

Programa de  
Especialización en Teoría,  
Métodos y Técnicas de

**INVESTIGACION  
SOCIAL**



# Investigación evaluativa

**Santiago Correa Uribe**

**Antonio Puerta Zapata**

**Bernardo Restrepo Gómez**

INSTITUTO COLOMBIANO PARA EL FOMENTO  
DE LA EDUCACIÓN SUPERIOR, ICFES

DIRECCIÓN GENERAL  
Calle 17 No. 3-40 A.A. 6913  
Teléfonos: 3387338 - 3387360  
Fax: 2836778  
Bogotá, Colombia

HEMEROTECA NACIONAL UNIVERSITARIA CARLOS LLERAS RESTREPO  
SUBDIRECCIÓN DE FOMENTO Y DESARROLLO DE LA EDUCACIÓN SUPERIOR  
Avenida Eldorado No. 44A-40  
Teléfono: 3689780  
Telefax: 3680028  
Bogotá, Colombia

DIRECTOR DEL ICFES: Daniel Bogoya Maldonado

SUBDIRECTORA DE FOMENTO  
Y DESARROLLO DE LA EDUCACIÓN SUPERIOR: María de Jesús Restrepo

COPYRIGHT: ICFES 1996  
Módulos de Investigación Social

ISBN: 958-9329-09-8 Obra completa  
ESPECIALIZACIÓN EN TEORÍA, MÉTODOS  
Y TÉCNICAS DE INVESTIGACIÓN SOCIAL

ISBN: 958-9329- Módulo seis  
INVESTIGACIÓN EVALUATIVA

Composición electrónica: ARFO Editores e Impresores Ltda.  
Diciembre de 2002

El proyecto para la formación de recursos humanos en la educación superior titulado “Especialización en Teoría, métodos y técnicas de investigación social” con la colaboración de ASCUN, fue realizado por el ICFES, para lo cual se conformó el siguiente equipo de trabajo:

<b>Dirección y coordinación académica:</b>	GUILLERMO BRIONES
<b>Dirección y coordinación del proyecto:</b>	MARÍA DE JESÚS RESTREPO ALZATE Profesional especializada Coordinadora Grupo de Fomento del ICFES
<b>Asesor:</b>	MIGUEL RAMÓN MARTÍNEZ Asesor Subdirección Gral. Técnica y de Fomento ICFES

#### GRUPO DE AUTORES

GUILLERMO BRIONES Universidad de Chile	MÓDULO 1	EPISTEMOLOGÍA DE LAS CIENCIAS SOCIALES
GUILLERMO HOYOS VÁSQUEZ Universidad Nacional de Colombia GERMÁN VARGAS GUILLÉN Universidad Pedagógica Nacional	MÓDULO 2	LA TEORÍA DE LA ACCIÓN COMUNICATIVA COMO NUEVO PARADIGMA DE INVESTIGACIÓN EN CIENCIAS SOCIALES
GUILLERMO BRIONES Universidad de Chile	MÓDULO 3	METODOLOGÍA DE LA INVESTIGACIÓN CUANTITATIVA EN LAS CIENCIAS SOCIALES
CARLOS A. SANDOVAL CASILIMAS INER Universidad de Antioquia	MÓDