrifican los intereses de los excluidos, será ése el precio que se pagará en aras del progreso. Hay que ser fuerte y tomar "decisiones duras".

Por otra parte, hay quienes se oponen al utilitarismo y dicen que los intereses de todos los afectados por la evaluación deben tener voz en ella y deben estar representados de forma directa. Hay incluso quienes afirman que los intereses de los menos favorecidos deben constituir una preocupación fundamental. El hecho de que *Follow Through* fuese un programa diseñado para ayudar a los menos afortunados refuerza aún más esta cuestión. Desde estas perspectivas, la evaluación fue injusta.

Una revisión de la política federal de evaluación

La evaluación de *Follow Through* fue la culminación de una década caracterizada por una política peculiar y quizá estuvo muy cerca de incorporarla a sí misma. Dado el hecho de que la evaluación fuera sesgada, carente de equidad e injusta (además de costar cincuenta millones de dólares), ya es hora de cuestionar en serio esa política.

La política federal de evaluación se ha basado en el enfoque del análisis de sistemas. Sus principales destinatarios son directivos y economistas. Da por supuesto el consenso respecto a los objetivos, a causas y efectos conocidos y a unas pocas variables cuantificadas de resultados. Su metodología está constituida por la experimentación de variaciones planificadas y el análisis de coste/beneficio. Su fin es la eficiencia y se plantea la siguiente cuestión: "¿cuáles son los programas más eficientes?"

Como la articularon sus principales defensores, como RIVLIN y EVANS,

© Ediciones Morata, S. L.

supone la existencia de un paralelismo directo entre la producción de servicios sociales y la manufactura. Se aplican las mismas técnicas de análisis. El único conocimiento verdadero es una función de producción que especifica relaciones estables entre entradas *(inputs)* y salidas *(outputs).* La única forma de adquirir ese conocimiento pasa por los métodos experimentales y las técnicas estadísticas. Es posible ponerse de acuerdo sobre unas pocas medidas de resultados. El problema está constituido por la asignación eficiente de recursos.

Las decisiones clave procederán de las más altas instancias gubernamentales y la tarea puede llevarse a cabo mediante una gestión estricta. La justificación final es utilitarista: conseguir la máxima satisfacción en la sociedad. Para llegar a ese máximo, es preciso conocer cuáles son los programas más eficientes. Esto sólo puede hacerse comparando alternativas, por lo que hace falta una medida común de resultados. Es un trabajo de expertos.

Hay lugares en los que este enfoque puede aplicarse de manera satisfactoria, pero los Estados Unidos, en conjunto, no es uno de ellos. El enfoque puede llevarse a cabo sin problemas cuando los objetivos y las medidas de resultados sean pocos. Y es probable que suceda esto cuando los destinatarios de la evaluación constituyan un grupo muy estrictamente definido y se pongan de acuerdo sobre unos pocos criterios de comparación. Asimismo, es útil cuando los criterios pueden representarse mediante un indicador cuantitativo razonablemente válido. [Pueden verse aplicaciones satisfactorias de esta técnica en la evaluación de una grabación de entrenamiento para investigadores educativos, de GLASS (1972), o en la evaluación de un *curriculum* elemental de matemáticas para el *National Institute of Education,* de HERBERT (1975).]

En un programa social de base amplia en una sociedad pluralista, como los Estados Unidos, no es aplicable casi ninguno de los supuestos de partida del enfoque del análisis de sistemas. Preguntar: "¿qué modelo de educación de la primera infancia es mejor para los niños menos favorecidos en todos los lugares de los Estados Unidos?" sólo conduce a evaluaciones erróneas. En pocas palabras, no es cierto que existieran escasos objetivos o medias de resultados aceptados en relación con el *Follow Through.* El consenso se logró por imposición del gobierno. No es cierto que existieran alternativas distintas ni que causas y efectos fuesen conocidos o fáciles de determinar. Los modelos sólo explicaban una pequeña fracción —mucho menor del 10%— de la varianza de las medidas seleccionadas por el gobierno.

No es cierto que haya una función estable de producción que relacione entradas *(inputs)* y salidas *(outputs).* En un lugar, un modelo iría muy bien y en otro, muy mal. No es cierto que las puntuaciones obtenidas en los tests sean "datos puros y duros", como el número de botellas producidas por hora. Los efectos desaparecen a los pocos años. Los modelos no producen efectos uniformes en todas partes. Esto hace sospechar de las prescripciones

© Ediciones Morata, S. L.

Scanned by CamScanner

obligatorias de modelos dictadas por el gobierno. Tampoco es fácil ni fiable la transferencia de los modelos de unos lugares a otros.

No es cierto que la producción de servicios sociales sea como la manufactura de bienes. Ésta produce artefactos físicos; aquélla no. No es cierto que la única forma de conocer pase por la experimentación y la estadística. Hay muchas formas de conocimiento. No es cierto que no se sepa nada sobre la educación. Los profesionales saben mucho, aunque su conocimiento pueda ser tácito.

No es cierto que la recogida de cantidades muy grandes y costosas de datos elimine los sesgos. Tampoco es cierto que los directivos y economistas sean los únicos destinatarios de las evaluaciones. Ante cada evaluación de este tipo, los periódicos toman los resultados y los interpretan de manera muy distinta de la de los expertos.

En programas sociales a gran escala, habrá muchos grupos que traten de satisfacer intereses muy divergentes. La evaluación del programa debe reflejar de alguna manera los intereses de estos grupos divergentes. Es probable que esto suponga múltiples criterios y métodos de recogida de datos. Significa que la información no puede resumirse por completo en una simple escala cuantitativa. Significará, en cambio, que el programa producirá resultados y satisfacciones diferentes para distintos grupos. Estos grupos tendrán que estar representados en la evaluación y hay diversas formas de hacerlo. Existen, al menos, otros ocho enfoques de la evaluación.

Quizá el principal error del enfoque del análisis de sistemas consista en la confusión entre objetividad e imparcialidad. Como señala ELMORE:

> En su formulación primitiva e ingenua, el PPBS coloca al "personal experto muy especial" del brazo del responsable de la política (el presidente, el jefe de gabinete, el director del organismo). Su consejo debe ser objetivo y desapasionado, basado en el análisis estricto e incontaminado por intereses burocráticos estrechos. Por su objetividad, los responsables políticos deberán considerarlo superior a otros tipos de consejos y, por tanto, desempeñará un papel determinante en la elaboración de la política.
>
> (ELMORE, 1976, pág. 23.)

Los analistas pensaban que la objetividad era suficiente para asegurar la superioridad y la influencia. Con mayor frecuencia, supone falta de relieve. La objetividad trata los intereses excluyéndolos del mapa. Hace falta la imparcialidad, que cuenta con los intereses, los incluye y los equilibra.

© Ediciones Morata, S. L.

CAPÍTULO XI

Crítica de los enfoques

A lo largo de este libro, he expuesto y criticado diversos enfoques de la evaluación. Este capítulo es un breve resumen de las ventajas e inconvenientes de cada enfoque. Cualquiera de ellos puede ser adecuado o inadecuado, dependiendo de las circunstancias de su aplicación y de la correspondiente validez de las premisas en las que se base. Aunque un enfoque sea adecuado, la evaluación concreta puede ser buena o mala, exacta o inexacta, coherente o incoherente, equitativa o no, democrática o antidemocrática, justa o injusta, útil o inútil, eficiente o ineficiente. Si determinado enfoque se aplica a una situación para la que resulta inadecuado, se multiplican las posibilidades de una evaluación incorrecta.

Una forma de desarrollar el resumen consistiría en hacer una lista de las normas y principios expuestos en el libro y juzgar cada modelo en relación con cada uno de los criterios. Sin embargo, este procedimiento sería demasiado tedioso, además de poder inducir a error. En realidad, sólo puede someterse a juicio una evaluación concreta llevada a cabo en circunstancias determinadas. En sí, los enfoques básicos sólo muestran ciertas tendencias. Por ejemplo, el análisis de sistemas tiende a pasar por alto los intereses de algunos participantes en el programa, pero no tiene por qué hacerlo. En consecuencia, criticar el enfoque del análisis de sistemas como esencialmente antidemocrático es excesivo, aunque muestra cierta debilidad en el sentido citado. Si disponemos de un mapa que realce o deforme, incluso, algunas características a expensas de otras no tenemos por qué perdernos necesariamente. El hecho de que un enfoque de evaluación haga hincapié en la fiabilidad de los datos en detrimento de la representatividad de intereses pertinentes no tiene por qué traducirse en la realización de una evaluación antidemocrática.

La estructura de este capítulo es más relajada, pareciéndose más a una discusión casual sobre los puntos fuertes y las debilidades de cada enfoque, según convenga. Yo mismo he utilizado o criticado por completo cada

© Ediciones Morata, S. L.

Scanned by CamScanner

uno de los tipos y pondré estos ejemplos personales para ilustrar mis críticas.

Análisis de sistemas

Ningún enfoque de evaluación ha recibido en este libro críticas más severas que el análisis de sistemas, sobre todo en el capítulo dedicado a la política federal, a causa de que el análisis de sistemas constituye el enfoque dominante, sancionado de forma oficial, y porque sus defensores hacen gran hincapié en que es el único enfoque científico. Estas protestas de cientificidad se tornan un tanto embarazosas porque, desde el punto de vista epistemológico, son ingenuas e inocentes (vestigios de un positivismo desacreditado). Por supuesto, no todos defienden el enfoque de modo tan extravagante.

No es preciso que nos extendamos sobre las falacias y limitaciones del análisis de sistemas. El enfoque en cuestión, tal como lo presentan teóricos como RIVLIN (1971) y ROSSI y cols. (1979), asume el punto de vista de dirigentes gubernamentales y economistas. Debe obtenerse información sobre la planificación, supervisión, impacto y costes de un programa a efectos de gestión. En cierto sentido, se trata de un punto de vista realista, dado que la mayor parte de las evaluaciones corresponden a programas gubernamentales pero, por desgracia, con demasiada frecuencia, este enfoque es dificitario desde las perspectivas política y moral. Excluye los intereses y preocupaciones de los participantes en el programa, sobre todo de quienes están situados en los niveles inferiores de la jerarquía social. La evaluación del programa *Follow Through* constituye un ejemplo excelente de esta postura de pasar por alto los intereses de los pobres, los padres y los profesores.

En el peor de los casos, el enfoque de análisis de sistemas lleva al cientifismo (punto de vista que sostiene que el único modo de acceder a la verdad pasa por el uso de determinadas metodologías). La objetividad se equipara a la fiabilidad, produciéndose información a partir de ciertos tipos de instrumentos exclusivamente. Se sacrifican la imparcialidad y la validez. Al reducir todo a algunos indicadores, de manera que se pueda: demostrar la fiabilidad, efectuar análisis de coste/beneficio y descubrir cuáles son los programas más eficientes (consiguiendo, de esto modo, la máxima utilidad), los resultados de programas sociales complejos se restringen a unas pocas medidas cuantitativas. En general, la aplicación del enfoque del análisis de sistemas supera con mucho el ámbito para el que resulta adecuado.

No obstante, a veces, el enfoque del análisis de sistemas es adecuado, como cuando las medidas pueden reducirse razonablemente a una pocas posibilidades y a relaciones sencillas de causa a efecto. Por ejemplo, si nos referimos al número de camas hospitalarias de una comunidad, parece adecuado un indicador cuantitativo. Sin embargo, no se nos ocurriría utilizar esto como única medida de la atención prestada a la salud. O bien, si el

© Ediciones Morata, S. L.

propósito y único objetivo del programa consiste en elevar las puntuaciones de los estudiantes en los tests, tiene sentido utilizar éstas como medida del éxito del mismo. (Véase una exposición en profundidad sobre los tests y su relación con la evaluación y la política en BAKER y QUELLMALZ, 1979.)

Un modo de reforzar el enfoque de análisis de sistemas consiste en ampliar el número y los tipos de indicadores utilizados. Por ejemplo, para evaluar el impacto de un nuevo programa de matemáticas, HERBERT (1975) elaboró varias docenas de escalas especiales que median capacidades que los matemáticos creen indican un pensamiento de alto nivel, en vez de ítems de tests normalizados. En un diseño clásico con grupo control, se administraron tanto los tests especiales como los normalizados para demostrar que el programa obtenía mejores resultados que los materiales tradicionales en los tests especiales y no peores en los normalizados. Sin los tests especiales, habría estado justificada la protesta de los autores del programa acerca de la falta de validez de la evaluación. HERBERT llevó a cabo también una serie de estudios colaterales para eliminar la posibilidad de que otros factores, como el determinado por mejores profesores, influyeran en los resultados. He aquí un estudio bien razonado. Por desgracia, muchos estudios de análisis de sistemas son poco ingeniosos, quedándose en meras aplicaciones mecánicas de métodos al uso. Se han efectuado buenos estudios de análisis de sistemas, coincidentes en el buen razonamiento que los guía y en no tratar de hacer pasar elefantes por el ojo de una aguja.

Objetivos conductuales

El enfoque de la evaluación por objetivos conductuales es el más conocido y, si se votase al respecto, quizá fuese el más utilizado por los evaluadores profesionales. Respecto a cualquier evaluación, es el enfoque que con mayor frecuencia se plantea. También son muchas las críticas que sostienen que la especificación de objetivos no es, en realidad, útil para la enseñanza ni para el desarrollo curricular (ATKIN, 1968; STAKE, 1973). Estas críticas han ido aumentando, pero el enfoque por objetivos sigue siendo popular.

Este enfoque se basa en la premisa de que la educación, o cualquier actividad social, constituye, en realidad, una tecnología, un conjunto de técnicas que conduce a fines especificados de antemano. WAKS (1975) señala que este modelo se funda en dos doctrinas: la "doctrina de la planificación y evaluación técnicas" y la "doctrina de la especificación conductual". La primera sostiene que si se confunden los fines u objetivos, no es posible aplicar las técnicas de manera adecuada. Los objetivos deben ser precisos y exhaustivos. Para especificar la acción correcta, el "espacio de evaluación" ha de quedar clausurado por un completo conjunto de objetivos que determinen con exactitud los resultados. La segunda doctrina se refiere a la especificación conductual, de donde toma su nombre el enfoque. Según ella, los objetivos de comportamiento han de estar especificados conductualmente, o sea deben ser mensurables. En conjunto, ambas

© Ediciones Morata, S. L.

Scanned by CamScanner

doctrinas llevan a una evaluación que define el éxito en términos de objetivos especificados de antemano y mensurables, y no respecto a los valores que en realidad se han conseguido.

La crítica básica de WAKS consiste en que "actuar con un propósito" no equivale a "poner los medios para llegar a un fin determinado". Cualquier actividad social, como la enseñanza, puede ser coherente y eficaz fundándose en las tradiciones culturales, la preparación, la experiencia y la socialización de los profesores. Ninguna de estas características requiere una especificación de objetivos en forma conductual. Según WAKS, el profesor participa de una forma completa de vida y de una manera de actuar y pensar. Como los profesores disponen de mitos, historia, tradiciones y costumbres en las que basarse, no necesitan un "flujo de objetivos" que den sentido u orientación. Además, al descontextualizar los objetivos respecto de su trasfondo de práctica y de conocimientos compartidos en los que se utilizan, los defensores de los objetivos conductuales introducen un elemento artificial en la situación de enseñanza. Las listas de objetivos, fuera de contexto, hacen que las conductas de los profesores parezcan carentes de sentido. La sustitución de la conducta contextualizada por objetivos "artificiales" puede deformar gravemente las acciones.

Otras críticas del enfoque por objetivos son más prácticas. ¿Quién define, en realidad, los objetivos? ¿Qué intereses reflejan? ¿Constituyen los objetivos un conjunto completo de las conductas previstas? ¿Cómo pueden reducirse y medirse los objetivos? ¿Quién lo hace? ¿La especificación antecedente de los objetivos deja de lado resultados importantes?

Si nos detenemos a examinar la definición y especificación de los objetivos, a menudo descubriremos que han sido elegidos y especificados de forma arbitraria. Las listas de objetivos no suelen ser exhaustivas y los métodos para seleccionar los más importantes resultan sospechosos. La definición de los resultados previstos de un programa social es mucho más difícil de lo que parece.

Por ejemplo, en el *Michigan Accountability system* —una tentativa compleja y con fuerte respaldo financiero de definición de objetivos y de elaboración de tests basados en ellos y relacionados con los distintos grados escolares del estado de Michigan— la especificación de objetivos siempre fue de algún modo arbitraria. Para el cuarto grado de matemáticas se identificaron cuatrocientos objetivos. Sin embargo, la administración de tests que abarcasen todos esos objetivos habría consumido varias semanas de tiempo de los alumnos, así como cientos de miles de dólares para el desarrollo y administración de las pruebas. La lista de objetivos se redujo a treinta y cinco, más o menos, pero la selección fue arbitraria (HOUSE y cols., 1974).

En Michigan, las primeras pruebas de tipo test se efectuaron con tests referidos a una norma, elaborados por editoriales importantes, pero, como no se correspondían bien con los objetivos estatales, se elaboraron tests referidos a los objetivos. Se realizaron cinco ítems de test para medir cada objetivo. Respecto a estos tipos de medida existe una gran controversia.

© Ediciones Morata, S. L.

Unos dicen que los tests basados en objetivos son superiores mientras otros dicen que no son válidos (POPHAM y EBEL, 1978).

En su haber, los enfoques objetivos de evaluación gozan de gran validez de constructo. ¿Acaso un programa social no debe responder de aquello que afirma hacer? Después de todo, en ello se basa la financiación pública del programa. La evaluación tiene una legitimidad evidente cuando parte de los objetivos y metas declarados de un programa. Esto resuelve el dilema del evaluador de la forma más directa. A fin de cuentas, eso es lo que *dijeron* que harían. Esta afirmación relativa a lo enunciado resulta muy atractiva para los mismos responsables del desarrollo de un programa, sobre todo al principio del mismo, cuando tienen grandes aspiraciones en relación con sus trabajos.

Otra ventaja del enfoque por objetivos consiste en la gran facilidad de definición de su tecnología. Primero, el evaluador hace esto, después esto otro y así sucesivamente. Las fases de la realización se establecen con todo detalle (MORRIS y cols., 1978). Esta característica atrae al gran número de novatos que cada año lleva a cabo evaluaciones. Ningún otro enfoque de evaluación dispone de una tecnología tan elaborada. Por regla general, como los analistas de sistemas, los objetivistas conductuales siguen procedimientos "científicos", aunque no insistan en ello tanto como los analistas de sistemas.

Teniendo en cuenta todo, en una sociedad tecnológica, el enfoque por objetivos de la evaluación resulta muy atractivo, como pone en evidencia su omnipresencia. Se deriva, en realidad, de la tradición del análisis de tareas de la psicología industrial. Como ocurre con gran parte de la tecnología, sus dificultades radican tanto en sus dimensiones filosóficas como en las prácticas. Cuando pensamos seriamente sobre él o tratamos de aplicarlo, surgen ciertos defectos.

Decisión

Una gran ventaja del enfoque de decisión consiste en que se centra en la evaluación, importante cuestión práctica para el evaluador. En cualquier proceso de evaluación, podríamos recoger razonablemente una enorme masa de información, por lo que es preciso efectuar algún tipo de selección. El enfoque de decisión lo hace usando el criterio de utilidad. ¿Qué información será más útil? La respuesta del teórico de la decisión será que la información que resulte más útil para el responsable de las decisiones. Han de presentarse las alternativas de decisión, por lo que habrá que recoger la información pertinente al respecto.

Por ejemplo, en la evaluación del *Illinois Gifted Program*, el programa constaba de cinco partes (centros de prueba, centros de reinserción, programas de entrenamiento, proyectos de investigación y apoyo administrativo). Los centros de prueba y de reinserción absorbían más de cuatro millones de dólares anuales del presupuesto, desconociéndose los efectos de estos

© Ediciones Morata, S. L.

Scanned by CamScanner

proyectos, aunque se creía que los centros de prueba eran eficaces y los de reinserción no. Como ayuda para las decisiones previstas por la administración del Estado, el consejo asesor estatal y el legislativo del mismo, centramos nuestra evaluación en las áreas en las que más se invertía y más desconocidas, descubriendo que los centros de prueba eran menos eficaces y los centros de reinserción más de lo que se pensaba.

Esta información se tradujo en una reestructuración de los centros de prueba y en la aportación suplementaria de un millón de dólares a la parte de reinserción del programa estatal. En la evaluación, se previó que los tres grupos responsables tomarían decisiones respecto a una reorganización, reasignación presupuestaria y financiación básica. A los tres grupos principales se les entregaron versiones diferentes de la información de evaluación, preparándose una cuarta versión destinada al público en general. No se recabó información sobre los proyectos de investigación (que se creían ineficaces y se modificaban sin datos) ni sobre los programas de entrenamiento (sobre los que se disponía de otras informaciones, como las clases magistrales), como tampoco se trató información sobre las operaciones del personal estatal, dado que no veíamos cómo utilizar de forma razonable esa información (HOUSE y cols., 1971).

Por supuesto, diversas preguntas surgen de inmediato. ¿Por qué ha de darse tanta preferencia al responsable de las decisiones, que suele ser el administrador del programa? ¿Acaso esto no pone al evaluador al servicio de la alta dirección, convirtiéndolo en el "brazo armado" de la dirección del programa? ¿No hace esto que la evaluación sea en potencia desigual e, incluso, antidemocrática? La respuesta es que éstas son debilidades potenciales del enfoque de decisión.

El teórico de la decisión podría responder a estas críticas diciendo que el responsable de las decisiones es el único que, en realidad, puede modificar las cosas, empleando de nuevo la idea básica de la utilidad de la información. A largo plazo, hemos de admitir que, aunque esta utilidad sea un criterio importante de la evaluación, no puede ser el único. No querríamos una evaluación útil si fuese también incierta o desigual. La utilidad debe equilibrarse en relación con otras consideraciones.

Como cuestión práctica, hay diversas formas de prevenir que los poderes decisorios utilicen la evaluación de manera poco ética. Una de ellas consiste en definir de modo más amplio los grupos decisorios. Otra es proteger a los informadores de las represalias o acciones de las autoridades superiores. Aún hay otra, consistente en proporcionar la misma información a grupos que no formen parte de los círculos decisorios. No obstante, el problema sigue siendo que la información proporcionada a los responsables de las decisiones imprime a la evaluación un fuerte sesgo a favor de la dirección. No puede sorprender que este enfoque sea el más popular entre los administradores.

Por ejemplo, en la *Illinois Gifted Evaluation*, los responsables superiores de las decisiones quedaron muy satisfechos de la evaluación. No obstante, podríamos preguntarnos por los participantes en los programas y las personas

© Ediciones Morata, S. L.

situadas en los distritos escolares. ¿Quedaron bien reflejados sus intereses y preocupaciones? No existe una respuesta automática a esta pregunta. Es preciso examinar la evaluación concreta y sus circunstancias. Pero la situación del evaluador, del brazo del responsable de las decisiones, siempre nos deja perplejos.

Otro conjunto de cuestiones surge cuando el evaluador trata de definir las decisiones específicas a las que contribuirá. Las decisiones no suelen expresarse mediante alternativas claras y, de hecho, sucede a menudo que las decisiones para las que se utiliza la evaluación no surgen hasta después de efectuada. Asimismo, las decisiones cambian de forma espectacular con el tiempo. Con frecuencia, las que determinan la suerte de un programa no son las previstas por el evaluador. En otras palabras, especificar y prever las decisiones que dependan de la evaluación es mucho más difícil de lo que parece.

Por ejemplo, el enfoque de GUTTENTAG-EDWARDS utiliza un complejo procedimiento de escala para llegar a un conjunto de preferencias del responsable de las decisiones, pero el procedimiento depende en gran medida de cómo definan las alternativas los evaluadores. En el sondeo de preferencias sólo se incluyen los responsables máximos de las decisiones. Podríamos preguntarnos por qué hay que otorgar tal peso a las preferencias subjetivas del responsable de las decisiones. El enfoque de PATTON restringe la búsqueda del usuario más probable de la información y se plantea cuáles son las preguntas "correctas" para ese usuario. (Los datos deberían recogerse de acuerdo con esas preguntas; para ellas debe haber más de una respuesta; los responsables de las decisiones deben creer que las pueden emplear y deben poder especificar su forma de utilización.) Todos éstos son criterios de utilidad.

A pesar de estas dificultades, el enfoque de decisión proporciona una valiosa visión de la evaluación. Hace hincapié en la importancia de la utilidad de la información. Supone que se empleará la información de la evaluación. La conexión entre la evaluación y la decisión subraya el objetivo de la primera. También es útil, desde el punto de vista práctico, configurar la evaluación de acuerdo con consideraciones concretas de decisión. Aunque no podamos definir de manera precisa las alternativas de decisión, podemos eliminar una serie de líneas de investigación por carecer de relieve. Esta tarea constituye una disciplina beneficiosa para quienes la abordan desde una perspectiva de investigación. En otras palabras, el enfoque de decisión subestima la evaluación como actividad *práctica*. El peligro estriba en convertirse en instrumento del responsable de la decisión.

Independencia de los objetivos del programa

El enfoque de independencia de los objetivos es uno de los que más se habla y menos se utiliza en la evaluación de programas sociales. Su propósito consiste en reducir los sesgos, sobre todo los que aparecen a causa del

© Ediciones Morata, S. L.

Scanned by CamScanner

conocimiento que el evaluador tenga sobre los objetivos del programa y de su relación con el personal que lo desarrolla. Es obvio que el sesgo aparece merced al contacto social. El evaluador debe comprobar todos los resultados del programa y no sólo los pretendidos por el personal que lo lleva a cabo. La *Consumers' Union* evalúa con regularidad productos de consumo sin tener en cuenta los objetivos e intenciones del fabricante, sin embargo no es habitual que este enfoque se utilice en programas sociales.

La primera dificultad consiste en la falta de una tecnología específica de la evaluación. En cierto sentido, la "atención a todos los resultados" se opone a un enfoque programado. SCRIVEN considera al evaluador como un cazador, un detective, un investigador armado con diversas técnicas, pero no con una tecnología de paso fijo, como la que ofrece el enfoque por objetivos conductuales.

El auténtico razonamiento empleado en la evaluación que prescinde de los objetivos del programa resulta un tanto extraño a los evaluadores. En su mayoría, están entrenados en una lógica de medios a fines: "He aquí los objetivos. ¿Se han cumplido?" El razonamiento que prescinde de los objetivos es más pragmático: "He aquí el programa. ¿Cuáles son sus efectos?" La idea tradicional de objetividad de los evaluadores formados en las ciencias sociales es cuantitativa. Para ellos, la objetividad significa emplear medidas repetibles, o sea, que tengan una fiabilidad elevada. La idea de objetividad, cuando no se tienen en cuenta los objetivos, es cualitativa: independencia de las diversas influencias que pueden sesgar el propio juicio.

Hasta cierto punto, SCRIVEN ha tratado de responder a la demanda de una metodología construyendo listas de comprobación de lo que ha de cumplir un programa. Los críticos replican que se limita a sustituir los objetivos del programa por sus propios criterios, sus propios objetivos ocultos, cosa que él niega. Para entender su razonamiento, conviene pensar en la evaluación de un producto físico, como un coche. Para evaluarlo, podemos especificar las funciones que debe llevar a cabo y valorar el coche de acuerdo con esas características. Debe ser barato, de bajo consumo de combustible, sin demasiadas necesidades de reparación, etcétera. La comprobación física de estas cualidades no tiene nada de arbitrario ni subjetivo, y el evaluador no necesita conocer las intenciones del fabricante para evaluar el coche. Podemos aplicar el mismo razonamiento a los programas sociales. Sin embargo, ¿acaso un programa social es igual que un producto físico? ¿Pueden determinarse sus funciones sin referencia a las intenciones del programa? En cierto modo sí y en cierto modo no. (HOGBEN y SIMPSON, 1978, mantienen una postura que se opone a la de SCRIVEN y presentan como alternativa la evaluación basada en la teoría.)

Estas dificultades se reflejan en las evaluaciones concretas. WELCH (1978) utilizó un tribunal externo para juzgar unos materiales didácticos elaborados en un proyecto de desarrollo curricular. Los criterios utilizados fueron criterios de producto. El proyecto fue juzgado según la calidad de sus productos, pero los creadores no estaban contentos con la evaluación porque no reflejaba los procesos del mismo ni el progreso y desarrollo efectuados por

© Ediciones Morata, S. L.

los participantes en el proyecto y por la institución. Los cambios que se produjeron en el personal del centro y en los participantes que desarrollaron los materiales no se reflejaban en la calidad de los productos. Tanto los evaluadores como los participantes en el programa estaban en desacuerdo con la metodología. No se tenían en cuenta los múltiples destinatarios de la evaluación ni la posibilidad de que algunos de ellos valoraran de forma distinta los productos.

La segunda dificultad, aparte de la metodología, está constituida por la interacción social. Los resultados de los programas de acción social y, en consecuencia, su conducta, son muy ambiguos y están sometidos a interpretaciones muy diversas. En la vida diaria, cuando tratamos de interpretar y comprender los actos ambiguos de otros —caso habitual y necesario—, lo hacemos por medio del concepto de persona que subyace a esas acciones. Suponemos que existe una estructura estable que denominamos "persona" y esa estructura sirve de contexto para la interpretación de las acciones. Sabemos lo que significa un acto concreto porque lo interpretamos como derivado de la persona que conocemos (PERELMAN y OLBRECHTS-TYTECA, 1969).

Una característica primordial de la persona es su intención. Interpretamos un acto sobre todo a través de las intenciones que suponemos tiene una persona. El mismo acto puede estructurarse de forma diferente, según sea nuestra presunción de esas intenciones. Los juicios morales se basan en las intenciones y moralidad del agente, frente a la moralidad del acto. Como las intenciones se suponen, sólo podemos escuchar las expresadas por el agente y observar los actos correspondientes. De hecho, algunos filósofos creen que la ambigüedad de la conducta, interpretada por necesidad de acuerdo con la intenciones, es lo que distingue las ciencias humanas de las ciencias naturales.

La importancia de la evaluación que prescinde de los objetivos del programa consiste en que: impide que el personal que participa en el mismo manifieste sus intenciones a los evaluadores. Esto supone que los evaluadores juzguen las acciones con independencia del contexto y de las intenciones de la persona. Además, el personal del programa desconoce los intereses del evaluador, que es en todo momento un personaje enigmático. La ruptura de este contrato social tanto por parte del evaluador como del personal del programa acaba con la posibilidad de la evaluación. Sólo las personalidades más seguras pueden aguantar esta tensión.

Así, aunque el argumento a favor de la evaluación que prescinde de los objetivos del programa pueda tener sentido, el evaluador tiene ante sí una metodología desconocida y una situación de valoración ambigua, y los evaluados se enfrentan a la perspectiva de una evaluación no amistosa a cargo de un extraño con quien no pueden comunicarse. Pocas evaluaciones independientes de los objetivos del programa que se evalúa se han llevado a cabo. Es significativo que en la única que yo he realizado, la persona que contrató la evaluación y la puso en marcha desde dentro era un buen amigo mío con quien había trabajado durante años. Aún así, yo estaba

© Ediciones Morata, S. L.

Scanned by CamScanner

preocupado sobre cómo proceder y el personal que participaba en el programa se mostraba inquieto por mi desconocimiento de sus intenciones.

Teniendo en cuenta todos los aspectos, no se equivoca SCRIVEN respecto al sesgo que supone estar informado sobre las intenciones del programa. El sesgo es más profundo de lo que parece. Sin embargo, puede que la confraternización sea necesaria, desde el punto de vista social, para que una evaluación pueda siquiera llevarse a cabo. Mis tentativas posteriores de realizar evaluaciones independientes de los objetivos han sido rechazadas por los posibles clientes.

Antes de que puedan llevarse a cabo evaluaciones independientes de los objetivos de los programas que evaluar, han de cumplirse ciertas condiciones. O bien se establece de antemano una relación de confianza, o las relaciones de autoridad son tales que el personal del programa no puede impedir la evaluación. Esto puede ocurrir si los evaluadores y el personal del programa permanecen alejados entre sí, como la *Consumers' Union* y *General Motors;* o si ambos están sometidos a una autoridad superior. En otro caso, las tácticas dirigidas a establecer unas relaciones de confianza pueden llevar consigo la utilización de otros enfoques junto con la evaluación independiente de los objetivos. En este libro, la evaluación se estructura como proceso social y la evaluación independiente de los objetivos pone de manifiesto hasta qué punto es cierto.

Crítica de arte[1]

Merece la pena considerar en serio este enfoque como forma de evaluación. La idea básica es atrevida e imaginativa. Hacen falta otros métodos de evaluación que complementen la evaluación científica de los programas sociales, y las largas tradiciones críticas que han favorecido las artes parecen constituir un rico manantial de ideas. Sin embargo, las dificultades que se plantean para el establecimiento de la crítica educativa como forma legítima de evaluación son formidables.

Sus virtudes consisten en que el evaluador puede basarse en su propia experiencia y en su razonamiento intuitivo para juzgar lo que ocurre y manifestar estos juicios en un lenguaje y con unos conceptos inteligibles para los no versados. Las dificultades se plantean a la hora de justificar estos juicios y controlar los posibles sesgos que surjan en la evaluación.

Según la concepción de EISNER (1979), la crítica incrementará la conciencia y la apreciación y consta de tres aspectos: descriptivo, interpretativo y evaluativo. En el primero, se da cuenta de las "cualidades pertinentes de la vida educativa". En el segundo, se "explican" los hechos y sus consecuencias por medio de las teorías de las ciencias sociales. En el tercero, se juzgan los procesos educativos de acuerdo con criterios educativos. Este último distingue, según EISNER, al crítico del científico social.

[1] Este apartado se basa en Ernest R. HOUSE y Rochelle S. MAYER: "A Critique of the Educational Imagination in Evaluation": *Journal of Aesthetic Education,* julio, 1980.

© Ediciones Morata, S. L.

Es curiosa la mezcla de conceptos y teorías de las ciencias sociales en la forma crítica de evaluación. Es fácil darse cuenta de cómo puede el crítico "exponer" el ambiente de clase mediante categorías artísticas, así como evaluarlo, pero la forma de introducir los conceptos de las ciencias sociales constituye un auténtico rompecabezas. Los ejemplos de crítica educativa que brinda el libro de EISNER no sirven al efecto porque las categorías evaluativas son muy personales en vez de científicas.

Los dos procedimientos para determinar la validez son: la medida en que la crítica "configura una totalidad coherente y persuasiva" y la comprobación empírica de la crítica respecto al fenómeno. Otra idea interesante —la auténtica cuestión no consiste en que sea verdadera, sino útil— sólo queda apuntada, pero no se estudia. No es difícil imaginar una crítica "útil" que no cumpla criterio alguno de validez. Teniendo en cuenta todos los aspectos, la discusión de la validez es incompleta y debe revisarse. No obstante, para hacer justicia a EISNER, hay que decir que la validez constituye un problema primordial en todos los enfoques de evaluación.

EISNER reconoce las limitaciones de la analogía entre la crítica de arte y la evaluación cuando indica que podemos examinar una escultura o una pintura para ver lo que ve el crítico, dado que la obra de arte no cambia con el tiempo. Sin embargo, una clase sí cambia. La conclusión de EISNER es que ésta no cambiará en cuanto a sus características clave. Quizá la postura del evaluador fuese algo distinta si el paradigma artístico utilizado fuera el teatro, la música o la literatura en vez de las artes visuales.

Hay también una diferencia entre la evaluación de un aula y lo que suele hacer un evaluador. Por regla general, éste valora un programa completo en vez de una única clase. Quizá hubiera que establecer la analogía con la revisión de una exposición de arte en vez de con una única pintura. Además, a pesar del argumento de EISNER sobre la estabilidad de la clase (cuestión en la que está en lo cierto), el evaluador de programas se ocupa más de los cambios que de la estabilidad.

La cuestión de la validez nos lleva a preguntas más profundas de la evaluación. ¿Qué criterios emplea el evaluador? ¿En qué valores se fundan? ¿Cómo se justifican? EISNER sostiene que el evaluador-crítico debe aplicar "criterios educativos", pero no especifica en qué consisten. Indica que los criterios provienen de un examen de la historia y la filosofía de la educación, así como de la experiencia práctica. Él plantea dos preguntas: ¿se ayuda a los niños?, ¿adquieren hábitos que conduzcan a posteriores desarrollos? Son, en efecto, dos buenas preguntas y las respuestas dependerán en gran medida de la propia teoría de la educación y de las preferencias personales. EISNER aborda el problema de forma directa: "Sin una visión de lo que sea la educación, nadie está en condiciones de hacer juicios sobre la calidad educativa del proceso de enseñanza o de sus consecuencias" (EISNER,

© Ediciones Morata, S. L.

Scanned by CamScanner

1979, pág. 44). Es evidente que el mismo EISNER piensa en la educación como desarrollo (KOHLBERG y MAYER, 1972).

¿Qué ocurre con quienes no tienen una visión semejante de la educación? ¿Estarán sometidos a los valores del evaluador? En cuanto crítico y experto, ¿tiene el evaluador autoridad para emitir juicios basándose en su dominio de la cuestión? Si es así, ¿pone de manifiesto los fundamentos de sus juicios de valor? EISNER se muestra ambivalente respecto a estos problemas, pero su resolución es fundamental en relación con la forma de actuar del crítico. En nuestra propia evaluación, parece que resultamos personalmente creíbles a los administradores del programa, tras unas horas de entrevista. ¿Acaso el patrocinador tenga que encontrar un crítico compatible? ¿Y qué ocurre con las personas evaluadas?

Los tres ejemplos de crítica educativa que aparecen en el libro de EISNER nos invitan a detenernos en ellos. Los llevan a cabo sus alumnos y revisten un carácter marcadamente judicial. El primero comienza de este modo:

> *Esta clase es casi una caricatura de la sociedad.*
> *El curriculum* está servido como los *Big Macs*. Lectura, matemáticas, lengua, incluso la educación física y afectiva están precocinadas, preenvasadas, con sabor artificial.
>
> (Pág. 229.)

El segundo ejemplo asume la forma narrativa en primera persona de un alumno de elemental, pero la narración está muy marcada por los valores del crítico. El tercero llega a la conclusión de que el profesor va a abandonar la enseñanza a causa de las presiones ejercidas por los padres respecto a la estructura del *curriculum*, aunque no queda claro si se han estudiado explicaciones alternativas. Los tres ejemplos proponen los intereses de los estudiantes en contra de los presuntos valores de eficiencia de los padres y de la sociedad.

No es raro que un crítico de arte presente puntos de vista controvertidos (el lector puede optar por dejarlos de lado). De hecho, el lector puede escoger sólo las críticas con las que esté de acuerdo. Sin embargo, una evaluación pública de un programa educativo no puede desestimarse con tanta facilidad. Es preciso justificar de alguna manera al crítico, los principios de éste o la crítica misma. Las exigencias de equidad y justicia son más rigurosas en la evaluación de programas públicos.

El problema final del enfoque crítico de la evaluación es el siguiente: ¿quién es el destinatario de la crítica? Leemos la crítica de una exposición o de una obra para decidir si vamos a verla o no. Leemos la crítica de una obra de arte para aumentar nuestros conocimientos y apreciaciones. En el caso que nos ocupa, ¿de quién es la apreciación que se mejora? La respuesta directa de EISNER consiste en que la crítica se dirige a los profesores a quienes puede resultarles útil ese *feedback* sobre su ejercicio

© Ediciones Morata, S. L.

docente. No obstante, los ejemplos que propone suscitan dudas respecto a que ese tipo de críticas resulte aceptable a los profesores. Por supuesto, las críticas efectuadas para su presentación en tesis doctorales pueden diferir considerablemente de las llevadas a cabo en situaciones reales de evaluación.

En resumen, la idea de la evaluación como crítica es fructífera y merece que se estudie y, en potencia, es importante. El libro de EISNER constituye el punto de partida lógico de tal ejercicio. Desde un punto de vista práctico, nuestra propia crítica educativa se desarrolló bien. Aunque haya que tratar con mayor detenimiento las difíciles cuestiones relativas a la validez, los valores y la justificación, no son insuperables. Habrá que ir más allá de estas formulaciones iniciales.

Revisión profesional

A medida que se profesionalizan más ocupaciones laborales, los procedimientos de revisión cobran mayor importancia. En el seno de las profesiones se mantiene la idea de que sólo quienes pertenecen a la misma están cualificados para juzgar las actividades de sus iguales. Fuera de cada una de ellas, casi nadie está de acuerdo con tal afirmación, aunque sí se defienda respecto de la propia.

Desde hace algún tiempo, se observa en el público en general la tendencia a negar a los profesionales el derecho a controlar sus propios asuntos. Los usuarios demandan un puesto en los consejos directivos de los centros de salud, en los consejos asesores y en los órganos de dirección de las escuelas. Existe la creencia de que los profesionales no vigilan sus propias operaciones con suficiente atención. Suele creerse que los médicos ni se oponen en privado ni testifican en público contra otros médicos.

La desconfianza pública ante las evaluaciones controladas por los propios profesionales se refleja en la disminución de la credibilidad de los organismos encargados de la acreditación profesional. En otras épocas, bastaba que una institución recibiera la correspondiente acreditación del organismo competente para que el público considerase garantizada su calidad, pero ya no. Los padres no siempre están convencidos de que el programa escolar sea de gran calidad porque esté reconocido por la *North Central Association*. Además, el control político de las actividades de reconocimiento y acreditación va pasando a manos de los gobiernos estatales. Las revisiones no profesionales, como las de NADER, son cada vez más frecuentes, aunque presenten sus propios problemas.

En efecto, el público tiene razón cuando cuestiona algunos procedimientos de evaluación. Las visitas de los equipos de revisión para la concesión de la acreditación son muy variables y su rigor depende de los miembros del equipo. Con frecuencia, son escogidos por la institución que va a ser evaluada. Los propios organismos consideran a veces la visita como una intrusión no deseada e incluso ilegítima en sus asuntos y, en ocasiones, utilizan

© Ediciones Morata, S. L.

Scanned by CamScanner

la situación con fines de relaciones públicas más que como una oportunidad de efectuar una auténtica revisión interna. Si la evaluación no se toma en serio, el organismo acreditador tendrá muy poco poder sobre la institución.

Por otra parte, a consecuencia de la presión pública a favor de la rendición de cuentas, está aumentando el número de revisiones profesionales. Muchos departamentos estatales de educación y organismos estatales de educación superior están llevando a cabo sus propias visitas de inspección en vez de dejarlas en manos de instituciones privadas. Aunque los procedimientos de revisión suelen ser semejantes a los de instituciones privadas de acreditación, la amenaza de la desaprobación oficial y la consiguiente supresión de ayudas financieras se toman más en serio.

Otro ejemplo es la popularidad de los sistemas de revisión profesional interna en las universidades y centros de enseñanza superior. Las universidades siempre han evaluado a los individuos a efectos de promoción o de ejercicio, pero los nuevos sistemas evalúan programas, departamentos y unidades administrativas. Muchos sistemas de este tipo se pusieron en marcha para soslayar la intervención directa de los organismos de la enseñanza superior, así como para tomar decisiones internas y reasignar fondos en tiempos difíciles. La actitud generalizada consiste en pensar que es mejor que la misma institución efectúe evaluaciones de sus unidades y no que las lleve a cabo un organismo externo. Muchas autoridades de la enseñanza superior utilizan procedimientos de revisión profesional.

No obstante, podemos suponer que los criterios utilizados por los profesionales en el seno de una institución diferirán sustancialmente de los empleados por organismos públicos de supervisión, como los de la educación superior. A veces, se cuestionan tanto el contenido como el rigor de la evaluación. Por ejemplo, el *Council on Program Evaluation* (COPE) de la Universidad de Illinois, en la que permanecí durante tres años, utilizaba como criterios para evaluar los departamentos la calidad de los programas de enseñanza, la investigación, el servicio, la cooperación con otras unidades, el valor respecto al Estado y la sociedad y las previsiones futuras.

Sin embargo, en las deliberaciones concretas del *Council* la calidad de la investigación, puesta de manifiesto en las clasificaciones nacionales y similares, solía primar sobre los demás criterios cuando se evaluaba un departamento de la universidad. Sospecho que esto mismo sucedería en todas las principales universidades de investigación. Aunque se tuviesen en cuenta los indicadores de los demás criterios, nada podía compensar un programa endeble de investigación y nada se opondría a uno importante, desde el punto de vista del *Council*.

Podríamos pensar que un organismo estatal de supervisión de enseñanza superior, al ser responsable de la asignación de fondos a las universidades estatales, estaría relativamente más interesado por los costes y la producción de créditos que por la calidad de la investigación. Aunque ambos grupos utilizaran los mismos criterios, los pesos relativos a ellos asignados diferirían de modo espectacular en los juicios finales. No es difícil imaginar cómo los

© Ediciones Morata, S. L.

profesores se considerarían a sí mismos como los mejores jueces de la calidad de los departamentos, mientras que el organismo supervisor se consideraría garante de los intereses públicos.

Incluso entre los profesores existen grandes discrepancias respecto a quién debe y puede emitir juicios. En el ámbito de actuación del COPE, algunos departamentos, como el de física, pensaban que los profesores que pertenecían a otros distintos carecían de competencia para juzgar el suyo. Algunas de las más duras batallas políticas se libraron en relación con la competencia del *Council* para juzgar a otros departamentos. Los defensores de los procedimientos del COPE sostenían que se trataba de actividades, criterios y competencias comunes a todos los profesores en cuanto tales. Aunque las personas no pertenecientes a un departamento pudieran no estar capacitadas para juzgar el contenido de la investigación, los miembros del *Council* tenían competencia para juzgar cómo consideraban la calidad de la investigación otros profesores de física de diversas partes del país. También eran competentes, como docentes, para juzgar la calidad de la enseñanza, etcétera. La evaluación interdisciplinar de los departamentos siguió adelante, aunque se llevara a cabo a veces tras largas y extensas negociaciones, y por la insistencia en ello de la administración superior de la universidad.

En la realización de revisiones profesionales, tienen gran importancia los procedimientos concretos que se utilicen. Por su medio, se determina si la evaluación es equitativa. Por ejemplo, en los primeros tiempos del sistema del COPE, se revisaron docenas de departamentos sin que se hicieran recomendaciones relativas a la disolución de ninguna unidad. De repente —o así les pareció a los afectados—, el Council recomendó la eliminación del *College of Communication,* asignando algunos de sus componentes a unidades distintas y disolviendo otras.

Esta acción despertó un enorme furor. El rector recibió más correspondencia que en ninguna otra ocasión desde que se eliminó la separación de sexos en la residencia universitaria. Durante el período de matrícula, quienes se oponían a la medida recogieron diez mil firmas de estudiantes en contra de la misma. Lo que unía a los opositores no era tanto el fondo del informe de evaluación como el tratamiento injusto deparado a la unidad en cuestión. Otras unidades habían recibido con anterioridad informes negativos, pero ninguna había sido amenazada de disolución. No había precedentes ni advertencias al respecto, ni tampoco se le había dado oportunidad a la unidad amenazada para defenderse antes de efectuarse la recomendación. Tanto en el enfoque de revisión profesional como en el cuasijudicial de la evaluación, las bases de la equidad vienen dadas por el procedimiento adecuado y el proceso debido. El *College of Communication* aún sigue allí.

La confidencialidad plantea otro problema. Para infundir confianza e informar públicamente, debe emitirse un comunicado público también. Sin embargo, los profesionales son reacios a que sus actividades salgan por completo a la luz pública. Por regla general, como en el caso del COPE, se redactan dos informes, uno interno, de carácter confidencial, en el que

© Ediciones Morata, S. L.

Scanned by CamScanner

se ponen de manifiesto las miserias y defectos —el informe "real"—, y otro público, "maquillado" de alguna manera. Parece que esta duplicidad de informes es necesaria para la cooperación profesional, pero es obvio que hace que el público desconfíe.

El sistema de evaluación de una universidad del oeste, que evaluaba los departamentos y publicaba los informes correspondientes de forma semejante a un periódico estudiantil, pone de relieve este aspecto de manera espectacular. Era previsible que este sistema crease un malestar considerable entre los profesores. Llegó el momento de elegir un nuevo rector de dicha universidad. El principal tema de discusión durante las entrevistas de los profesores con los candidatos a rector consistió en las actitudes de los mismos respecto al sistema de evaluación. El método desapareció. Por regla general, tanto en el plano profesional como en el personal, aceptamos la crítica pública de nosotros mismos mucho menos de lo que pensamos.

Cuasijudicial

La fuerza del enfoque cuasijudicial de la evaluación reside en que incorpora los mecanismos y la autoridad de la ley. Su debilidad está en que también hace suyos algunos de sus inconvenientes. En cierto sentido, la correspondencia entre los procedimientos judiciales reales y los cuasijudiciales no es buena. Por ejemplo, no existe cuerpo legal que permita decidir en terrenos como el educativo. Sin embargo, los procedimientos cuasijudiciales han demostrado su valor para dilucidar cuestiones públicas controvertidas.

Los procedimientos cuasijudiciales de evaluación están constituidos por los tribunales de expertos en los certámenes, la mayor parte de los procedimientos de audiencia y los simulacros de juicios. Una importantísima ventaja consiste en que los problemas públicos candentes pueden tratarse con rapidez mediante el nombramiento de una comisión o tribunal. Las personas involucradas en el asunto de forma más significativa pueden ser llamadas a declarar y presentar sus puntos de vista. La participación puede ser muy amplia, incluyendo grupos que quedarían excluidos en la mayoría de los demás enfoques. Un atractivo fundamental del cuasijudicial es su potencial apertura a opiniones diversas.

Por ejemplo, en el tribunal de audiencia patrocinado por la *National Education Association* para evaluar el *Michigan Accountability System,* declararon más de veinte grupos, desde la *League of Women Voters* hasta la asociación de profesores de artes. Para tomar declaración se siguieron procedimientos acordados de antemano (HOUSE y cols., 1974). La audiencia adoptó las formas de los procedimientos judiciales: el tribunal se acomodaba tras una mesa frente a los testigos, las declaraciones se grababan en cinta magnetofónica, los documentos escritos se anotaban y clasificaban, etcétera. En estas audiencias, los símbolos desempeñan un papel no despreciable.

© Ediciones Morata, S. L.

Otro elemento crítico es la constitución de la comisión, tribunal o jurado. En el caso de Michigan, el tribunal estaba compuesto por tres conocidos evaluadores. Sin embargo, estos componentes del tribunal fueron elegidos por la *National Education Association,* que, en cierto sentido, era parte en el caso. En la investigación sobre evaluación de tests de la *Florida Competency* (TYLER y cols., 1978), se siguió un procedimiento de selección semejante. En ambos casos, se suscitó el problema de la credibilidad. La identidad del tribunal y sus procedimientos se equilibra con el modo de seleccionarlo y lo que pueda suponer de ganancia o pérdida, es decir, en estos temas, su imparcialidad. La cuestión de la imparcialidad no tiene que ver con que los miembros del tribunal hayan expresado opiniones previas sobre cuestiones similares. Si son expertos, seguro que lo habrán hecho. El problema estriba en si sus propios intereses entrán en juego y en cómo esté constituido el tribunal en su totalidad. En los juicios simulados, los jurados han estado constituidos hasta ahora por profesionales que, de algún modo, participan con el público.

Algunos evaluadores (POPHAM y CARLSON, 1977) han criticado mucho los mismos juicios simulados, o procedimientos contradictorios, como se los denomina a veces. La evaluación contradictoria se basa en la idea de que dos defensores presenten las versiones opuestas sobre la cuestión. Los críticos achacan que es muy raro que ambos defensores tengan igual habilidad, por lo que el caso lo ganarán los más dotados, con independencia de los méritos intrínsecos de cada opción.

Otra crítica se refiere a que los jueces son falibles en extremo. Sin embargo, en educación y en otros campos sociales, no existen tribunales de apelación que puedan corregir los errores de procedimiento. Una tercera crítica alude a que el modelo contradictorio no es eficaz para esclarecer la verdad respecto a una cuestión. Aunque pueda resolver conflictos, su potencial ilustrador es limitado. Ramsey Clark, antiguo fiscal general, lo dijo con la máxima claridad: "Si hay algún procedimiento peor para llegar a la verdad, no sé cuál es."

Otra crítica se refiere a que los problemas sociales no permiten el tipo de formulaciones de "sí o no" que requiere el modelo judicial. Los problemas son demasiado complejos y suponen gradaciones más matizadas de los juicios. Asimismo, responsables de decisiones poco escrupulosos pueden hacer uso del enfoque para confirmar y justificar lo que ya hubieran decidido. Pueden dejarse persuadir por los argumentos de menor peso en un juicio simulado. Por último, los críticos dicen que los procedimientos contradictorios son extremadamente caros, exigiendo a menudo los gastos correspondientes a la recogida de datos y al juicio.

En defensa de los procedimientos contradictorios (recordando que no todos los enfoques cuasijudiciales son de carácter contradictorio), THURSTON (1978) dice que el juicio con jurado tiene tres ventajas principales: la publicidad del juicio pone de manifiesto el problema de que se trate; al menos, pueden defenderse dos aspectos de un problema, y puede presentarse cualquier tipo de prueba, debiendo hacerse de manera comprensible. Dice

© Ediciones Morata, S. L.

Scanned by CamScanner

THURSTON que se ha exagerado la diferencia de habilidades de los contendientes y la falibilidad de los jueces. En todo caso, cualquier evaluador que utilice algún enfoque es falible. También pueden utilizarse mal otros puntos de vista y costar demasiado. Además, los métodos estadísticos de evaluación no son más adecuados para llegar a la "verdad" que los enfoques contradictorios.

Afirma THURSTON que, en los procedimientos contradictorios la definición de los problemas y la utilización del jurado constituyen problemas más graves. En terrenos sociales, como la educación, no hay tradiciones ni precedentes de contextualización de problemas ni hay acuerdo sobre las normas, como ocurre con las jurídicas. En un caso judicial concreto, los abogados conocen las normas. Sin embargo, ante los tribunales de justicia no todas las cuestiones son blancas o negras. Pueden ser más complejas.

Según THURSTON, otro dilema importante radica en el carácter y papel del jurado. Los jurados educativos pueden desenvolverse mejor si sus miembros son individuos que dominan el tema y se les permite deliberar sobre pruebas ajenas al juicio, prácticas que las normas judiciales no permiten. Teniéndolo todo en cuenta, desde su punto de vista como jurista e investigador educativo a la vez, THURSTON reconoce ciertas dificultades de los procedimientos contradictorios, pero manifiesta que pueden paliarse si se toman como modelos los tribunales de apelación o los procedimientos administrativos en vez de los más conocidos juicios ante los tribunales de justicia.

En resumen, podemos decir que los procedimientos cuasijudiciales, como las audiencias y tribunales, constituyen formas útiles de efectuar evaluaciones, dependiendo de la naturaleza del problema de que se trate. Las cuestiones públicas controvertidas que deben tratarse con rapidez y que requieren el juicio entre diversas opciones reclaman la actuación de tribunales constituidos por personas, que sean o no profesionales, que escuchen a todas las partes antes de resolver la cuestión (en la medida en que una controversia de este tipo pueda resolverse). La actuación de estos tribunales es rápida y da una sensación de atención y de resolución. Por supuesto, como los miembros de los tribunales suelen escogerse entre personas de renombre en el asunto de que se trate, las conclusiones suelen reflejar la postura del estamento dirigente.

Por ejemplo, el jurado que evaluó el *Michigan Accountability System* suscitó un interés considerable y atrajo la atención pública sobre las características de ese sistema. Como consecuencia de las audiencias públicas y de la amplia difusión del informe del jurado, se modificaron diversas características del sistema de Michigan. Asimismo, otros muchos estados que estudiaban programas similares, revisaron en consecuencia sus propios planes. Los resultados de la evaluación del tribunal recibieron el apoyo de la misma comunidad profesional de evaluadores. Por tanto, las repercusiones de dichas audiencias, sobre el público en general y sobre el estamento profesional, fueron muy grandes. Es probable que ningún otro enfoque hubiese suscitado tanta publicidad ni interés.

© Ediciones Morata, S. L.

Por otra parte, la eficacia de los juicios de este tipo se presenta como más cuestionable. Parece que adolecen de graves problemas, aunque quizá no tantos como afirman muchos críticos. Es probable que los juicios a gran escala sean relativamente caros y exijan una preparación larga, aunque lo mismo ocurre con otras evaluaciones. Los juicios (y demás procedimientos cuasijudiciales) permiten la participación abierta de todos los individuos del público. Sin embargo, no hay garantía de que se participe, aunque se dé la ocasión para ello. Muchos procedimientos de este tipo se desarrollan para grupos reducidos de destinatarios. Quizá los juicios a gran escala debieran limitarse aún a casos deliberadamente experimentales. De momento, parecen más eficientes otros procedimientos cuasijudiciales.

Estudio de casos

El enfoque del estudio de casos se está haciendo cada vez más popular, aunque, como los demás enfoques subjetivos, es difícil establecer su credibilidad en una comunidad predominantemente científica. Algunos teóricos dicen que, en realidad, es científico, pero los estudiosos de casos más serios niegan que dos trabajos distintos lleguen a descubrir lo mismo.

Por ejemplo, en un estudio de casos de un sistema de enseñanza basado en ordenador, GJERDE y yo hicimos hincapié en las características técnicas y políticas (HOUSE, 1974). No cabe duda de que esas peculiaridades estaban presentes, pero unos evaluadores distintos podrían haber seleccionado otras igualmente pertinentes, dando al estudio de casos una perspectiva muy diferente. La mayoría de los problemas en relación con el estudio de casos se refiere más a la importancia que se otorgue a unas u otras cuestiones que a la verdad o falsedad de algo. Distintos observadores referirán hechos diferentes. Aunque estas cuestiones se hayan discutido a menudo, los principales problemas que atañen a la credibilidad de los estudios de casos son los relativos a la consistencia metodológica y a la interpretación.

En su haber, el estudio de casos proporciona información rica y persuasiva que no facilitan otros enfoques. Pueden reflejarse todos los aspectos, desde las personalidades de los participantes hasta las perspectivas de otros sujetos muy alejados del programa. Esto confiere fuerza y utilidad a la información. Nuestro estudio de casos pasó con toda rapidez de mano en mano de los administradores de la oficina central, tomándose decisiones acerca de los sistemas de enseñanza basados en los ordenadores.

La presentación de los hechos de manera tan personal suscita también problemas de confidencialidad, equidad y justicia. Es difícil desdibujar a individuos y personalidades de forma que sea imposible su identificación. Como nunca llegamos a tener una imagen completa de los participantes, también es difícil representarlos con exactitud. Todo estudio de casos plantea esos problemas, que nunca son tan graves en formas de evaluación más impersonales.

© Ediciones Morata, S. L.

Scanned by CamScanner

Un aspecto muy positivo del estudio de casos consiste en que permite la representación de distintos puntos de vista y de intereses diferentes. En este sentido, puede que sea uno de los enfoques más democráticos. Dependiendo de la destreza y del esfuerzo del evaluador, la evaluación puede ser extremadamente democrática en cuanto a la representación de los intereses de las partes implicadas, en contraste con el enfoque del análisis de sistemas, que tiende a excluir los distintos puntos de vista e intereses. Si llevamos a cabo un trabajo razonablemente decente para localizar y entrevistar a los diversos grupos que participan en un programa, se reflejará la mayor parte de los intereses. Como he señalado en otro lugar de este libro, quienes con mayor frecuencia quedarán fuera de este enfoque pluralista serán los menos favorecidos y menos poderosos. Por supuesto, el evaluador *puede* excluir distintos intereses. El enfoque del análisis de sistemas exige un esfuerzo deliberado para no hacerlo así.

La representación de la diversidad de intereses conduce a otro formidable problema: ¿cómo puede el evaluador equilibrar y resolver la divergencia de intereses? Los teóricos del estudio de casos no se ponen de acuerdo al respecto. Unos creen que los evaluadores deben equilibrar los intereses de acuerdo con su propio sentido de justicia. En particular, los antropólogos tienen una larga tradición de "hacerse indígenas", de situarse entre los desheredados. Sin embargo, esto hace que se cuestione la imparcialidad de la evaluación. Por regla general, los evaluadores no apelan a principios, sino que se basan en sus juicios intuitivos.

Otra postura sostiene que el evaluador debe permanecer al margen de los intereses. No debe favorecer ningún interés concreto en la evaluación, sino tratar todos por igual. Desde este punto de vista, el evaluador representa y refleja con neutralidad, dejando que el lector sopese y equilibre los diversos intereses a su juicio. El evaluador no lo hace. Un exponente cualificado de esta postura diría que la información contenida en el estudio de casos debe ser de igual utilidad para todos los intereses. Un acceso igual al conocimiento dará a los destinatarios iguales oportunidades para influir en la práctica educativa. En consecuencia, el evaluador debe ser neutral, de oficio, con respecto a los distintos destinatarios (ELLIOTT, 1979). No obstante, esto supone que las personas pueden hacer uso real de la información. Podría razonarse que los más poderosos tienen más probabilidades de hacerlo que los que ostentan menos poder. En otras palabras, esta postura neutraliza al evaluador, pero no las relaciones de poder vigentes. En la práctica, el desinterés favorece al poderoso.

Una cuestión relacionada con la anterior es si el evaluador debe plantear recomendaciones explícitas a partir del estudio o si esto debe dejársele al lector, o sea, si el evaluador debe sacar conclusiones por su cuenta. Me parece que ambas posturas son aceptables y que la opción dependerá de los destinatarios. Es cierto, como afirman quienes sostienen esta postura, que el evaluador quizá no conozca bastante las circunstancias del programa para poder extraer conclusiones acertadas. Sólo en el caso de que las alternativas evaluadas sean idénticas a las recomendaciones comprendería

© Ediciones Morata, S. L.

bastante el evaluador las circunstancias concretas. Por otra parte, no creo que las recomendaciones del evaluador intimiden a los lectores si no son acertadas.

En la vida moderna, se dan muchas circunstancias en las que estaríamos más dispuestos a leer y aceptar las recomendaciones que a leer toda la evaluación y extraer nuestras propias recomendaciones. Por ejemplo, yo estoy dispuesto a aceptar los consejos de la *Consumers' Union* sobre muchos productos de consumo sin leer todas las evaluaciones de batidoras eléctricas o de aparatos de radio de onda corta. Si tuviéramos que leer cada evaluación en la que se funda cada una de las recomendaciones sobre cuya base actuamos en la vida moderna, emplearíamos en ello demasiado tiempo, aún en el caso de que pudiéramos comprender la evaluación y extraer nuestras propias recomendaciones. Obligar al lector a que saque sus propias conclusiones y recomendaciones en cada una de las áreas en las que tiene que actuar es una idea interesante y democrática, pero muy poco práctica si la aplicamos a todos los campos especializados de la sociedad moderna. En ésta, parece inevitable cierta especialización a la hora de decidir. Me parece que el estudio de casos debe presentar al lector la información que le permita juzgar el fundamento de las conclusiones y recomendaciones, si desea hacerlo.

En relación con esta cuestión está la afirmación que la mayoría de sus exponentes hacen respecto de los estudios de casos. Afirman que éstos son más útiles que otros tipos de evaluación. Y creo que, en cuanto al tipo de información que proporciona un estudio de casos, la cual nos permite conocer mejor las interioridades de un programa, algo tiene de verdad. Pero la afirmación de que los estudios de casos son más útiles en todas las situaciones es muy exagerada. Dichos estudios siempre son largos y están muy inmersos en la situación correspondiente. El lector no tiene tiempo de leer un estudio para preparar cada decisión. Es preciso aceptar ciertas transacciones en aras de la eficiencia; y la creencia de que los estudios de casos son más útiles, por regla general, es más un artículo de fe que un hecho empírico.

Otro problema: no todo el mundo puede dedicarse a hacer estudios de casos. Su redacción constituye una tarea difícil y sólo algunas personas tienen el talento suficiente para hacerla. Parece fácil, igual que parece fácil escribir una novela, y muchos están tentados de emprenderlos cuando no les atraen otros enfoques. Por desgracia, la falta de orientaciones, críticas y procedimientos conducen a menudo a un trabajo de baja calidad. Y un estudio de casos malo es, en efecto, malo. La capacidad de centrarse en las cuestiones pertinentes y de ponerlas en perspectiva de modo que no sea una exposición escrita de las opiniones personales no son cualidades universales. Quienes se dedican a hacer estudios de casos necesitan una preparación más rigurosa de la que ahora reciben. Por supuesto, otros enfoques también son susceptibles de aplicaciones de mala calidad.

A pesar de estas advertencias y reservas, el enfoque del estudio de casos es uno de los más prometedores y dignos de desarrollo. A igualdad

© Ediciones Morata, S. L.

Scanned by CamScanner

de circunstancias, con frecuencia prefiero este enfoque. Si es creíble para los destinatarios a quienes se dirige, un estudio de casos bien estructurado constituye la evaluación más potente. Tiene la posibilidad de ser persuasivo, preciso, coherente y justo, en cuanto representación de diversos puntos de vista en situaciones complejas. Por otra parte, no es ninguna panacea y plantea de por sí un conjunto de problemas característicos. Ni su teoría ni su metodología se han desarrollado por completo.

© Ediciones Morata, S. L.

CAPÍTULO XII

La realización de evaluaciones válidas

En este libro, el concepto de validez es clave. En sentido amplio, entiendo por "validez" algo así como: "cualidad de merecer el reconocimiento". La definición del diccionario es: "la cualidad de estar bien fundado en los hechos o basado en principios sólidos y completamente aplicables al caso o circunstancias; solidez y fuerza (de razonamiento, prueba, autoridad, etc.)" *(Oxford E.D.)**. El concepto de validez que he aplicado a la evaluación es mucho más amplio que la idea tradicional de validez como previsión, aunque la incluye.

La práctica moderna de la evaluación se considera, con razón, como un procedimiento de decisión social. En el Capítulo Primero de este libro, delineamos tres situaciones básicas de evaluación: personal, interpersonal y pública. En la situación personal, el evaluador y el destinatario de la evaluación coinciden. Como mínima exigencia de validez, preveríamos que la evaluación fuese veraz. Existen diferentes ideas y métodos para llegar a la verdad.

En la situación interpersonal, el evaluador trabaja al servicio de los destinatarios, pero tanto él como aquéllos son personas o entidades privadas, como en las evaluaciones de la *Consumers' Union*. En este caso, no sólo pediríamos que la evaluación fuese veraz, sino también creíble para los destinatarios. Éstos deben confiar en ella. Hay varias formas de demostrar la credibilidad y hacerla digna de confianza.

En la tercera situación básica, el evaluador valora un programa público para unos destinatarios externos. En este caso, no basta con que la evalua-

* En español, "validez" se define como la *"calidad de válido"*. A su vez, "válido" (o "válida") se define como *"firme, que vale legalmente"*, y en segunda acepción: *"robusto, fuerte o esforzado"*. Podemos decir, por tanto, que, en líneas generales denota lo mismo que el término inglés *"validity"*. En sentido estadístico, la "validez" se refiere a la correspondencia de la medida o apreciación con la magnitud real que se mide o aprecia. *(N. del T.)*

© Ediciones Morata, S. L.

Scanned by CamScanner

ción sea veraz y creíble; también debe ser correcta, desde el punto de vista normativo, pues ahora la evaluación ha pasado de ser un asunto privado a un hecho público. El tercer caso es el único al que se enfrenta la mayoría de los evaluadores. Ellos tienen que hacer frente a la triple exigencia de validez que requiere que la evaluación sea veraz, creíble y ajustada a derecho. De nuevo, hay varias formas de establecer la corrección normativa de la evaluación. La tarea a la que se enfrenta el evaluador es formidable porque un fallo en cualquiera de los tres aspectos invalida la evaluación.

Cada uno de los ochos enfoques de la evaluación hace una reivindicación de validez, afirma "ser merecedora de reconocimiento". En cada afirmación de validez se revela, como en taquigrafía, el haber y el debe de cada enfoque. El enfoque del análisis de sistemas afirma que es científico, que, de acuerdo con procedimientos explícitos, produce información fiable, "datos puros y duros". Al centrarse de forma tan exclusiva en el aspecto de veracidad de la validez, con frecuencia estas evaluaciones no son creíbles para los evaluados y, en ocasiones, resultan antidemocráticas, desiguales o incorrectas, desde el punto de vista normativo.

El enfoque por objetivos conductuales afirma su validez de otra manera, basándose en la idea de la técnica. Trata de determinar la evaluación especificando los resultados que se pretenden y definiendo, a continuación, los medios para alcanzar esos fines. Los métodos son especificaciones del campo y de la conducta que hay que observar. La validez se deriva de que el programa dé cuenta de sus objetivos especificados de antemano. El enfoque no contempla métodos para juzgar la corrección de los objetivos.

El enfoque de decisión reclama su validez en virtud de su utilidad para los responsables de las decisiones. La utilidad de los datos recogidos constituye la consideración primaria, de manera que la credibilidad de la evaluación ante sus destinatarios sea alta. Debemos preguntarnos, una vez más, si el servicio prestado al responsable de las decisiones es suficiente para asegurar su corrección normativa. El evaluador que prescinde de los objetivos del programa que evalúa sostiene la validez de su evaluación por su falta de sesgos. La protección respecto de influencias inadecuadas lo erige en juez objetivo de las consecuencias del programa. No obstante, con frecuencia se cuestiona su credibilidad porque los destinatarios externos y quienes son evaluados se preguntan a menudo si en realidad aprecia lo que hace el programa.

En el Capítulo III, denominé "objetivistas" a estos cuatro enfoques por su epistemología. En mayor o menor grado, todos basan su protesta de validez en métodos explícitos. El método primordial de recogida de datos es la observación. La réplica constituye un criterio clave y se consigue mediante procedimientos exteriorizados y explicitadores, de modo que diversos observadores puedan atestiguar los mismos hechos. Se valora mucho la fiabilidad de la medida. La validez suele evaluarse por la réplica o por la concordancia entre instrumentos. Su fuerza está en el rigor del método, al que estos enfoques prestan gran atención. Se valora mucho la posibilidad de comprobar las hipótesis. Su debilidad estriba en que se centran en el

© Ediciones Morata, S. L.

aspecto de veracidad de la validez, pasando por alto la base normativa de lo que hacen. Si esto constituye una debilidad en el desarrollo de la ciencia, en la evaluación es un defecto fatal.

En general, los enfoques objetivistas de la evaluación se derivan de la tradición teórica empírica de las ciencias sociales. La evaluación se considera como una ciencia social aplicada. Los defensores de la teoría empírica sostienen que las ciencias sociales se diferencian sólo en grado, pero no en cuanto clase, de las ciencias naturales. Sólo existen dos modelos de conocimiento justificado: las ciencias naturales y las disciplinas formales, como la lógica y las matemáticas. Todo lo que no cumpla las normas de conocimiento fijadas por estas disciplinas es sospechoso (BERNSTEIN, 1978).

Un punto de vista muy extendido y simplificado consiste en que la base de todo conocimiento empírico es el ámbito de los hechos "puros y duros", no interpretados. Las afirmaciones empíricas se justifican en virtud de esos hechos. Por tanto, la ciencia consiste en recoger datos y efectuar generalizaciones a partir de ellos. Podemos suponer que el evaluador de este tipo investigará los efectos de los programas, utilizando los métodos de las ciencias sociales para garantizar la objetividad y, en consecuencia, la validez de la evaluación; la validez se basa en la utilización de la metodología prescrita.

En esta perspectiva, es fundamental la distinción categórica entre teoría y práctica. El científico o el evaluador descubre relaciones empíricas y este conocimiento se aplica en la práctica. La metodología explícita conduce a un conocimiento explícito, validado. La acción es la aplicación técnica de este conocimiento validado. En consecuencia, el papel adecuado del investigador, como el del científico natural, se caracteriza por el "desinterés". Es neutral, desde el punto de vista axiológico. Por medio de su metodología, descubre los hechos, pero no los critica ni los traduce a la práctica. Corresponde a otros, en especial a los responsables de las decisiones, extraer consecuencias para la acción. En cierto sentido, la idea de mercado libre de las ideas y la de que cada uno saque sus propias conclusiones y derivaciones son intrínsecas a la filosofía del liberalismo.

Son muchos los críticos de esta epistemología objetivista. Los que aquí hace al caso es que se centra en el aspecto de veracidad de la validez, dejando de lado la credibilidad y los aspectos normativos. Supone la validez por su metodología y sólo resulta creíble a quienes abogan por esa metodología. Con frecuencia, no es creíble para los evaluados, a menos que crean en la "ciencia" como fuerza legitimadora. Por ejemplo, en la evaluación del *Follow Through*, las personas evaluadas pensaban que ésta no examinaba resultados pertinentes, aunque se había llevado a cabo con todo rigor.

La segunda crítica consiste en que el enfoque objetivista encubre valores implícitos de los que no tienen conciencia los investigadores. Por ejemplo, los tests normalizados de rendimiento no son axiológicamente neutrales, sino que se basan en ciertas premisas de valor, como la de elevar al máximo las diferencias individuales. Los conceptos, categorías y técnicas,

© Ediciones Morata, S. L.

Scanned by CamScanner

como la categorización de "destrezas básicas" de la evaluación *Follow Through*, se inclinan a favor de ciertos valores. El hecho de que los evaluadores objetivistas no sean conscientes a menudo de los sesgos de sus técnicas y supongan la neutralidad axiológica de las mismas convierte esas evaluaciones en potencialmente peligrosas. En realidad, el empleo de esas técnicas presupone ciertos axiomas de base, como un sistema ético y un sistema de justicia que indica cómo han de determinarse las pretensiones e intereses. Estas cuestiones se dan por supuestas. Se parte de la base de que la información del investigador se transfiere al presunto sistema de decisión social como, por ejemplo, a individuos que utilizan la información para satisfacer sus propias aspiraciones.

Una idea de objetividad más compleja que la de correspondencia de la teoría con un campo de hechos no interpretados es la de normas intersubjetivas de racionalidad o normas de indagación mediante las que se trata de eliminar diversas formas de sesgos y deformaciones (BERNSTEIN, 1978). Hasta cierto punto, la evaluación de SCRIVEN que prescinde de los objetivos del programa evaluado utiliza una idea ampliada de objetividad y constituye una excepción respecto a la crítica general de los enfoques objetivistas. En realidad, va mucho más allá y sostiene que el evaluador puede determinar "necesidades" objetivas que existen con independencia de las preferencias de los destinatarios. Por tanto, su postura es objetivista en la medida en que el evaluador no sólo determina de manera objetiva la información derivada del proceso, sino posiblemente también las mismas normas de evaluación. Los valores son hechos y el evaluador tiene derecho a emitir juicios de valor sobre el programa. Por regla general, ni evaluadores ni filósofos comparten este punto de vista.

Los otros cuatro enfoques —el que sigue el estilo de la crítica de arte, la revisión profesional, el cuasijudicial y el estudio de casos— son "subjetivistas" porque fundan su presunta validez en la apelación a la experiencia y no al método científico. Suponen que el saber es en gran medida tácito más que explícito. En el enfoque de estilo crítico artístico, el fundamento consiste en la experiencia del profesionnal. Tienen gran importancia su preparación y sus credenciales. La validez de la evaluación depende de sus percepciones. La enseñanza y la misma actividad social se consideran artes en las que los fines se realizan en los medios, sin separarlos como sucede en la técnica (ELLIOTT, 1979). El método del crítico consiste en la percepción holística. Establece la relación entre las partes y el conjunto de forma integrada, elaborando de este modo sus juicios. La clave de la validez reside en su habilidad para hacerlo. El punto débil de este enfoque está en que es fácil que distintos críticos lleguen a críticas del programa muy diferentes.

El enfoque de revisión profesional se funda en la experiencia colectiva, tradicional de la profesión. El juicio profesional, basado en las normas profesionales, hace que la evaluación merezca la atención de los destinatarios. El trabajo manual y no el arte ni la técnica constituye el modelo de la profesión. Existen reglas, conocimientos y paradigmas interiorizados mediante los cuales puede juzgarse el ejercicio profesional. Por tanto, la validez

© Ediciones Morata, S. L.

depende del acatamiento de tales reglas de competencia aceptadas de forma intersubjetiva, aunque tácita. Sólo el profesional que haya participado de la cultura profesional puede haber adquirido este conocimiento y ser un juez competente. Conviene señalar que los enfoques de crítica de arte y de revisión profesional gozan de gran credibilidad entre los destinatarios en ejercicio. No ocurre lo mismo con los científicos positivistas, que consideran que el método científico "invalida" las reglas de la práctica, y entre el gran público, que teme que los fuertes intereses profesionales pesen más que el interés general.

El enfoque cuasijudicial se fundamenta en la experiencia de la profesión jurídica, adoptando procedimientos "procesales" desarrollados en el ámbito jurídico inglés. El grado de adaptación de la evaluación a los procedimientos adecuados constituye una medida de su validez. La clave consiste en la equidad del procedimiento. El fundamento de la validez es la experiencia judicial, tal como se estructura en la tradición jurídica. Si los procedimientos son equitativos, con independencia de los resultados, la evaluación será válida.

Por último, el enfoque del estudio de casos se basa en la experiencia de los participantes y de los destinatarios. Un aspecto primordial que se debe tener en cuenta es la comprensión por parte de los receptores. El evaluador consigna la experiencia de los participantes de manera que los destinatarios puedan entenderla. La validez depende de la correspondencia entre la evaluación y las experiencias de los participantes y de los destinatarios.

Es obvio que los subjetivistas también afirman que sus evaluaciones son veraces, pero tratan de conseguir una visión perspicaz en el marco de referencia de los destinatarios y de los mismos participantes. Es importante la significación. La evaluación ha de ser comprensible. Los subjetivistas asumen que cada investigador y cada lector son capaces de establecer conclusiones de forma individual. Estos enfoques tienen a su favor la característica de comunicar con frecuencia perspectivas importantes. En su contra tienen que sus conclusiones son muy variables e incluso contradictorias y que no es fácil reconciliarlas. Lo que es válido para una persona puede no serlo para otra.

La epistemología subjetivista y, en especial, el enfoque del estudio de casos, están íntimamente relacionados con la visión fenomenológica del hombre. En contraste con el punto de vista de que sólo la ciencia es la medida de la realidad y la norma del conocimiento justificado, los fenomenólogos sostienen que es más fundamental la propia visión cotidiana del mundo, centrada en las personas y en su conducta. Sobre todo si pensamos en la ación humana, debemos hacerlo en términos de intenciones y de las personas que orientan su conducta de acuerdo con principios y normas basados en los valores comunitarios. Es el dominio de la razón práctica más que de la teórica (BERNSTEIN, 1978).

De modo similar, los enfoques subjetivistas de la evaluación apelan a la experiencia en la situación concreta en cuanto fundamento de validez. La experiencia puede manifestarse en el experto, en una profesión, en la tradi-

© Ediciones Morata, S. L.

Scanned by CamScanner

ción que acoge puntos de vista encontrados o en los participantes y destinatarios concretos de la evaluación. La exactitud de la descripción de la situación y la utilidad para los destinatarios son conceptos clave. Es decir, la evaluación será "adecuada" si es comprensible para el actor y refleja la realidad social en la que vive.

Los críticos de la epistemología fenomenológica señalan que con frecuencia surge la confusión respecto a qué percepciones de sentido común han de constituir el fundamento de la comprensión. Es más, si tomamos como base de indagación el conocimiento cotidiano, ¿no nos limitaremos a reconstruir cualesquiera ideologías, sesgos y falsas creencias ya existentes? ¿Cómo podemos distinguir los determinantes causales y las regularidades —los puntos fuertes de la epistemología positivista— de las creencias percibidas? ¿Cómo podemos evaluar las interpretaciones opuestas? La fenomenología no proporciona medio alguno para hacerlo; tiene algo de adivinanza, pues lo que tomamos como determinantes causales de la acción social depende de nuestras propias teorías relevantes. Por tanto, la interpretación depende, en parte, de nuestro mismo análisis causal (BERNSTEIN, 1978).

Por último, tanto para los fenomenólogos como para los positivistas, el papel ideal del investigador es "desinteresado" y alejado de los intereses del mundo cotidiano. Los positivistas establecen sus regularidades empíricas y los fenomenólogos dilucidan estructuras subjetivas fundamentales, para llegar a la pura descripción. Si puede acusarse a los positivistas de reflejar de modo encubierto los sesgos institucionales, a los fenomenólogos puede achacárseles que reflejan los sesgos de quienes describen. En ambos casos, al menos a los efectos de la evaluación, el papel del evaluador "desinteresado", alejado de todos los problemas prácticos, es defectuoso, desde el punto de vista moral.

El evaluador debe ser imparcial. Esto no es lo mismo que ser indiferente ante los intereses que se reflejen o estar fuera del mundo real. El evaluador está comprometido con el mundo. Su trabajo afecta de forma directa a lo que cada uno pueda conseguir. Por naturaleza, participa de su función social. La evaluación debe ser imparcial en el sentido de que estén representados todos los intereses pertinentes. Esto debe constituir una preocupación permanente del evaluador, que no cumpliría con sus obligaciones morales si se aislara de los intereses externos. En la ciencia liberal existe una tradición que sostiene que el aislamiento de las influencias del mundo garantiza la veracidad y la equidad, pero no cabe duda de que esto es erróneo. La indiferencia ante los intereses que afloren no es lo mismo que representar los intereses de todos.

En resumen, la validez de una evaluación depende de que sea veraz, creíble y correcta, desde el punto de vista normativo. Todos los enfoques de la evaluación aseguran su validez a su modo respectivo, pero la utilización de un enfoque concreto no la garantiza. Cuanto más se corresponda la situación de evaluación con las premisas de un enfoque determinado, más probable será su validez. En la mayoría de los casos, ésta se identifica con

© Ediciones Morata, S. L.

la veracidad, pero no basta para evaluar un programa social. Como demuestra el contraste entre las situaciones personal, interpersonal y pública, el evaluador tiene obligaciones adicionales.

No cabe duda de que una evaluación no veraz no sería válida, y la verdad puede establecerse de muchos modos. La postura positivista equipara la veracidad con la réplica y la predicción. Si conocemos la causa de algo, podemos observarlo y preverlo con exactitud. Sin embargo, la veracidad abarca mucho más de lo que reconoce la concepción positivista. Los criterios de verdad dependen del acuerdo intersubjetivo de la comunidad interesada y cambian con el tiempo. En la situación individual, personal, podemos imaginar la búsqueda de la verdad como empeño individual, aunque no sea así.

En la situación interpersonal, descubrir los hechos no es suficiente. Es preciso comunicarlos a un público determinado. La evaluación ha de ser creíble, de manera que inspire la confianza del público. La credibilidad es función tanto del evaluador como de la evaluación. La credibilidad del evaluador se refuerza cuando no entran en juego sus propios intereses. Por ejemplo, dudaríamos de una evaluación de automóviles si la *General Motors* pagase al evaluador, con independencia de la calidad de su trabajo. La evaluación ha de ser auténtica, en el sentido de que merezca la aceptación o la confianza por su concordancia con los hechos o la experiencia conocidos. Los objetivistas hacen hincapié en la concordancia con los hechos y los subjetivistas insisten en la concordancia con la experiencia. En ambos casos, la evaluación debe suscitar la confianza.

Por regla general, si el público comprende y cree las intenciones del evaluador, se reforzará la confianza en éste. La manifestación de las intenciones sirve como signo de sinceridad. Con frecuencia, ésta se juzga por la coherencia de las acciones del evaluador. En el mismo informe de evaluación, la coherencia, el tono y otros elementos estéticos analizados en el Capítulo V influyen en la credibilidad y autenticidad. Por encima de todo, el evaluador debe dar pruebas de que es digno de confianza.

Sin embargo, todo esto no es suficiente en la evaluación de programas públicos. Con total exactitud y sinceridad, podemos favorecer el bienestar de un grupo, dañando gravemente el bienestar público. El criterio definitivo no puede consistir en el servicio o la utilidad a un grupo concreto. Una evaluación pública debe ser correcta, desde el punto de vista normativo. Junto a la verdad y la belleza debe ir la justicia, aunque, de nuevo, esto se conciba de formas diferentes. En la Tercera Parte, he tratado de indicar cuáles deben ser estas consideraciones normativas. La evaluación pública debe ser democrática, equitativa y, en último término, basada en los valores morales de equidad, autonomía, imparcialidad y reciprocidad. No obstante, ésta sólo es una formulación inicial de tales consideraciones.

Ningún enfoque de la evaluación, ningún método garantizará de antemano la validez. Tenemos que mirar hacia atrás y examinar la evaluación concreta en su situación para ver si es válida. En la mayoría de los casos, resultarán adecuados diversos enfoques de evaluación y el evaluador puede escoger

© Ediciones Morata, S. L.

Scanned by CamScanner

uno sobre la base de sus propias preferencias o las de su cliente. Muchas evaluaciones serán mezclas de varios enfoques. Teóricamente, el evaluador debe estar preparado para utilizar varios enfoques y no aplicar de forma mecánica el que haya aprendido. Debe conocer los puntos débiles de sus enfoques preferidos con el fin de evitar aquello que ponga en peligro su validez.

Una forma de hacer patentes los intereses del cliente y de las partes afectadas consiste en el acuerdo de evaluación equitativa. Constituye una salvaguardia contra la conducta autocrática del evaluador o del cliente. Puede llegarse a un enfoque o combinación de enfoques de evaluación de manera moralmente aceptable. Por desgracia, la idea contractual implícita en el acuerdo de evaluación equitativa no siempre protege en grado suficiente los intereses de quienes no participan en el establecimiento del acuerdo.

El defecto más frecuente en las evaluaciones es de orden moral. Una sociedad liberal se ve a sí misma compuesta por unidades independientes y autónomas que sólo cooperan cuando se favorecen los fines de las partes. En una sociedad de este tipo, es fácil que se conciba la evaluación como una cuestión privada o interpersonal, como un acuerdo entre dos partes en beneficio mutuo, sin tener en cuenta a los demás. A menudo se pasan por alto las obligaciones con respecto a la sociedad en general. La evaluación de un programa público constituye una situación en la que existen tales obligaciones y el concepto de validez debe ampliarse de manera que pueda aplicarse a esta significativa práctica social.

© Ediciones Morata, S. L.

APÉNDICE A

Un análisis de la lógica de una evaluación

Formas de razonamiento

En este apartado, analizaremos la estructura argumentativa de un informe de evaluación. El informe *"Educational Product Evaluation"*, de GLASS (1972) corresponde a la evaluación de unas grabaciones magnetofónicas elaboradas por Michael SCRIVEN. Tras estudiar el informe, analizamos también la estructura lógica de la respuesta de SCRIVEN (1972) a dicha evaluación. No obstante, en primer lugar es preciso establecer las categorías de análisis, las formas de razonamiento que emplean ambos evaluadores.

No hay duda de que, en ciertas circunstancias de la evaluación, es aplicable la lógica formal. Por ejemplo, la lógica deductiva es apropiada para determinar la coherencia interna de los modelos matemáticos y la lógica inductiva está indicada en los problemas de inferencia estadística. En las circunstancias apropiadas, este razonamiento puede ser aplicado. Sin embargo, la mayoría de los evaluadores han de basarse en modelos de razonamiento no formales. Enumeraré algunas técnicas de argumentación basadas en el tratado de PERELMAN y OLBRECHTS-TYTECA (1969) sobre este tema. La relación no es en absoluto exhaustiva de los poderes de razonamiento informal del hombre. En el apartado siguiente, mostraré el uso de estas argumentaciones mediante el análisis de un estudio de evaluación que goza de gran aceptación.

Las técnicas de argumentación que presentamos aquí se dividen en tres clases: argumentos cuasilógicos, argumentos basados en la estructura de la realidad y argumentos que establecen la estructura de la realidad. La primera clase, la de los argumentos cuasilógicos, deriva su credibilidad de su semejanza con el razonamiento lógico formal o matemático. No obstante, el argumento cuasilógico sólo parece formal por reducción. Este argumento es esencialmente más informal que formal y, en último extremo, se sostiene recurriendo a otras formas de argumentación.

© Ediciones Morata, S. L.

Scanned by CamScanner

El primer tipo de argumentación depende de su semenjanza con las relaciones lógicas. Estas relaciones son las de contradicción e incompatibilidad, identidad y definición, transitividad y reciprocidad. El otro grupo de argumentos cuasilógicos depende de su semejanza con el razonamiento matemático, cuyas relaciones son las de inclusión de la parte en la totalidad, división de la totalidad en partes, comparación y razonamientos de probabilidad.

Incompatibilidad. En un sistema lógico, dos tesis en contradicción mutua demuestran que el sistema es lógicamente inconsistente. El análogo cuasilógico es la incompatibilidad, en la que nos vemos obligados a escoger entre dos tesis que son incompatibles desde el punto de vista práctico, aunque no desde el lógico, a causa de las circunstancias. En casos extremos, el mantenimiento de tesis incompatibles puede caer en el ridículo, equivalente argumental del absurdo lógico. Por ejemplo, en una evaluación, el director del proyecto puede presentar una determinada visión del mismo, mientras que un profesor que trabaja en él presenta otra perspectiva muy diferente. Ambos puntos de vista lógicamente no son contradictorios, pues pueden ser ciertos considerados desde circunstancias diferentes. Sin embargo, la incompatibilidad puede constituir una cuestión importante en el conjunto de la evaluación. De hecho, el director cuyo punto de vista es incompatible con los de otros participantes en el proyecto empieza a parecer ridículo.

Identidad total y definición. En la medida en que las definiciones se enuncien de forma inequívoca y carentes de ambigüedad, pertenecen a sistemas de lógica formal. Tan pronto se aplican a problemas del mundo real, las definiciones se convierten en cuasilógicas. Tenemos que escoger entre muchos significados posibles. Sólo los sistemas exclusivamente convencionales pueden evitar estos problemas de identidad. Por ejemplo, la validez se define de cinco formas, al menos, desde una justificación general hasta la capacidad de prever un hecho a partir de otro. Podemos utilizar cualquiera de las definiciones, pero, si se discute, hemos de defenderla como adecuada y aplicable.

Identidad parcial. La "regla de justicia formal" requiere que se otorgue idéntico tratamiento a seres o situaciones del mismo tipo. Esto proporciona coherencia a la acción, fundamento de la justicia formal. La "reciprocidad" de la conducta se basa en la definición de situaciones simétricas. Estos argumentos exigen reducciones parciales, como en relación con el prestigio y la categoría de las partes implicadas, que dependen de las posturas discutidas. Por ejemplo: "Era realmente equitativo que la profesora proporcionase una ayuda especial al niño, pues ya había prestado ayuda extraordinaria a otros". Más discutible sería: "Merecían notas iguales, pues habían realizado el mismo esfuerzo, aunque con resultados muy distintos". Estos enunciados se basan en definiciones de identidades parciales.

© Ediciones Morata, S. L.

Transitividad. A es mayor que B y B, mayor que C; por tanto A es mayor que C (pero el fundamento de "mayor que" es discutible). Por ejemplo: "El programa A es mejor que el B porque las puntuaciones en el test son más altas. A debe ser mejor que C porque los resultados de B en el test son mejores que los de C." Por supuesto, los criterios de comparación son discutibles, como también la transitividad de la relación. Aunque la primera relación se mantenga, el programa A puede no ser mejor que el C.

Los argumentos basados en la semejanza con el razonamiento matemático son los siguientes:

Inclusión de la parte en la totalidad. La totalidad es mayor que cada una de las partes. Por ejemplo: "Obtener una puntuación total superior en el test es mejor que una puntuación elevada en una de las partes, porque la puntuación total incluye las parciales."

División de la totalidad en partes. La división exhaustiva en partes conduce a la conclusión de que la parte que queda es de algún modo necesaria. "Haré una relación de los sesgos a favor y en contra del estudio." "En ambos casos tenemos un error de Tipo I o un error de Tipo II."

Comparación. La comparación directa de objetos se basa en la idea de medida, pero toda norma de medida es defectuosa. A menudo, se mencionan criterios. La elección supone siempre comparación. El "razonamiento del sacrificio" es una forma de comparación: ¿qué sacrificio estaríamos dispuestos a hacer para conseguir determinado fin? Quizá toda evaluación sea básicamente comparativa.

Probabilidades. El argumento de probabilidad y variabilidad suele llevar consigo una reducción de los datos a valores monistas y homogéneos, y a elementos mediante los cuales compararlos. Pero suele ser un argumento potente porque infunde un carácter empírico, aun cuando no sea cuantitativo (p. ej., la teoría de la decisión, que requiere que ésta se reduzca a un modelo concreto de decisión).

Una clase de argumentos totalmente distinta se basa en la "estructura de la realidad". Se parte de una concordancia suficiente respecto a la realidad, como los hechos y verdades, y de su carácter indiscutido, de modo que tratamos de establecer una conexión entre las ideas aceptadas y las que promovemos. Estos razonamientos pueden clasificarse con más detalle bien como relaciones de sucesión, que vinculan un fenómeno con sus causas o consecuencias, bien como relaciones de coexistencia, que enlazan una "esencia" con sus manifestaciones, p. ej., una persona con sus acciones. Entre las relaciones de sucesión, en las que el tiempo constituye un factor principal, están las siguientes:

Causalidad. La demostración de relaciones causales puede basarse en métodos muy diferentes y, como es obvio, desempeñan un papel esencial

© Ediciones Morata, S. L.

Scanned by CamScanner

en la argumentación evaluativa. Al tratar de establecer una relación causal, hay que relacionar dos hechos sucesivos, razonando desde un hecho dado hacia su presunta causa o proyectando una consecuencia causal a partir de un hecho. En todo caso, el enunciado causal requiere ciertos juicios de valor (véase ENNIS, 1973).

Pragmatismo. Un hecho se evalúa por sus consecuencias. El valor de éstas se transfiere a la causa. O se acepta el valor de las consecuencias o hará falta recurrir a otros argumentos para establecerlo.

Fines y medios. La determinación de los medios mejores depende de la definición precisa del fin que se busca y del acuerdo respecto al mismo. Sólo es probable que se discutan los valores relacionados con el fin. En una sociedad de orientación tecnológica, los razonamientos fines-medios tienen una fuerza especial. Por ejemplo, los programas por objetivos conductuales son buenos si consiguen los fines previstos. La separación analítica de medios y fines facilita una concordancia máxima, aunque es dudoso que unos medios determinados produzcan un único efecto. En la práctica, los medios y los fines están mucho más relacionados entre sí.

Pérdida. Dado el esfuerzo realizado hasta cierto momento, interrumpirlo ahora constituiría una pérdida. "Sería una lástima no reanalizar estos datos, teniendo en cuenta lo que ha costado reunirlos." "Desarrollad al máximo el talento del niño."

Dirección. Si cedemos en este momento, ¿qué es lo que esto ocasionará? La teoría del dominó. "El saber puede incrementarse de forma indefinida. El aprendizaje no tiene límites."

Desarrollo ilimitado. Más es mejor y puede lograrse.

Aunque las relaciones de sucesión están en el mismo nivel fenomenológico, las relaciones de coexistencia conectan dos objetos o hechos de los que uno es más básico y explicativo que el otro. El orden de los hechos tiene una importancia secundaria. Estos hechos son los siguientes:

La persona y sus actos. Las acciones de una persona suelen influir en nuestro concepto de ella, aunque, por regla general, no se equiparen como en el conductismo. En los estudios de evaluación, la interpretación de un hecho adscribiéndolo a la personalidad es práctica corriente. La forma de tratar las "intenciones" de la persona resulta especialmente crítica. Con frecuencia, la intención se infiere por la correspondencia entre acciones, pero siempre existe ambigüedad. La mayoría de las atribuciones de motivación constituyen un ejemplo de este tipo de razonamiento.

Autoridad. Aunque se excluye por completo de las demostraciones lógicas, pues la lógica debe sostenerse por sí misma, en la argumentación

© Ediciones Morata, S. L.

es importante el prestigio de la persona que hace una afirmación. En el razonamiento judicial es esencial. Sólo si un público universal acepta la afirmación, considerándose, por tanto, como un "hecho", supera el dominio de la autoridad.

La "objetividad" se logra a menudo separando a la persona de su acto, p. ej., retirando el nombre del autor de las propuestas antes de juzgarlas. Sin embargo, la persona puede ser el mejor elemento de predicción del éxito del proyecto. La imparcialidad puede conseguirse mediante técnicas de reducción de los sesgos mejor que por la escisión completa del agente de su acto (SCRIVEN, 1975). En la argumentación y en la evaluación, la relación entre la persona y su afirmación es importante.

Persona y grupo. "Hizo eso porque es conductista." En esta categoría se incluyen los argumentos que expresan la preocupación por mantener o establecer relaciones con otros. En la evaluación, la caracterización de una persona por su pertenencia a un grupo es mucho más corriente de lo que parece. No sólo se llevan a cabo estudios cuantitativos para poner de manifiesto diferencias entre grupos; a menudo, las evaluaciones cualitativas interpretan el sistema social que se estudia como un conjunto de grupos interactivos. Además, con frecuencia el evaluador se ve en apuros para demostrar su preocupación, imparcialidad o ambas poniendo de manifiesto a qué grupos pertenece o no.

Acto y esencia. ¿En qué consiste ser un buen director? Un buen director es el que se ajusta al ideal del director. Si no existe esa correspondencia, se produce una "deficiencia". La esencia de un objeto sometido a evaluación se define a menudo mediante un conjunto de criterios intuitivos que parecen aplicables al caso. Por ejemplo, se espera que un "buen director de proyectos" sea de cierto modo y haga determinadas cosas. El evaluador puede hacer explícito este conjunto de criterios, implícito por regla general, con el fin de juzgar al director. Lo mismo puede hacerse en relación con un buen programa, un buen libro de texto, etc. La lista de criterios nunca es inclusiva y siempre discutible. Sin embargo, con frecuencia la lista es eficaz para persuadir a los destinatarios de la calidad. Ejemplo: los informes de la *Consumers' Union* sobre productos manufacturados.

Relación simbólica. Sólo los miembros de un grupo concreto creen en la relación mágica entre el símbolo y la cosa, como la bandera nacional. Las relaciones simbólicas son importantes para describir ciertos aspectos de los sistemas y categorías sociales. Estas relaciones son algo diferentes porque no pueden justificarse ante otros. Los educadores vinculan con frecuencia significados especiales de este tipo con determinadas facetas de su programa y con ciertos líderes carismáticos dentro del mismo. Personas y cosas se convierten en objetos de fe en sí mismos y por sí mismos. Esta situación constituye un rompecabezas habitual para el evaluador, que puede tratar de buscar en vano más relaciones materiales subyacentes a la fe.

© Ediciones Morata, S. L.

Scanned by CamScanner

La tercera clase de argumentos asume el menor número de premisas. Estas argumentaciones no se basan en la semejanza con la lógica formal ni se desarrollan a partir de acuerdos previos sobre la estructura de la realidad. En cambio, tratan de *establecer* la realidad. El ejemplo y las ilustraciones lo hacen recurriendo a casos concretos, las analogías y las metáforas poniendo de manifiesto ante los destinatarios relaciones conceptuales nuevas. La evaluación "naturalista" se funda en gran medida en esta forma de argumentación.

Ejemplo. El recurso al ejemplo supone la falta de unanimidad respecto a una regla concreta, aunque, en principio, se esté de acuerdo en poder llegar al entendimiento. Una serie de ejemplos invita a la generalización. A veces, el razonamiento procede de lo particular a lo particular, sin establecer regla alguna. Los ejemplos operan de forma implícita. La técnica del "caso cerrado" y del "precedente" legal se funda en esta fórmula. Esta argumentación valora lo concreto y lo habitual. Para que el ejemplo sea eficaz, ha de aceptarse su verosimilitud.

Ilustración. Mientras que el ejemplo se utiliza para establecer una regla, la ilustración se emplea para aclararla y fortalecer la adhesión a ella. Facilita la comprensión. Las ilustraciones de las formas de razonamiento de este apartado tratan de aclarar las categorías, pero éstas no dependen de las ilustraciones.

Analogía. La analogía establece una relación entre dos esferas hasta entonces no relacionadas y, en consecuencia, es esencial en la invención y la imaginación. Desarrolla y amplía el pensamiento.

Metáfora. La afirmación metafórica abre nuevos campos de pensamiento pasando de lo conocido a lo desconocido y ayudando a señalar cosas inespecificables en el lenguaje corriente. En la mayoría de los casos, la afirmación metafórica se utiliza en conjunción con los ejemplos y las ilustraciones. En la parte dedicada a la evaluación naturalista expondremos cómo actúa ampliando las ideas de los destinatarios.

Esta relación de técnicas de argumentación no es exhaustiva y no pretende ser una lista de técnicas a partir de las cuales se construyan evaluaciones. Intenta, en cambio, ilustrar el tipo de razonamiento utilizado, en la práctica, en las evaluaciones.

Análisis de la "Educational Product Evaluation" de GLASS

He elegido *"Educational Product Evaluation: A Prototype Format Applied"*, de GLASS (1972), para analizarlo en relación con los argumentos enumerados en la sección anterior. He seleccionado esta evaluación porque es

© Ediciones Morata, S. L.

muy accesible, es sucinta, la autoridad del autor es incuestionable, muestra diversos y complejos argumentos evaluativos y me parece muy persuasiva.

La técnica que utilizo consiste en parafrasear la obra de GLASS e indicar entre paréntesis los argumentos a medida que aparezcan. No digo que haya encontrado todos los argumentos en el trabajo de GLASS, que los que he reseñado no pudieran categorizarse de otra manera, ni que los tipos de argumentaciones enumerados en el apartado anterior sean exhaustivos. Sería imposible hacer una lista de todos los argumentos o sus tipos, o clasificarlos sin ambigüedad. Sólo pretendo ilustrar, a partir de un trabajo muy bueno, que estas argumentaciones desempeñan un papel crítico en el razonamiento evaluativo. La lógica general del trabajo de GLASS es algo más compleja que las argumentaciones expuestas y prescindiré de ella hasta haber estudiado las argumentaciones particulares.

GLASS empieza con una breve introducción que plantea el carácter provisional de las técnicas de evaluación y describe lo que pretende hacer. El cuerpo del escrito se divide en diez partes.

La *Parte I* es una descripción de la grabación del AERA que pretende evaluar. En sí, constituye un estudio de la evaluación de Michael SCRIVEN.

La *Parte II* menciona los tres objetivos del producto y los evalúa. Preparar a evaluadores es bueno, dado que hacen falta destrezas de evaluación a causa de la legislación que ordena efectuar evaluaciones (causa y efecto). La producción de una *cassette* que pueda utilizarse mientras se va a trabajar puede o no ser deseable, puesto que invade el tiempo de una persona dedicado a sí misma de improviso (argumento pragmático: se valora un hecho por sus consecuencias). La experimentación con nuevos medios es recomendable si no es un "mero pasatiempo técnico" (la persona y sus acciones: intención del actor). La prueba consistirá en la adecuada evaluación de la *cassette* (intención inferida a partir de la coherencia de las acciones: la persona y sus acciones).

La *Parte III* describe el estado de la cuestión cuando entra en escena el evaluador. Ya estaban decididas previo acuerdo las siguientes cuestiones: el director, los temas de la grabación, el profesor, la materia y las copias iniciales. Todavía no se habían decidido la venta de las *cassettes*, la selección de los materiales y los planes de promoción y venta. Esto delimita dónde tiene sentido que GLASS centre su atención. Está aquí implícito el argumento de que sería una pérdida de tiempo y esfuerzo del evaluador y de los destinatarios ocuparse de cuestiones ya decididas (argumento de pérdida).

La *Parte IV* se titula "alternativas" y constituye una forma brillante de dar la vuelta a la argumentación general. GLASS enumera lo que podría adquirirse con los recursos empleados en la producción de la *cassette:* una sesión de preparación de una jornada para cien investigadores, la impresión de veinte mil ejemplares de materiales escritos, una gratificación semestral para un investigador en formación o cuatro becas para las sesiones de entrenamiento para investigadores procedentes de minorías étnicas. Éstas serían las alternativas para el patrocinador, la *U.S. Office of Education.* Asimismo, se rela-

© Ediciones Morata, S. L.

Scanned by CamScanner

cionan alternativas para los demás destinatarios principales: el director, el AERA y el usuario.

El razonamiento comienza planteando lo que se dejaría a cambio de la adopción del enfoque basado en la *cassette* (argumento de sacrificio). Establece la equivalencia de las alternativas en relación con lo que podría comprarse con los recursos dedicados al plan de producción de la *cassette* (argumento de identidad). Las alternativas también son equivalentes porque todas ellas son coherentes con la intención del productor. Sin poner de manifiesto las razones, GLASS escoge la alternativa del texto mecanografiado por ser la que cuenta "con mayores ventajas" (argumento de comparación). ¿Por qué escoge la alternativa de mayor peso para proseguir las comparaciones? En el razonamiento está implícita la idea de que se debe optar por la técnica que cumpla *mejor* el fin que pretende el productor (argumento de fines y medios).

Seleccionado el competidor más fuerte, GLASS, en la *Parte V* del estudio, amplía al máximo la comparación de costes entre el proyecto de producción de la *cassette* y el del texto escrito (argumentos de comparación y sacrificio). Estudiando las comparaciones de costes, afirma que el coste sería aceptable para grupos de entre diez y quince personas; que la grabación es demasiado cara [para ello, menciona como autoridad el departamento de enseñanza audiovisual de Colorado (argumento de autoridad)]; que los textos escritos serían más fáciles de almacenar, y que el coste del texto escrito podría reducirse más. Todos estos argumentos son variaciones basadas en comparaciones entre los dos enfoques y en lo que podría costar cada uno ante diversas contingencias.

La *Parte VI* constituye la evaluación "intrínseca", que el evaluador califica como "secundaria". Es una evaluación de la calidad técnica, contenido y "utilización del carácter exclusivo" del medio. Esta serie de argumentos se ocupa de cuestiones que son secundarias respecto a la comparación completa entre el enfoque mediante la *cassette* y el del texto escrito, aunque puedan ser importantes para un usuario potencial que desee adquirir las *cassettes*.

La evaluación de la calidad técnica y del contenido se basan en el ideal de lo que debieran ser la calidad técnica y el contenido [las desviaciones de estos ideales son deficiencias (argumento del acto y la esencia)]. El evaluador elabora una relación de los criterios que considera pertinentes y sobre los que suele haber acuerdo, dado que no trata de justificarlos. La calidad técnica engloba la calidad de la cinta magnética, de la fidelidad de la grabación, la calidad estética, el montaje y la envoltura externa. Cada criterio va acompañado por un juicio y algunas notas que mencionan observaciones en las que se funda el juicio. Similares criterios "a posteriori" se aplican al contenido.

La segunda parte de la evaluación intrínseca versa sobre la "utilización del carácter exclusivo" del medio utilizado *(cassette)*. Se trata, de nuevo, de un argumento basado sobre todo en la "esencia" de la *cassette* (acto y esencia). Se estudian dos afirmaciones del productor. El evaluador refuta la

© Ediciones Morata, S. L.

presunción de la ventaja que para el usuario suponga la posibilidad de detener la grabación contando el número de detenciones. La segunda afirmación, acerca de que un número significativo de personas tiene un magnetófono de *cassette* y tiempo durante el que escuchar las grabaciones se confirma mediante una encuesta por correo dirigida a cien miembros del AERA (argumento de probabilidad). Sabiendo que se dirige a un público constituido por investigadores educativos, GLASS indica los intervalos de confianza en una nota a pie de página. Durante la segunda parte, la comparación latente con el texto escrito se utiliza para refutar las afirmaciones del productor acerca de que la lectura del texto escrito no permite hacer las mismas cosas. El razonamiento de GLASS se opone a la afirmación del productor respecto a las "características exclusivas" del enfoque basado en la *cassette* (argumento de acto y esencia).

La *Parte VII* es la evaluación de "resultados", que el evaluador califica de "primaria". La comparación entre la *cassette* y el texto escrito se establece en términos de resultados. Dice que, aun cuando el medio auditivo fuera tan eficaz en la transimisión de información, es más lento. Esta comparación supone la medida. Es una comparación basada en consecuencias pragmáticas (argumento de comparación; argumento pragmático). El acceso a la información también es mucho más lento con la *cassette* (argumento de sacrificio).

GLASS cita una revisión de estudios experimentales de comparación entre el medio auditivo y el visual, diciendo que no es concluyente porque la eficacia relativa de cada uno depende de diversas contingencias. Esto no favorece su agumentación, salvo para aumentar la credibilidad del evaluador, pero permite a GLASS describir con detalle un estudio concreto que demuestra la superioridad del aprendizaje visual (argumento de ilustración).

La *Parte VIII* es un resumen de conclusiones y una serie independiente de recomendaciones para cada uno de los conjuntos de destinatarios. Las recomendaciones son muy directas, explícitas y específicas en cuanto a su orientación. De hecho, establecen una jerarquía de acciones que cada destinatario puede adoptar, dependiendo de las contingencias.

La *Parte IX* muestra una relación de aquellos destinatarios concretos a quienes podría beneficiar el enfoque basado en la *cassette*. Los argumentos consisten en que las *cassettes* benefician a los estudiantes invidentes y a los grupos grandes, *"Reverse Luddites"*. Todos los argumentos de las Partes VIII y IX son variaciones de los de coste/beneficio (argumentos de fines y medios; argumentos pragmáticos).

La *Parte X* es rara por su reflexividad. Se titula: *"Evaluating the Evaluator"* ("Evaluación del evaluador") y estudia los propios sesgos del evaluador. Por supuesto, el simple hecho de emprender esta labor refuerza la credibilidad del evaluador. GLASS indica que las evaluaciones suponen un coste, sobre todo cuando destruyen la sensación de comunidad (argumentos de persona y grupo). En este caso, él emprendió el estudio porque se lo pidió el autor del producto (persona y acto). Demuestra su credibilidad manifestando que llevó a cabo acciones contrarias a sus propios intereses, poniendo de manifiesto su imparcialidad.

© Ediciones Morata, S. L.

Scanned by CamScanner

GLASS divide sus motivos en las categorías exclusivas de "motivos para una evaluación favorable" y "motivos para una evaluación desfavorable" (argumento de división de la totalidad en partes). Los sesgos a favor de una revisión positiva se derivan del hecho de que GLASS es miembro del comité ejecutivo del AERA, los benefactores; y del hecho de que los productores son colegas directos suyos (argumento de persona y su grupo).

Los motivos a favor de una revisión negativa son que él mismo se negó a participar, basándose en que el proyecto de la *cassette* no es rentable en cuanto a la relación coste/eficacia, y en el hecho de que, en cierta ocasión, el director del proyecto le venció en un partido de tenis de mesa. Estos argumentos dependen del constructo de la persona que subyace a sus actos (argumento de persona y actos). Concluye la evaluación indicando que no ha recogido datos relativos a las actitudes ante el producto ni sobre su eficacia. Deja que los destinatarios saquen sus propias conclusiones sobre la proporción entre los sesgos y sobre la credibilidad global.

Merece la pena examinar la estructura *general* del estudio. Consiste en una forma compleja de razonamiento denominada argumentación de "jerarquía doble" (PERELMAN y OLBRECHTS-TYTECA, 1969). La argumentación de jerarquía doble consiste en dos jerarquías de valores u objetos que suelen estar conectadas por relaciones a partir de la estructura de la realidad. Por ejemplo, el enunciado de LEIBNIZ de que "dado que [Dios] cuida de los gorriones, no se olvidará de las criaturas racionales que le son mucho más queridas" se basa en las jerarquías implícitas de las criaturas y de las atenciones de Dios, conectadas por la causa y el efecto implicados. Las argumentaciones de jerarquía doble adoptan con frecuencia la forma de enunciados condicionales "si... entonces" y, por regla general, están implícitas.

La estructura lógica general de la evaluación de GLASS parece ser la de una argumentación de jerarquía doble. Una jerarquía es la jerarquía de costes. La otra es la de beneficios. Ambas están conectadas por la relación entre medios y fines. De hecho, todo el estudio se basa en el establecimiento de esta estructura lógica y de organizar las argumentaciones subordinadas dentro del gran diseño general.

Por ejemplo, después de definir el contexto del estudio por la descripción del producto, los objetivos del productor y el punto en el que entra en acción el evaluador, GLASS establece una jerarquía de alternativas en la Parte IV. En la Parte V, selecciona el competidor más fuerte, estableciendo la jerarquía de comparación de costes entre los dos enfoques. En la Parte VII elabora la jerarquía de beneficios, basada de nuevo en comparaciones entre los dos enfoques. La relación medios-fines conecta ambas jerarquías. Ello exige escoger los mejores medios para conseguir los fines establecidos. Las contingencias que aparecen en las Partes VIII y IX son estudios de lo que ocurriría si nos desplazáramos hacia arriba o hacia abajo en cualquiera de las jerarquías de costes y beneficios.

De este modo, GLASS ha efectuado un análisis coste/beneficio sin hacer ninguna medida precisa de costes ni de beneficios. Además, es persuasivo.

© Ediciones Morata, S. L.

Creo que resulta tan irresistible a causa de la integración de argumentos. Todas las argumentaciones operan de manera económica dentro de la estructura general. Existen muy pocos movimientos extraños. Sólo la introducción y la sección final sobre la credibilidad del evaluador no contribuyen de forma directa a la línea argumental. Desde el punto de vista estético, estos dos apartados están situados adecuadamente al principio y al final. Nos inclinamos a pensar, con POLANYI, que la prueba definitiva de veracidad consiste en la coherencia y en la belleza de la estructura.

La parte más difícil de manejar en el contexto del diseño general es la VI, que versa sobre la calidad de la *cassette*. En realidad, a GLASS le pidieron que evaluase la *cassette* en sí. Supongo que el problema básico de incompatibilidad intelectual del que surgió la evaluación consistía en que la *cassette* en sí era buena, pero GLASS no consideraba justificada la inversión. Él redefinió el problema de manera que la evaluación se refiriese al enfoque mediante la *cassette* y no sólo sobre ésta. No obstante, no podía evaluar el producto sin una evaluación directa de la grabación. Asimismo, uno de los grupos destinatarios estaba constituido por los posibles usuarios que compraran la grabación y no sólo los miembros de la ejecutiva del AERA, que querían saber si la actividad en su conjunto estaba justificada. Él consideró secundaria la evaluación de la *cassette*, frente a la evaluación primaria de resultados. Desde el punto de vista estético, también le restó importancia situando la evaluación intrínseca en medio de la presentación general.

Además de la coherencia lógica de la evaluación, también es persuasiva porque las premisas del acuerdo habían sido bien seleccionadas en relación con los destinatarios. Las proporciones coste/beneficio constituyen valores poderosos para los destinatarios y es raro que las personas versadas en la ética utilitarista cuestionen las relaciones medios-fines. GLASS lleva a los destinatarios desde valores con los que comulgan hasta conclusiones que quizá no aceptaran en principio. Conoce a la perfección a sus destinatarios, dirigiéndose, incluso, directamente a cada sector y dándoles recomendaciones específicas para cada uno. No obstante, sospechamos que estas argumentaciones no resultan igualmente persuasivas para todos. Es probable que algunos grupos tengan en cuenta valores y condiciones que la evaluación no contempla. Sin embargo, resuelve con elegancia, tanto lógica como estéticamente, el problema de los grupos complejos de destinatarios con preocupaciones diferentes.

¿Cómo rechazar una evaluación de este tipo? Podríamos atacar la estructura argumental básica negando la equivalencia de las alternativas y cuestionando la selección del competidor más fuerte, rechazando, en consecuencia, la relación medios-fines. Podríamos atacar los costes y negar los beneficios comparativos que se derivan del enfoque del texto escrito. Un ataque dirigido contra la evaluación secundaria de la calidad de la grabación no significa mucho porque no afecta a la lógica general del estudio. GLASS podría hacer ahí algunas concesiones y llegar aún a conclusiones negativas. Asimismo, podríamos decir que el evaluador se halla excesivamente sesgado,

© Ediciones Morata, S. L.

Scanned by CamScanner

atacando, en consecuencia, la credibilidad del estudio, aunque la exposición que hace GLASS de sus propias inclinaciones hace difícil abordarlo de ese modo. Toda evaluación es discutible, aunque sea muy persuasiva.

Hemos de hacer notar que, en esta magistral evaluación, GLASS ha utilizado la mayoría de los tipos argumentales antes mencionados. Se basa especialmente en argumentaciones a partir de la estructura de la realidad, sobre todo en relaciones de sucesión que vinculan fenómenos con consecuencias, como los argumentos de medios y fines, y los cuasilógicos, como las comparaciones. Presenta muy pocas argumentaciones que traten de establecer la estructura de la realidad, como ejemplos y metáforas.

Los procedimientos formales de recogida de datos sólo se utilizan de manera moderada y, cuando se emplean, no son fundamentales respecto a la importancia de la evaluación. La mayoría de los datos consiste en "hechos" ya aceptados. Los procedimientos formales de recogida de datos no son esenciales para la evaluación; la argumentación sí lo es.

Análisis de la respuesta de SCRIVEN a la evaluación de GLASS

Escribí este apartado cinco meses después que el resto del texto porque no conocía la respuesta de SCRIVEN a la evaluación de GLASS hasta que este último me habló de ella. La escala temporal es importante porque SCRIVEN atacó la evaluación de GLASS precisamente del modo que habría que hacerlo, según indicamos en el anterior apartado. Hace falta negar la equivalencia de las alternativas y cuestionar la selección del competidor más fuerte, rechazando de este modo la relación entre medios y fines, así como atacar los costes, negando los beneficios comparativos de la alternativa del texto escrito. Esto es lo que hace SCRIVEN y resulta interesante la comparación de su razonamiento con el de GLASS.

SCRIVEN (1972) comienza diciendo que le han invitado a responder a la evaluación de su *cassette* llevada a cabo por GLASS (intenciones del actor: argumento que relaciona a una persona con sus actos). Esquematiza las condiciones básicas que rodean su decisión de rehacer todo el segundo ciclo en vez de revisar el primer producto. El argumento se presenta de forma racional como una elección entre tres alternativas (argumento pragmático). Sin embargo, SCRIVEN dedica tanto espacio a desarrollar el contexto de su acción que resulta obvio que pretende que sus destinatarios comprendan sus motivaciones (argumento de relación de una persona con sus actos).

Por otra parte, el problema de SCRIVEN en relación con los sesgos es mucho mayor que el afrontado por GLASS, pues SCRIVEN responde a una evaluación de un producto suyo, por lo que de inmediato se hace sospechoso. Es interesante su actuación al respecto: afirma que la dirección del sesgo es tan evidente que no puede resultar lesiva (argumento pragmático). Este argumento trata de *romper* la relación entre el acto de contraargumentación y la motivación (sesgo) del actor (argumento de relación de una

© Ediciones Morata, S. L.

persona con sus actos). Es una tentativa de reducir el sesgo que se percibe en el actor. SCRIVEN refuerza su imparcialidad demostrando su capacidad de distinguir entre "excusas" y "críticas" (en sí, un argumento de división de la totalidad en partes). Aunque en las dos primeras partes de su respuesta aparecen varios tipos de argumentos, SCRIVEN los organiza con el fin de quedar disociado del sesgo.

En la tercera parte, SCRIVEN pasa a considerar las conclusiones sobre los aspectos materiales de la grabación. Acepta la mayoría de las críticas de GLASS (reforzando su credibilidad) pero desestima la conveniencia de cintas magnetofónicas baratas porque su comportamiento es peor (argumento pragmático) debido a los efectos de calentamiento y de fricción (causa y efecto). Asimismo, niega la deformación del sonido si las *cassettes* se reproducen en magnetófonos adecuados. La mención de los aparatos "Advent" y "MacIntosh" atrae de inmediato la atención de los destinatarios aficionados a los equipos de reproducción de sonido y pone de manifiesto que SCRIVEN conoce el tema (argumento de autoridad). Queda claro, pues, por qué empieza por el análisis de un aspecto de poca importancia relativa como es la materialidad de la grabación: SCRIVEN conoce mejor que GLASS estos aspectos. Constituye también un ataque al análisis de costes de GLASS en relación con la magnitud de la audiencia a la que podría llegarse y del coste de las cintas magnetofónicas.

Poco hay que decir respecto a la calidad de la grabación, pues la evaluación que GLASS hace de la cinta de SCRIVEN es un rosario de "alabanzas". SCRIVEN desestima la crítica de que la falta de citas constituya un déficit, basándose en la información recibida procedente del campo (argumento de probabilidad).

A continuación, llega el ataque fundamental de SCRIVEN contra la lógica de la evaluación de GLASS. SCRIVEN concede que "el procedimiento general de efectuar en concreto estimaciones comparativas de coste/eficacia me parece absolutamente correcto y, en realidad, es el método que debe elegirse en toda evaluación educativa". Pero no está de acuerdo con la valoración que hace GLASS de los costes y de los beneficios y, sobre todo, con la forma utilizada por él para relacionarlos. El alegato fundamental de SCRIVEN consiste en que GLASS ha seleccionado para la comparación un competidor erróneo (el texto escrito).

Sostiene SCRIVEN que la *cassette* pretende fines diferentes de los propios del texto escrito; es más útil que escuchar la radio en el coche y puede constituir un sustituto barato de un profesor visitante en un curso (estos dos objetivos exclusivos se establecen por definición). Estas argumentaciones niegan la equivalencia de resultados establecida por GLASS (argumento de identidad). Las *cassettes* cumplen fines diferentes y, por tanto, las alternativas no son equivalentes. La *cassette* constituye un elemento motivador en lugares en los que el material escrito no lo es (argumento pragmático). Por otra parte, los costes son similares a los de las cintas magnetofónicas comerciales (comparación con la norma).

SCRIVEN admite las ventajas del texto escrito en cuanto a coste, velocidad

© Ediciones Morata, S. L.

Scanned by CamScanner

y repetición, pero la *cassette* introduce un nuevo elemento del que carece el material escrito. SCRIVEN da varias razones para utilizar la *cassette* en clase: permite escuchar en directo a la autoridad; son preferibles varias voces a una sola, y la grabación proporciona variedad (argumento pragmático, de la totalidad superior a las partes y de desarrollo ilimitado). Aunque en general no sea superior, sí "ensancha el repertorio". Nótese que, sobre todo, SCRIVEN se manifiesta a favor del *carácter exclusivo de la cassette*, mientras que GLASS afirma que el texto escrito cumplimenta mayor cantidad de los objetivos comunes (cualitativo frente a cuantitativo).

SCRIVEN considera problemáticas las alternativas de costes. Quizá el dinero para la producción de las *cassettes* no estuviera disponible para ninguna otra cosa en las circunstancias concretas (¿argumento de pérdida?). Aunque si lo estuviera, el AERA debería experimentar (acto y esencia: la "experimentación", criterio implícito del AERA); y éste es un experimento razonable, habida cuenta de otros intentos (argumento de comparación con la norma). Por otra parte, es mejor probarlo en educación para poder utilizarlo en ella (¿identidad parcial?).

Pero la principal objeción de SCRIVEN en contra de la evaluación de GLASS se refiere al objeto de comparación: "Por tanto, mi principal crítica de la evaluación de GLASS se refiere a la elección de la principal comparación fundamental. No debería haber sido el texto escrito, sino el mejor contenido y envoltorio más barato de *cassette*". El desacuerdo no sólo tiene que ver con la comparación, versa también sobre si la conexión de costes y beneficios debe hacerse por medio de un argumento medios-fines, que señale el *mejor* competidor (el texto escrito) o mediante un argumento pragmático, que apunte al competidor *menor*.

SCRIVEN insiste en el carácter exclusivo del medio. Aunque GLASS refuta el argumento de dicho carácter contando el número de veces que SCRIVEN detiene la cinta, éste dice que no le persuade la refutación porque GLASS no ofrece algo que suponga una utilización exclusiva. SCRIVEN apunta como factor único la "comprehensibilidad", admitiendo que el número de detenciones de la *cassette* es mal indicador de ese criterio (argumento de acto y esencia).

En la última parte de su respuesta, SCRIVEN indica que la categoría *"Reverse Luddites"* de posibles audiencias, que señala GLASS, está concebida de forma demasiado restringida y que hay muchas personas normales que sacarían provecho de la *cassette* porque hay gente que prefiere escuchar a leer (argumento de frecuencia). En realidad, todo el mundo lo hace en algún momento del día (causa y efecto). SCRIVEN respalda estas argumentaciones preguntándole a su mujer (argumento de ilustración), del mismo modo que GLASS utiliza un estudio de uno de sus alumnos graduados. Por último, SCRIVEN dice que también debemos tener en cuenta los beneficios adicionales correspondientes a lo aprendido por la población intermediaria (él mismo y el productor) (argumento pragmático). Todos estos argumentos incrementan los beneficios, haciendo más aceptable la proporción coste/beneficio.

La lógica general de la evaluación original de GLASS consiste en un

© Ediciones Morata, S. L.

argumento de jerarquía doble de costes y beneficios conectada por una relación medios-fines. SCRIVEN percibe con toda claridad esta estructura y acepta la comparación básica de costes y beneficios como método preferible en toda evaluación. Trata de demostrar que los costes no son extravagantes ni poco razonables y que GLASS ha subestimado de forma significativa los beneficios de la *cassette*. Pero la crítica principal consiste en rechazar el argumento medios-fines de GLASS, reemplazándolo, como enlace entre costes y beneficios por un argumento pragmático. El argumento medios-fines requiere la comparación de la *cassette* con la *mejor* alternativa disponible. El argumento pragmático de SCRIVEN sólo requiere que la *cassette* sea la mejor en relación con todas las demás *cassettes* existentes. La estrategia de SCRIVEN consiste en sostener las características exclusivas de la *cassette*, de manera que no tenga que entrar en una competición frontal y en todos los aspectos con el texto escrito. SCRIVEN defiende un terreno de comparación cualitativamente diferente.

En su forma elemental, el argumento pragmático consiste en evaluar un hecho en relación con sus consecuencias. El argumento de medios a fines, por su parte, depende del acuerdo sobre estos últimos. La determinación de los *mejores* medios para conseguir los fines depende de la definición *exacta* de los resultados que se pretenden. No se tienen en cuenta los valores no relacionados con los fines. Si éstos están definidos con precisión y se está de acuerdo en ellos, la determinación de los medios mejores se convierte en un problema técnico. Este razonamiento, adecuado para las disciplinas técnicas, es muy diferente del cotidiano.

Hablando en general, el trabajo de GLASS en conjunto tiende a centrarse en el razonamiento de medios a fines y a orientarse en un sentido más técnico, mientras que el de SCRIVEN tiende a basarse más en la argumentación pragmática. De hecho, la evaluación independiente de los objetivos del programa evaluado de SCRIVEN puede considerarse como la expresión última del argumento pragmático. No nos preocupamos en absoluto de los fines manifiestos sino sólo de las consecuencias del objeto sometido a evaluación. En general, la concepción de un problema de evaluación según la lógica de "medios-fines" tiende a devaluar los medios en relación con los fines, mientras que la concepción del mismo problema según la lógica de "hecho-consecuencias" tiende a dar mayor importancia relativa al hecho. La oposición de SCRIVEN a GLASS culmina en una discusión sobre los fines del proyecto de la *cassette*.

En un nivel más abstracto, la disputa se plantea entre dos principios de opción racional: el principio de los medios eficaces y el principio de inclusividad (RAWLS, 1971). El principio de los medios eficaces estipula que, dado un objetivo, hay que conseguirlo con el menor gasto posible de medios o, dados los medios, ha de cumplirse el objetivo en la mayor medida posible. En otras palabras, han de adoptarse las mejores alternativas.

El principio de inclusividad estipula que ha de preferirse un plan a otro alternativo si, además de los objetivos del segundo, cumple otros adicionales. Al defender el enfoque de la *cassette* como modo de ampliación del "reper-

© Ediciones Morata, S. L.

Scanned by CamScanner

torio", pero no como sustituto total del texto escrito, SCRIVEN se expresa del modo antes citado.

Las pocas diferencias entre GLASS y SCRIVEN no deben ocultar las muchas semejanzas de sus argumentos evaluativos. Ambos aceptan la comparación de costes y beneficios como método preferible. Ambos se basan especialmente en argumentos de "estructura de la realidad", fundándose algo más GLASS en las relaciones de coexistencia, p. ej., las existentes entre una persona y sus actos y entre una persona y su grupo. SCRIVEN se basa algo más en argumentos de relaciones de sucesión, sobre todo en los pragmáticos. A pesar de los argumentos de estructura de la realidad, se recaba poca información procedente de terceros. Ambos evaluadores se basan en sus propias observaciones personales para extraer los datos primarios.

De forma secundaria, ambos utilizan argumentos cuasilógicos, aunque sólo son la mitad de frecuentes que los argumentos antes mencionados. Ambos utilizan en una sola ocasión argumentaciones que tratan de establecer la estructura de la realidad, p. ej., ejemplos, analogías, etc. Un tipo de evaluación completamente distinto habría consistido en una utilización de campo de las *cassettes*, recogiendo anécdotas respecto a la forma de utilizarlas.

Tanto GLASS como SCRIVEN utilizan en sus respectivos artículos más de veinticinco argumentos, aunque el de GLASS ocupa el doble que el de SCRIVEN. La elevada densidad argumental de SCRIVEN refleja su estilo general: tiene la habilidad de presentar una serie de razones respecto a un juicio determinado, una tras otra, de manera profusa y lineal. Tanto aquí como en otros casos, GLASS presenta menos razones, pero están articuladas entre sí de manera más cuidadosa, basándose minuciosamente unos argumentos en otros.

En parte por eso, el trabajo de GLASS es más coherente y agradable, desde el punto de vista estético, que el de SCRIVEN. Éste tiene la desventaja de estar respondiendo al artículo de GLASS, en vez de crear una argumentación propia, plenamente estructurada, como hizo, por ejemplo, en su artículo sobre la evaluación independiente de los objetivos (SCRIVEN, 1973). El flujo algo irregular que se aprecia en el artículo de SCRIVEN cuando responde a diversas cuestiones del de GLASS resta algo de la persuasividad global de sus argumentos. Es una grave desventaja a la que tienen que hacer frente quienes responden a otro documento.

Por último, hay que señalar que este intercambio de opiniones entre dos de los principales teóricos de la evaluación no se refiere sobre todo a los datos. La disputa se plantea en torno a la comparación más adecuada al objeto sometido a evaluación, que puede llevarse hasta la forma argumental preferida y a los destinatarios a quienes se dirige. Algunos piensan que todas las disputas pueden resolverse acudiendo a los datos, pero no es así. A menudo, lo que se discute es la lógica de la evaluación.

© Ediciones Morata, S. L.

APÉNDICE B

La evaluación naturalista

Cuando leemos una novela o un poema, aprendemos algo. Si alguien nos preguntase qué hemos aprendido, nos resultaría difícil decirlo. A menudo, el conocimiento adquirido mediante este tipo de lectura no reviste forma proposicional. Sin embargo, al leer estas obras, la experiencia de la novela o del poema queda cartografiada en la mente del lector. Se ha dado en llamar "naturalistas" (STAKE, 1976) o "espontáneas" (PERELMAN y OLBRECHTS-TYTECA, 1969) a los tipos de generalizaciones que el lector consigue. La generalización naturalista utiliza un tipo especial de argumentación cualitativa.

Los argumentos que tratan de establecer la estructura de la realidad y suponen el mínimo acuerdo previo entre el autor y su público son los más utilizados en la evaluación "naturalista". Entre ellos se cuentan el ejemplo, la ilustración, la analogía y la metáfora. Yo considero "naturalista" una evaluación que procure que sus destinatarios lleguen a generalizaciones naturalistas; que se dirija a destinatarios no técnicos, como los maestros o el público en general; que utilice el lenguaje ordinario; que se base en el razonamiento informal cotidiano, y que utilice de forma generalizada argumentos que traten de establecer la estructura de la realidad. En esta categoría incluyo la mayor parte de las evaluaciones mediante el estudio de casos, así como las que emplean procedimientos de tipo judicial.

DENZIN (1971) describe el enfoque naturalista en sociología. Trata de mezclar las "características encubiertas e íntimas del acto social con sus homólogas públicas y observables, en sentido conductual. De este modo, va adelante y atrás entre la palabra y la acción, la definición y el acto". El observador forma parte del acto de investigación y las reflexiones sobre sí mismo pueden constituir datos importantes. La investigación comienza con cuestiones preocupantes y admite cualesquiera datos éticos pertinentes.

La atención se centra en la complejidad de la vida diaria y el naturalismo trata de comprender el mundo cotidiano según la experiencia de quienes lo

© Ediciones Morata, S. L.

Scanned by CamScanner

viven. El naturalista demuestra un profundo respeto hacia el mundo empírico. Los participantes constituyen fuentes constantes de ideas y elementos de verificación de éstas que elabora el naturalista. Para representar la totalidad del cuadro, son esenciales las perspectivas múltiples. Cuando trata de explicar los acontecimientos, el naturalista desarrolla diálogos encubiertos consigo mismo e, incluso, es probable que tome nota de ellos.

Como las diversas interacciones constituyen el centro de atención, el naturalista ha de seguir el desarrollo de los acontecimientos en el tiempo. Busca explicaciones más que previsiones, y las explicaciones se basan, por regla general, en las razones retrospectivas que dan las personas sobre sus propias conductas y las de otros. Esto hace necesaria la inmersión en la cultura y en la forma de expresión de los participantes. Las acciones conjuntas constituyen aspectos fundamentales a los que prestar atención, debiendo tenerse en cuenta en perspectiva histórica.

La validez se deriva de la comprobación cruzada de informaciones procedentes de distintas fuentes de datos y de la comprobación de las percepciones del evaluador respecto a las de los participantes. Los problemas y cuestiones surgen de las personas y situaciones que se someten a estudio más que de las preconcepciones del investigador. Los conceptos y los indicadores "se derivan del mundo de significados y acciones del sujeto". Al estructurar sus explicaciones, el naturalista busca la convergencia de sus fuentes de datos y elabora explicaciones en fases sucesivas que parten de la base de que ningún acontecimiento depende de una única causa. Un procedimiento habitual consiste en buscar los orígenes de los acontecimientos importantes. La introspección constituye una interesante fuente de datos.

Por supuesto, el sociólogo pretende elaborar una teoría generalizable. El evaluador naturalista sólo se preocupa por el caso que evalúa. El sociólogo tratará de justificar sus conclusiones ante un público universal. El evaluador naturalista debe ajustar su trabajo a un público concreto, que puede estar constituido por los participantes en el programa sometido a evaluación. Al presentar sus estudios, ambos han de fundarse de modo especial en ejemplos e ilustraciones extraídos del campo. El evaluador puede sacar o no conclusiones específicas a partir de los ejemplos. Si éstos están recogidos y presentados de forma sistemática, su lógica se parecerá a la del razonamiento inductivo. Sin embargo, en la evaluación naturalista, los destinatarios siempre tienen la opción de interpretar los descubrimientos y asignarles la credibilidad que les corresponda.

Cada vez son más corrientes las evaluaciones que utilizan ejemplos e ilustraciones de forma generalizada e, incluso, que consisten por completo en un único ejemplo ampliado. Tienen especial importancia cuando se dirigen a destinatarios no técnicos que no están acostumbrados a las formas más secretas de argumento cuantitativo, y a destinatarios de quienes el evaluador no puede presuponer las bases sobre las que llegar a un acuerdo. Los profesionales de la escuela se encuadran en ambas categorías. Es peligroso suponer de antemano que ellos parten de los mismos valores y

© Ediciones Morata, S. L.

ven la realidad del mismo modo que los evaluadores o los funcionarios del gobierno.

Las analogías y metáforas no suelen utilizarse en evaluación porque a menudo se perciben como meras figuras literarias y, por tanto, como datos poco fiables. No obstante, sí son formas importantes de llegar a establecer generalizaciones naturalistas. PETRIE (1976) dice que los paradigmas de KUHN transmiten categorías cognitivas esenciales para la iniciación en la comprensión de las teorías científicas. ORTONY (1975a, 1975b) expone de qué forma operan las metáforas para ampliar el marco de pensamiento.

ORTONY sostiene que las palabras no transmiten con precisión el flujo de experiencia tal como se presenta a la mente humana. La experiencia es continua y no discreta, y aunque las palabras no tengan significados exclusivos como los símbolos de los sistemas lógicos, tampoco reflejan con exactitud todas las formas de experiencia. Mediante la "particularización", las metáforas ayudan a salvar la distancia entre el lenguaje y la experiencia. La particularización transmite imágenes mentales a la mente del lector. Una expresión como "guerrero audaz" evoca su significado de forma más sucinta y compacta que una larga descripción. Además, las metáforas pueden recoger distinciones inexpresables de otra forma.

Según ORTONY, otra característica de las metáforas es su vivacidad. Están muy próximas a la experiencia y transmiten significados emocionales, tanto como cognitivos y sensoriales. Esta imaginabilidad se asocia con su facilidad de aprendizaje. Las metáforas facilitan la perspicacia y la comprensión personal pasando de lo conocido a lo menos conocido. Facilitan la generalización naturalista de los destinatarios. Sin embargo, es fundamental que el autor comprenda a su audiencia con el fin de saber si una afirmación metafórica ampliará la comprensión o dejará en suspenso a los destinatarios.

ORTONY extiende también esta concepción del lenguaje a la situación de enseñanza-aprendizaje. Basándose en la idea del conocimiento tácito de POLANYI, sostiene que el profesor debe saber siempre mucho más que lo que pueda expresar en forma proposicional. Este conocimiento tácito que, en parte, es de aplicación contextual, constituye la comprensión profunda de un campo o disciplina. Con el fin de comunicar el saber a un alumno, el profesor debe seleccionar aspectos de su conocimiento tácito y procurar representarlo en forma proposicional. Esta forma siempre se aleja algo de la plena comprensión tácita.

En principio, el estudiante sólo ve proposiciones. Es como aprender a montar en bicicleta leyendo una serie de instrucciones. La conducta del principiante está controlada por el conocimiento proposicional explícito, que resulta inadecuado. Aquí es donde el profesor puede ayudar al alumno por medio de ejemplos, metáforas y expresiones no literales.

Los científicos que tratan de aprender su disciplina encuentran problemas semejantes. Según críticos importantes, sería imposible aprender una disciplina científica limitándose sólo a seguir un conjunto de reglas (POLANYI, 1958; KUHN, 1970). Según KUHN, un científico aprende su disciplina por

© Ediciones Morata, S. L.

Scanned by CamScanner

medio de un conjunto de paradigmas (problemas concretos que permiten soluciones que hacen posible que el novato establezca comparaciones con otros problemas diferentes). El significado compartido se transfiere por medio de estas experiencias y no sólo mediante reglas.

La semejanza entre las generalizaciones naturalistas en la evaluación, mediante el uso de ejemplos y metáforas y otros argumentos que tratan de establecer la estructura de la realidad, es obvia. Aun a falta de proposiciones científicamente verificadas en el sentido de la lógica formal, se facilita la comprensión y el conocimiento de los destinatarios. Aunque las premisas epistemológicas y psicológicas en las que se basa sean diferentes de las de otros tipos de evaluación, la evaluación naturalista sigue siendo una forma de argumentación.

Ediciones Morata, S. L.

APÉNDICE C

Un contrato de evaluación
(Protocolo de contrato entre el equipo de evaluación y MEA/NE)

1. *Cometido*

El jurado de evaluación externa compuesto por el Dr. Wendell RIVERS, el Dr. Ernest HOUSE y el Dr. Daniel STUFFLEBEAM ha sido encargado por la *Michigan Education Association* y por la *National Education Association* para evaluar la validez y utilidad educativas en Michigan del *Michigan Accountability Model,* ocupándose de modo especial del componente de evaluación.

2. *Destinatarios* (en orden de prioridad)

NEA/MEA.
Responsables de decisiones del sistema educativo de Michigan *(State Board of Education* y *State Department of Education).*
Los medios de comunicación (el público).
Los usuarios (padres, PTA, público en general, etc.).
Personal técnico (sobre todo del área de la medición educativa).

3. *Informe y publicación del mismo*

El tribunal sólo es responsable de la realización de la evaluación y de su informe final. La NEA y la MEA pueden redactar y divulgar documentos independientes (como puntualizaciones, impugnaciones, comentarios o documentos de carácter descriptivo). Se conviene que el informe del equipo sea lo más reducido y directo posible y que ha de estar elaborado para comunicarlo a los destinatarios previstos del informe.

© Ediciones Morata, S. L.

Scanned by CamScanner

4. Divulgación

El tribunal externo tiene derecho a facilitar su informe a cualesquiera miembros de los grupos destinatarios o a otras personas tras su elaboración. La entrega del informe no supone el respaldo de la MEA o de la NEA. Ambas instituciones pueden optar por refrendar o no el informe, dependiendo de su juicio sobre la calidad y adecuación del informe. La MEA y la NEA decidirán, en su caso, la publicación de sus propios documentos que describan el informe, bien entendido que tales documentos se identificarán como propios y no como del equipo de evaluación. Únicamente el informe final, tal como lo prepare este equipo, será divulgado de manera que figuren en él los nombres de los componentes de dicho equipo.

5. Formato del informe

Se considera deseable que en el informe final del equipo de evaluación figuren los siguientes ítems:

a) mención de los acuerdos entre el equipo de revisión y NEA/MEA;
b) presentación de los principales descubrimientos;
c) presentación de las opiniones minoritarias, si las hubiera.

6. Cuestiones que se deben tratar en el informe

Se tratarán las siguientes cuestiones específicas:

a) validez y fiabilidad de los tests referidos a criterio;
b) utilización de los tests para evaluar al profesorado;
c) importancia de los objetivos en los que se basa la evaluación de Michigan;
d) participación de los profesores en la elaboración de dichos objetivos y de los tests;
e) recomendaciones del equipo de evaluación relativas a posibles cambios y posteriores estudios;
f) comentarios sobre el equilibrio entre el esfuerzo del Estado y la conveniencia de ampliar el ámbito de la evaluación, habida cuenta, sobre todo, de los factores de coste asociados con los proyectos para mejorar o ampliar la evaluación de Michigan;
g) calidad de la planificación en el *Michigan Accountability Program;*
h) proyecciones de coste/beneficio respecto al programa;
i) valor de los resultados e informes de evaluación para los distintos niveles de destinatarios de Michigan;
j) problemas de sesgo en el *Michigan Accountability Program;*

© Ediciones Morata, S. L.

7. Recursos (presupuesto) del programa

La MEA y la NEA aportarán al equipo externo de evaluación: recursos suficientes para cubrir ocho días de trabajo de cada miembro de dicho equipo en el programa, el apoyo administrativo que fuera necesario y cualesquiera requisitos de materiales y equipos, por ejemplo, magnetófonos, cintas magnetofónicas, etc., para las audiencias que se desarrollen en Lansing. Si cualquiera de los miembros del equipo tuviera que efectuar llamadas telefónicas interurbanas para solicitar a personas que trabajan en Michigan sus opiniones sobre el programa, tales gastos les serán reembolsados, siempre que presenten un informe exacto y completo sobre el objetivo de la llamada, indicando la persona con la que se contactó.

8. Calendario de entrega

El tribunal de evaluación ha de entregar su informe final el 1 de marzo o tan pronto como fuera posible a partir de esa fecha.

9. Acceso a datos

El *Michigan Department of Education* pondrá a disposición del tribunal de evaluación cuantos datos e informes solicitara éste para realizar su trabajo. Esta disponibilidad se limita, como es obvio, a los datos e informes de los que dispone en la actualidad el *Michigan Department of Education* respecto a los datos contabilizados en Michigan.

10. Procedimientos

De acuerdo con las condiciones anteriores, el tribunal externo de tres personas tendrá pleno control del proceso de evaluación que debe poner en práctica para cumplimentar responsablemente el trabajo a él encomendado y que acepta desarrollar. En consonancia, el tribunal acuerda implementar el siguiente procedimiento general:

Celebrará entrevistas exclusivamente con representantes del *Michigan Department of Education*, representantes de la NEA y la MEA, y de grupos seleccionados (profesores, administradores, consejeros de dirección y grupos de acción educativa). Asimismo, el tribunal revisará los documentos que pongan a su disposición la NEA, la MEA y el *Michigan Department of Education*. Por último, concederá audiencias para recabar información adicional sobre las cuestiones que considere de relieve en el transcurso de las entrevistas con diversos grupos interesados y al estudiar los distintos documentos.

© Ediciones Morata, S. L.

Scanned by CamScanner

Bibliografía

ALKIN, M. C., R. DAILLAK, y P. WHITE (1979). *Using Evaluations*. Beverly Hills: Sage Publications.

——— y F. S. ELLETT (1979). "The importance of category systems in evaluation theory: a personal viewpoint". *CEDR Quarterly 12* (Otoño), págs. 3-5.

AMAREL, M., E. R. HOUSE, D. LANGMEYER, D. LORTIE, R. MAYER, L. MCLEAN, y L. SEALEY (1979). *Reform, Response, and Renegotiation Transitions in a School Change Project*. Nueva York, Ford Foundation.

APPLE, M., (1979). *Ideology and Curriculum*. Londres, Routledge & Kegan Paul. (Trad. cast.: *Ideología y currículo*. Madrid, Akal, 1986.)

ATKIN, J. M. (1968). "Behavioral objectives in curriculum design: a cautionary note". *Science Teacher 35* (Mayo), págs. 27-30.

BAKER, E. L. y E. S. QUELLMALZ (eds.) (1980). *Educational Testing and Evaluation: Design, Analysis, and Policy*. Beverly Hills, Sage Publications.

BARNES, R. E. y A. L. GINSBURG (1979). "Relevance of the RMC models for Title I policy concerns". *Educational Evaluation and Policy Analysis 1* (Marzo-Abril), págs. 7-14.

BARRY, B. (1973). *The Liberal Theory of Justice*. Londres. Oxford Univ. Press.

——— (1965). *Political Argument*. Londres, Routledge & Kegan Paul.

BEARDSLEY, M. (1958). *Aesthetics*. Nueva York, Harcourt Brace Jovanovich.

BERGER, P. L. (1974). "Policy and the calculus of meaning", en *Pyramids of Sacrifice*. Nueva York: Basic Books.

BERNSTEIN, R. J. (1978). *The Restructuring of Social and Political Theory*. Filadelfia, Univ. Of Pennsylvania Press.

BEYER, L. E. (1977). "Schools, aesthetic meanings, and social change". *Educational Theory 27* (Otoño), págs. 274-282.

——— (1974). "Objectivity, autonomy, and aesthetic evaluation". *Journal of Aesthetic Education 8* (Julio), págs. 107-118.

BLOOM, B. S. (1956). *Taxonomy of Educational Objectives*. Nueva York, McKay. (Trad. cast.: *Taxonomía de los objetivos de la educación*. Tomo I: *Ámbito del conocimiento*. Alcoy, Marfil, 1975, 2.ª ed.)

© Ediciones Morata, S. L.

BLOOM, B. S., J. T. HASTINGS, y G. MADAUS (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. Nueva York, McGraw-Hill.

BOULDING, K. (1956). *The Image*. Ann Arbor, Univ. of Michigan Press.

BOWLES, S., y H. GINTIS (1972-1973). "I.Q. in the U.S. class structure." *Social Policy 3* (Noviembre-Diciembre, Enero-Febrero), págs. 65-96. (Trad. cast.: *La meritocracia y el "coeficiente de inteligencia": una nueva falacia del capitalismo. El I.Q. en la estructura de clases de los Estados Unidos*. Barcelona, Anagrama, 1976.)

BRONOWSKI, J. (1956). *Science and Human Values*. Nueva York, Harper & Row.

BROUDY, H. (1972). *Enlightened Cherishing*. Urbana, Univ. of Illinois Press.

BURKE, K. (1945). *A Grammar of Motives*. Nueva York; Random House.

CALLAHAN, R. E. (1962). *Education and the Cult of Efficiency*. Chicago, Univ. of Chicago Press.

CAMPBELL, D. T. (1975a). "Assessing the impact of planned social change", en G. M. Lyons (ed.) *Social Research and Public Policies*. Dartmouth College, NH: Public Affairs Center.

——— (1975b). "Degrees of freedom and the case study." *Comparative Political Studies 8*, 2.

——— (1974). "Qualitative knowing in action research." Presentado a the American Psychological Association. Nueva Orleans.

——— (1966). "Pattern matching as an essential in distal knowing", en K. R. Hammond (ed.) *The Psychology of Egon Brunswick*. Nueva York: Holt, Rinehart & Winston.

——— y J. STANELY (1966). *Experimental and Quasi-Experimental Design for Research*. Chicago, Rand McNally.

CARE, N. S. (1978). "Participation and policy." *Ethics 88* (Julio), págs. 316-337.

COLEMAN, J. S. (1975). "Racial segregation in the schools: new research with new policy implementations." *Phi Delta Kappan 57* (Octubre), págs. 75-78.

——— y cols (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

COOLEY, W. W. y P. R. LOHNES (1976). *Evaluation Research in Education*. Nueva York: Irvington.

CRONBACH, L. J. (1979). *Design of Evaluations*. Stanford: Stanford Evaluation Consortium.

——— (1974). "Beyond the two disciplines of scientific psychology." Presentado a the American Psychological Association, Nueva Orleans.

——— (1971). "Test validation", en R. L. Thorndike (ed.). *Educational Measurement*. Washington, DC, American Council on Education.

DATTA, L.-E. (1976). "The impact of the Westinghouse/Ohio evaluation on the development of Project Head Start: an examination of the immediate and longer-term effects and how they came about", en C. C. Abt (ed.) *The Evaluation of Social Programs*.Beverly Hills, Sage Publications.

DENZIN, N. K. (1971). "The logic of naturalistic inquiry." *Social Forces 50*, 2.

DICKIE, G. (1971). *Aesthetics*, Indianapolis, Bobbs-Merrill.

DIGGS, B. J. (1974). *The State, Justice, and the Common Good*. Glenview, Il, Scott, Foresman.

EGBERT, R. (1977). Interview conducted at the Center for Instructional Research and *Curriculum* Evaluation, University of Illinois, urbana, IL, Abril 28.

EISNER, E. (1979). *The Educational Imagination*. Nueva York, Macmillan.

ELLIOTT, J. (1977). "Democratic evaluation as social criticism: or putting the judgment back into evaluation", en N. Norris (ed.) *Theory in Practice*. University of East Anglia. Centre for Applied Research in Education, Norwich, England.

© Ediciones Morata, S. L.

Scanned by CamScanner

ELLIOTT, J. (1979). "Educational accountability and the evaluation of teaching", en A. Lewy (ed.) *Evaluation Roles.*
ELMORE, R. F. (1976). "Follow Through: decision making in a large-scale social experiment". Tesis Doctoral inédita. School of Education, Harvard University.
ENNIS, R. H. (1975). Equality of educational opportunity. Urbana; College of Education, University of Illinois. (*Trabajo multicopiado.*)
——— (1973). "On causality". *Educational Researcher 2,* 6.
EPPS, E. G. (s.f.). *Memorandum to Debbie K. Walker-Re.,* Test battery for national Follow Through evaluation.
FRANKE-WIKBERG, S. (1979). The LONG project: an example of educational evaluation given from Sweden. Umea University. (*Trabajo multicopiado.*)
GARDNER, J. (1978). *On Moral Fiction.* Nueva York, Basic Books.
GARDNER, M. (1976). "Mathematical games". *Scientific American 234,* 3.
GLASS, G. V. (1972). "Educational product evaluation: a prototype format applied". *Educational Researcher 9,* 1.
GOTTSHALK, D. W. (1962). *Art and the Social Order.* Nueva York, Dover.
GUBA, E. G. (1978). *Toward a Metodology of Naturalistic Inquiry in Educational Evaluation.* Los Ángeles, Center for the Study of Evaluation, UCLA.
GUSFIELD, J. (1976). "The literary rhetoric of science". *American Sociological Review 41* (Febrero), págs. 11-33.
GUTTENTAG, M. (1973). "Subjectivity and its use in evaluation research". *Evaluation 1, 2,* págs. 60-65.
——— y K. SNAPPER (1974). "Plans, evaluations, and decisions". *Evaluation 1, 2,* págs. 58-74.
HABERMAS, J. (1973). *Theory and Practice.* Boston, Beacon. (Trad. cast.: *Teoría y praxis. Estudios de filosofía social.* Madrid, Tecnos, 1987.)
HAMILTON, D. (1977). "Making sense of curriculum evaluation", en L. Shulman (ed.) *Review of Research in Education.* Vol. 5. Itasca, IL, F. E. Peacock.
——— (1976). A science of the singular? University of Illinois. (*Trabajo multicopiado.*).
HANEY, W. (1977). *The Follow Through Evaluation: A technical History.* Cambridge, MA, Huron Institute.
HARRINGTON, P. y J. R. SANDERS (1979). Guidelines for goal-free evaluation. Evaluation Center, Western Michigan University. (*Trabajo multicopiado.*)
HERBERT, M. (1975). *Extended Pilot Trials of the Comprehensive School Mathematics Program.* St. Louis, CEMREL, Inc.
HOGBEN, D. y K. SIMPSON (1978). Evaluator authority and responsibility: a delimitation. Flinders University of South Australia. (*Trabajo multicopiado.*)
HOUSE, E. R. (1979). The role of theories of justice in evaluation-justice on strike. *Educational Theory 29,* 4.
——— (1978a). "Assumptions underlying evaluation models." *Educational Researcher 7* (Marzo), págs. 4-12.
——— (1978b). "Evaluation as scientific management in United States school reform." *Comparative Education Review 22,* 3, págs. 388-401.
——— (1977). *The Logic of Evaluative Argument.* Los Ángeles; Center for the Study of Evaluation, UCLA.
——— (1976). "Justice in evaluation." en G. V. Glass (ed.) *Evaluation Studies Annual Review.* Vol. I. Beverly Hills, Sage Publications.
——— (1974). *The Politics of Educational Innovation.* Berkeley, McCutchan.
——— (1973). "The conscience of educational evaluation", en E. R. House (ed.) *School Evaluation.* Berkeley, McCutchan.

© Ediciones Morata, S. L.

HOUSE, E. R. y N. S. CARE (1979). "Fair evaluation agreement". *Educational Theory 29,* 3.
———, G. V. GLASS, L. D. MCLEAN y D. WALKER (1978). "No simple answer: critique of the Follow Through evaluation." *Harvard Educational Review 48* (Mayo), págs. 128-160.
——— y D. HOGBEN (1974). *A goal-free evaluation for "Me and My Environment"*, Formative Evaluation Informe n.º 3, Biological Sciences Curriculum Study, páginas 14-16.
——— y R. S. MAYER (1980). "A critique of the educational imagination in evaluation". *Journal of Aesthetic Education* (Julio).
———, J. STEELE, y C. T. KERINS (1971). *The Gifted Classroom.* Center for Instructional Research and Curriculum Evaluation, University of Illinois, Urbana.
———, W. RIVERS, y D. STUFFLEBEAM (1974). "An assessment of the Michigan accountability system". *Phi Delta Kappan 55* (Junio), págs. 663-669.
ITTELSON, W. y H. CANTRIL (1954). *Perception: A transactional Approach.* Garden City, NY: Doubleday.
JENKINS, D. y B. O'TOOLE (1978). "Curriculum evaluation, literary criticism, and the paracurriculum", en G. Willis (ed.) *Qualitative Evaluation.* Berkeley, McCutchan.
KALLOS, D. (1978). "Notes on schooling, curriculum, and teaching", en G. Willis (ed.) *Qualitative Evaluation.* Berkeley, McCutchan.
KARIER, C. J. (1973). "Testing for control and order in the corporate state", en C. J. Karier, P. Violas, y J. Spring (eds.) *Roots of Crisis: American Education in the Twentieth Century.* Chicago, Rand McNally.
KEARNEY, C. P., D. L. DONOVAN, y T. H. FISHER (1974). "In defense of Michigan's accountability program." *Phi Delta Kappan 51* (Septiembre).
KELLY, E. F. (1978). "Curriculum criticism and literary criticism: comments on the anthology", en G. Willis (ed.) *Qualitative Evaluation.* Berkeley, McCutchan.
KELLY, E. G. (1980). "Evaluation as persuasion". *Educational Evaluation and Policy Analysis 2,* 5.
KEMMIS, S. (1976). "Evaluation and evolution in knowledge about educational programs". Tesis Doctoral inédita. University of Illinois.
KOGAN, M. y T. PACKWOOD (1974). *Advisory Councils and Committees in Education.* Londres, Routledge & Kegan Paul.
KOHLBERG, L., y R. S. MAYER (1972). "Development as the aim of education." *Harvard Educational Review 42,* págs. 449-496.
KRULEE, G. K. (1973). *An Organizational Analysis of Project Follow Through Evanston,* IL, Northwestern University.
KUHN, T. (1970). The Structure of Scientific Revolutions. Chicago, Univ. of Chicago Press. (Trad. cast.: *La estructura de las revoluciones científicas.* Madrid, Fondo de Cultura Económica, 1980, 5.ª reimpr.)
LANGER, S. K. (1942). *Philosophy in a New Key.* Nueva York: Mentor.
LEVIN, H. M. (1972). "The social science objectivity gap." *Saturday Review* (Noviembre 11), págs. 44-51.
LUKES, S. (1974). *Power.* Londres, Macmillan. (Trad. cast.: *El poder. Un enfoque radical.* Madrid, Siglo XXI, 1985.)
LUNDGREN, U. P. (1977). *Model Analysis of Pedagogical Processes.* Estocolmo, Stockholm Institute of Education.
MACDONALD, B. (1977). "A political classification of evaluation studies", en D. Hamilton, D. Jenkins, C. King, B. MacDonald, y M. Parlett (eds.) *Beyond the Numbers Game.* Londres, Macmillan.

© Ediciones Morata, S. L.

Scanned by CamScanner

MacDonald, B. (1974). *Evaluation and the Control of Education.* Norwich, Inglaterra, Centre for Applied Research in Education. (Trad. cast.: "La evolución y el control de la educación", en J. Gimeno Sacristán y A. Pérez Gómez: *La enseñanza: su teoría y su práctica.* Madrid, Akal, 1983, págs. 467-478.)
——— y R. Walker (1974). "Case-study and the social philosophy of educational research." University of East Anglia, Norwich, Inglaterra.
Macpherson, C. B. (1966). *The Real World of Democracy.* Oxford, Inglaterra, Oxford Univ. Press.
Mager, R. F. (1972). *Goal Analysis.* Belmont, CA, Fearon.
——— (1962). *Preparing Objectives for Programmed Instruction.* San Francisco, Fearon.
Marcuse, H. (1978). *The Aesthetic Dimension.* Boston, Beacon.
McCutcheon, G. (1978). "Of solar systems, responsibility and basics: an educational criticism of Mr. Clement's fourth grade", en G. Willis (ed.) *Qualitative Evaluation.* Berkeley, McCutchan.
McLaughlin, M. W. (1975). *Evaluation and Reform.* Cambridge. MA: Ballinger.
Mill, J. S. (1893). *A System of Logic.* Nueva York, Harper. (Trad. cast.: *El utilitarismo.* Madrid, Alianza, 1984.)
——— (1861). *Utilitarianism.* Indianapolis, Bobbs-Merrill.
Morgan, T. (1976). "The good life (along the San Andreas Fault)". *New York Times Magazine* (Julio 4).
Morris, L., C. Fitz-Gibbon, y M. Henerson (1978). *Program Evaluation Kit.* Beverly Hills, Sage Publications.
National Study of Secondary School Evaluation (1969). *Evaluative Criteria* Washington, DC, Author.
National Study of School Evaluation (1978). *Evaluative Criteria.* Arlington, VA, Author.
Novak, M. (1971). *Ascent of the Mountain, flight of the Dove.* Nueva York, Harper & Row.
O'Connor, J. J. (1976). "Wiseman's latest film is another reality fiction". *New York Times* (Noviembre 7).
Ortony, A. (1975a). Knowledge, language and teaching. University of Illinois. (*Trabajo multicopiado.*)
——— (1975b). "Why metaphors are necessary and not just nice". *Educational Theory* (Invierno).
Owens, T. R. (1973). "Educational evaluation by adversary proceeding", en E. R. House (ed.) *School Evaluation.* Berkeley, McCutchan.
Parker, J. H. (1960). "The problem of esthetic form", en M. Rader (ed.). *A Modern Book of Esthetics.* Nueva York: Holt, Rhinehart & Winston.
Parlett, M. y D. Hamilton (1977). "Evaluation as illumination: a new approach to the study of innovatory programmes", en D. Hamilton y cols. (eds.) *Beyond the Numbers Game,* Londres, Macmillan. (Trad. cast.: "La evaluación como iluminación", en J. Gimeno Sacristán y A. Pérez Gómez: *La enseñanza: su teoría y su práctica.* Madrid, Akal, 1983, págs. 450-466.)
Patton, M. Q. (1978). *Utilization-Focused Evaluation.* Beverly Hills, Sage Publications.
Paulston, R. G. (En prensa) "Paradigm conflict in the assessment of educational reform: evaluation as social critique". *Educational Evaluation and Policy Analysis.*
Perelman, C. y L. Olbrechts-Tyteca. (1969). *The New Rhetoric: A Treatise on Argumentation.* Notre Dame. IN, Univ. Dame Press.
Petrie, H. G. (1979). "COPE begins second evaluation cycle." *Illini Week 1* (Octubre 12), pág. 5.

© Ediciones Morata, S. L.

Petrie, H. G. (1976). Metaphorical models of mastery: or, how to learn to do the problems at the end of the chapter of the physics textbook. University of Illinois. (*Trabajo multicopiado.*)
Polanyi, M. (1958). *Personal Knowledge.* Chicago, Univ. of Chicago Press.
——— y H. Prosch (1975). *Meaning.* Chicago: Univ. of Chicago Press.
Popham, W. J. (1975). *Educational Evaluation.* Englewood Cliffs, NJ, Prentice-Hall.
——— y D. Carlson (1977). "Deep dark deficits of the adversary evaluation model". *Educational Researcher 6* (Junio), págs. 3-6.
Popham, W. J. y R. L. Ebel (1978). "Annual meeting presidential debate." *Educational Researcher 7, 11,* pág. 3.
Rawls, J. (1971). *A theory of Justice.* Cambridge, MA: Harvard Univ. Press. Informe de the Fourth Annual Follow Through Working Conference (1974). High Scope Educational Research Foundation. (Trad. cast.: *Teoría de la justicia.* Madrid: Fondo de Cultura Económica, 1979.)
Education Daily (1977). "Results of Follow Through evaluation disappointing but useful, OE says." Agosto 10.
Rivlin, A. M. (1971). *Systematic Thinking for Social Action.* Washington, DC, Brookings Institution.
Rossi, P. H., H. E. Freeman, y S. R. Wright (1979). *Evaluation: A Systematic Approach.* Beverly Hills, Sage Publications.
Schön, D. (1979). "Generative metaphor: a perspective of problem-setting in social policy", en A. Ortony (ed.) *Metaphor and Thought.* Nueva York: Cambridge Univ. Press.
Scriven, M. (1976). "Bias control systems in evaluation." Trabajo presentado a The annual meeting of the American Educational Research Association, San Francisco.
——— (1976a). "Evaluation bias and its control", en G. V. Glass (ed.). *Evaluation Studies Review Annual,* Vol. 1. Beverly Hills, Sage Publications.
——— (1976b). "Maximizing the process of causal investigations: the modus operandi method", en G. V. Glass (ed.). *Evaluation Studies Review Annual,* Vol. I, Beverly Hills, Sage Publications.
——— (1973). "Goal free evaluation", en E. R. House (ed.). *School Evaluation.* Berkeley, McCutchan.
——— (1972a). "Educational product re-evaluation." *Educational Researcher 1,* 5.
——— (1972b). "Objectivity and subjectivity in educational research", en L. G. Thomas (ed.). *Philosophical Redirection of Educational Research.* National Society for the Study of Education.
Shapley, D. (1976). "Earthquakes: Los Angeles prediction suggests faults in federal policy." *Science 192* (mayo).
Simons, H. (1977). "Building a social contract: negotiation and participation in condensed field research", en N. Norris (ed.) *Theory in Practice. University of East Anglia.* Centre for Applied Research in Education, Norwich, Inglaterra.
Smith, L. M. y P. A. Pohland (1974). "Education, technology, and the rural highlands", en D. Sjogren (ed.). *Four Evaluation Examples: Anthropological,* Economic, Narrative, and Portrayal. AERA Colección Monografías. Curriculum Evaluation, No 7. Chicago, Rand McNally.
Smith, M. L., R. Gabriel, J. Schott, y W. L. Podia (1976). "Evaluation of the effects of outward bound", en G. V. Glass (ed.). *Evaluation Studies Review Annual,* Vol. 1. Beverly Hills, Sage Publications.

© Ediciones Morata, S. L.

Scanned by CamScanner

SNYDER, W. P. (1967). *Case Studies in Military Systems Analysis.* Washington, DC, Industrial College of the armed Forces.
SOAR, R. S. (1973). Informe final: *Follow Through Classroom Process Measurement and Pupil Growth (1970-1971).* Gainesville, FL, Institute for Development of Human Resources.
SONTAG, S. (1977). *On Photography.* Nueva York: Farrer, Straus & Giroux. (Trad. cast.: *Sobre la fotografía.* Barcelona, Edhasa, 1989, 3.ª ed.)
STAKE, R. E. (1978). "The case study method in social inquiry". *Educational Researcher 7* (Febrero), págs. 5-8.
——— (1976). *Evaluating Educational Programmes: The Need and the Response.* Washington DC, OECD Publications Center.
——— (1975a). *Evaluating the Arts in Education: A Responsive Approach.* Columbus, OH, Merrill.
——— (1975b). Some alternative presumptions. Center for Instructional Research and Curriculum Evaluation, University of Illinois. *(Trabajo multicopiado.)*
——— (1973). "Measuring what learners learn", en E. R. House (ed.) *School Evaluation.* Berkeley, McCutchan.
——— y J. EASLEY (1978). *Case studies in science education.* Center for Instructional Research and Curriculum Evaluation, University of Illinois.
STALLINGS, J. A. (1973). *Follow Through classroom observation evaluation 1971-72.* Menlo Park, CA, Stanford Research Institute.
Stanford Research Institute (1972). Summary of plans for longitudinal evaluation of the national Follow Through program. 1969-1970. *(Trabajo multicopiado.)*
Steering Committee for a National Follow Through Parent Advisory Council (1972). Parent Resolution. Asamblea de Denver, Noviembre 1. *(Trabajo multicopiado.)*
STENZEL, N. (1979). *Committee Hearings as an evaluation format.* Portland, OR, Northwest Regional Educational Laboratory.
STRAUCH, R. E. (1976). "A critical look at quantitative methodology." *Policy Analysis 2,* 1.
STRIKE, K. (1979). "The role of theories of justice in evaluation —why a house is not a home." *Educational Theory 29,* 3.
STUFFLEBEAM, D. L. (1973). "An introduction to the PDK book", en B. Worthen y J. R. Sanders (eds.). *Educational Evaluation: Theory and Practice.* Worthington, OH, Charles A. Jones.
——— (1969). "Evaluation as enlightenment for decision-making", en Beatty (ed.) *Improving Educational Assessment and an Inventory of Affective Behaviors.* Washington, DC, Association for Supervision and Curriculum Development, NEA.
SUCHMAN, E. A. (1967). *Evaluative Research.* Nueva York, Russell Sage Foundation.
SULLIVAN, E. V. (1977). Kohlberg's Structuralism. Ontario, Ontario Institute for Studies in Education.
TAYLOR, P. W. (1961). *Normative Discourse.* Englewood Cliffs, NJ, Prentice-Hall.
TEMPLIN, P. S. (1978). Photography: can it provide program portrayal? Presentado a American Education Research Association, Toronto, Marzo.
THURSTON, P. (1978). "Revitalizing adversary evaluation: deep dark deficits or muddled mistaken musings". *Educational Research 7* (Julio-Agosto), págs. 3-8.
TOLSTOY (1960). "The communication of Emotion", en M. Rader (ed.) *A modern Book of Esthetics.* Nueva York: Holt, Rinehart & Winston.
TURNER, T. (1973). "Piaget's structuralism". *American Anthropologist 75,* págs. 351-373.
TYLER, R. W. (1950). *Basic Principles of Curriculum and Instruction.* Chicago, Univ. of Chicago Press. (Trad. cast.: *Principios básicos del currículo.* Buenos Aires, Troquel, 1977, 2.ª ed.)
TYLER, R. S., LAPAN, J. MOORE, C. W. RIVERS, y D. SKIBO (1978). "Impact of minimun competence testing in Florida." *Today's Education 67* (Septiembre-Octubre), págs. 30-38.
U.S. Office of Education (1969). *Follow Through Program Manual.* Washington, DC: Department of Health, Education and Welfare.
VALLANCE, E. (1978). "Scanning horizons and looking at weeks: a critical description of 'The Great Plains Experience'", en G. Willis (ed.) *Qualitative Evaluation.* Berkeley: McCutchan.
WAKS, L. J. (1975). "Educational objectives and existential heros", en R. A. Smith (ed.). *Regaining Educational Leadership.* Nueva York; John Wiley.
WALKER, D. R. y J. SCHAFFARZICK (1974). "Comparing curricula." *Review of Educational Research 44* (Invierno).
WALLER, J. J. (1967). "Identification of problem drinkers among drunken drivers". *Journal of the American Medical Association,* págs. 124-130.
WEIKART, D. P. y B. A. BANET (1975). "Model design problems in Follow Through", en A. M. Rivlin y P. M. Timpane (eds.) *Planned Variation in Education. Washington,* DC, Brookings Institution.
WEIZENBAUM, J. (1976). *Computer Power and Human Reason.* San Francisco, W. H. Freeman. (Trad. cast.: *La frontera entre el ordenador y la mente.* Madrid, Pirámide, 1977.)
WELCH, W. (1978). Goal-free formative evaluation-an example. Presentado a The American Education Research Association. Toronto, Marzo.
WILLIS, G. (ed.) (1978). *Qualitative Evaluation.* Berkeley: McCutchan.
WIRTZ, W. (1977). *On Further Examination.* Nueva York: College Entrance Examination Board.
WOLCOTT, H. F. (1977). *Teachers vs. Technocrats.* Eugene. OR, Center for Educational Policy and Management, University of Oregon.
WOLF, R. L., (1974). "The use of judicial evaluation methods in the formulation of educational policy". *Educational Evaluation and Policy Analysis* 1 (mayo-Junio), págs. 19-28.
——— (1975). "Trial by jury: the process". *Phi Delta Kappan 57* (Noviembre), págs. 185-187.
WOLFF, R. P. (1968). *The Poverty of Liberalism.* Boston, Beacon.
WOODWARD, B. y C. BERNSTEIN (1976). *The Final Days.* Nueva York, Simon & Schuster.
WORTHEN, B. R. y J. R. SANDERS (1973). *Educational Evaluation: Theory and Practice.* Worthington, OH: Charles A. Jones.

© Ediciones Morata, S. L.

© Ediciones Morata, S. L.

Scanned by CamScanner

Otras obras de Ediciones Morata de interés

Aitken, J. y Mills, G.: *Tecnología creativa,* 1994.
Apple, M. W.: *Escuelas democráticas,* 1997.
——— *Política cultural y educación,* 1996.
Arnold, P. J.: *Educación física, movimiento y* curriculum, 1991.
Bale, J.: *Didáctica de la geografía en la escuela primaria,* (2.ª ed.), 1996.
Ball, S.: *Foucault y la educación,* (3.ª ed.), 1997.
Baudelot, Ch. y Establet, R.: *El nivel educativo sube,* 1989.
Bennett, N.: *Estilos de enseñanza y progreso de los alumnos,* 1979.
Bernstein, B.: *La estructura del discurso pedagógico,* (2.ª ed.), 1994.
Blanco, A.: *Cinco tradiciones en la psicología social,* (2.ª ed.), 1995.
Bruner, J.: *Desarrollo cognitivo y educación,* (2.ª ed.), 1995.
Carr, W.: *Una teoría para la educación,* 1996.
Connell, R.: *Escuelas y justicia social,* 1997.
Cook, T. D. y Reichardt, Ch.: *Métodos cualitativos y cuantitativos en investigación evaluativa,* (2.ª ed.), 1995.
Decroly, O.: *El juego educativo,* (3.ª ed.), 1993.
Dewey, J.: *Democracia y educación,* (2.ª ed.), 1996.
Donaldson, M.: *La mente de los niños,* (3.ª ed.), 1993.
Driver, R., Guesne, E. y Tiberghien, A.: *Ideas científicas en la infancia y la adolescencia,* (2.ª ed.), 1992.
Egan, K.: *Fantasía e imaginación: su poder en la enseñanza,* 1994.
Elliott, J.: *La investigación-acción en educación,* (2.ª ed.), 1994.
——— *El cambio educativo desde la investigación-acción,* (2.ª ed.), 1997.
Escuelas infantiles de Reggio Emilia: *La inteligencia se construye usándola,* 1995.
Fernández Dols, J. M.: *Patrones para el diseño de la psicología social,* 1990.
Fernández Pérez, M.: *Evaluación y cambio educativo: el fracaso escolar,* (4.ª ed.), 1995.
Freinet, C.: *La escuela moderna francesa. Una pedagogía moderna de sentido común. Las invariantes pedagógicas,* 1996.
Gimeno Sacristán, J.: *El* curriculum: *una reflexión sobre la práctica,* (6.ª ed.), 1997.
——— *La pedagogía por objetivos: obsesión por la eficiencia,* (9.ª ed.), 1997.
——— *La transición a la educación secundaria,* (2.ª ed.), 1997.
——— **y Pérez Gómez, A. I.:** *Comprender y transformar la enseñanza,* (5.ª ed.), 1996.
Goetz, J. P. y LeCompte, M. D.: *Etnografía y diseño cualitativo en investigación ed.,* 1988.
González Portal, M. D.: *Conducta prosocial: evaluación e intervención,* (2.ª ed.), 1995.

© Ediciones Morata, S. L.

Gore, J.: *Controversias entre las pedagogías,* 1996.
Graves, D. H.: *Didáctica de la escritura,* (2.ª ed.), 1996.
Grundy, S.: *Producto o praxis del* curriculum, (2.ª ed.), 1994.
Halliday, J.: *Educación, gerencialismo y mercado,* 1995.
Hargreaves, A.: *Profesorado, cultura y postmodernidad,* 1996.
Hargreaves, D. J.: *Infancia y educación artística,* 1991.
Harlen, W.: *Enseñanza y aprendizaje de las ciencias,* (2.ª ed.), 1994.
Hegarty, S.: *Aprender juntos: la integración escolar,* (2.ª ed.), 1994.
Hicks, D.: *Educación para la paz,* 1993.
Hyde, J.: *Psicología de la mujer,* 1995.
Inhelder, B.: *Aprendizaje y estructuras del conocimiento,* (2.ª ed.), 1996.
Jackson, Ph. W.: *La vida en las aulas,* (4.ª ed.), 1996.
Kemmis, S.: *El* curriculum: *más allá de la teoría de la reproducción,* (2.ª ed.), 1993.
Liston, D. P. y Zeichner, K. M.: *Formación del profesorado y condiciones sociales de la escolarización,* (2.ª ed.), 1997.
Loughlin, C. E. y Suina, J. H.: *El ambiente de aprendizaje: diseño y organización,* (3.ª ed.), 1995.
Lundgren, U. P.: *Teoría del* curriculum *y escolarización,* (2.ª ed.), 1997.
McCarthy, C.: *Racismo y curriculum,* 1994.
McCormick, R. y James, M.: *Evaluación del* curriculum *en los centros escolares,* 1995.
Moyles, J. R.: *El juego en la educación infantil y primaria,* 1990.
Oléron, P.: *El niño: su saber y su saber hacer,* 1987.
Orton, A.: *Didáctica de las matemáticas,* (2.ª ed.), 1996.
Perrenoud, Ph.: *La construcción del éxito y del fracaso escolar,* (2.ª ed.), 1996.
Piaget, J.: *La composición de las fuerzas y el problema de los vectores,* 1975.
——— *Psicología del niño,* (13.ª ed.), 1994.
——— *La representación del mundo en el niño,* (6.ª ed.), 1984.
——— *La toma de conciencia,* (3.ª ed.), 1985.
Pimm, D.: *El lenguaje matemático en el aula,* 1990.
Popkewitz, Th. S.: *Sociología política de las reformas educativas,* (2.ª ed.), 1997.
Pozo, J. I.: *Teorías cognitivas del aprendizaje,* (4.ª ed.), 1996.
Rivas, F.: *Psicología vocacional: enfoques del asesoramiento,* (2.ª ed.), 1993.
Sanuy, M.: *Aula sonora. Hacia una educación musical en primaria,* (2.ª ed.), 1996.
Saunders, R. y Bingham-Newman, A. M.: *Perspectivas piagetianas en la educación infantil,* 1989.
Scraton, Sh.: *Educación física de las niñas,* 1995.
Selmi, L. y Turrini, A.: *La escuela infantil a los tres años,* (3.ª ed.), 1997.
——— y ——— *La escuela infantil a los cuatro años,* (3.ª ed.), 1997.
——— y ——— *La escuela infantil a los cinco años,* (2.ª ed.), 1995.
Stenhouse, J.: *Investigación y desarrollo del* curriculum, (3.ª ed.), 1993.
——— *La investigación como base de la enseñanza,* (3.ª ed.), 1996.
Tann, C. S.: *Diseño y desarrollo de unidades didácticas en la escuela primaria,* (2.ª ed.), 1993.
Torres, J.: *Globalización e interdisciplinariedad,* (2.ª ed.), 1996.
——— *El* curriculum *oculto,* (5.ª ed.), 1996.
Turner, J.: *Redescubrir el grupo social,* 1990.
Tyler, W.: *Organización escolar,* (2.ª ed.), 1996.
Usher, R. y Bryant, I.: *La educación de adultos como teoría, práctica e investigación. El triángulo cautivo,* 1992.
VV.AA.: *Volver a pensar la educación* (2 vols.), 1995.
Walker, R.: *Métodos de investigación para el profesorado,* (2.ª ed.), 1997.
Willis, A. y Ricciuti, H.: *Orientaciones para la escuela infantil de cero a dos años,* (2.ª ed.), 1997.
Zimmermann, D.: *Observación y comunicación no verbal en la esc. inf.,* (2.ª ed.), 1992.

© Ediciones Morata, S. L.

Scanned by CamScanner

En esta obra Ernest R. HOUSE se propone explicar el caótico marco en el que frecuentemente se desenvuelve la evaluación. Presenta la primera taxonomía detallada de los distintos enfoques que existen en este campo y desarrolla una profunda crítica de los mismos. Analiza ocho modelos fundamentales a la luz de las normas, principios y técnicas de la evaluación, valorando la adecuación y debilidades de cada uno y explicitando sus supuestos epistemológicos, éticos y políticos.
Quienes están interesados en conocer el estado actual de toda la problemática de los procesos de evaluación educativa, y en cómo valorar la calidad y justicia de las evaluaciones, encontrarán de interés los análisis expuestos por el autor, que contribuyen a desarrollar un ejercicio profesional más responsable y reflexivo.

00000
9 788471 123893
ISBN 84-7112-389-4

Colecciones: Psicología
Pedagogía
Sociología
Psiquiatría
Matemáticas
Medicina
Sexología
Filosofía
Religión
Política

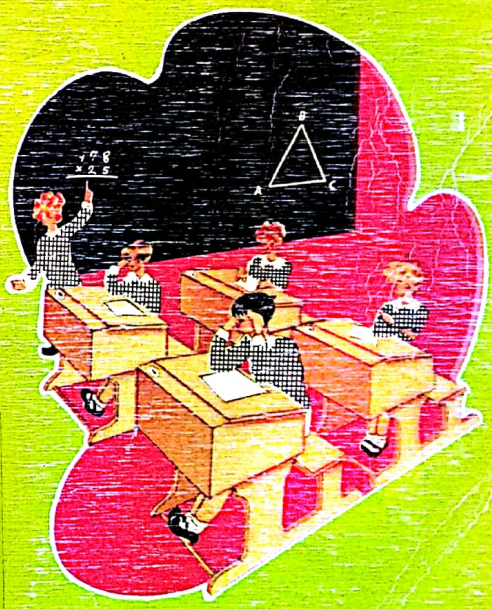
EDICIONES MORATA, S. L.
Mejía Lequerica, 12. 28004 - Madrid

Evaluación, ética y poder E. R. House

361.6
H 842
c.1

E. R. House

Evaluación, ética y poder



SEGUNDA EDICIÓN

Morata

Scanned by CamScanner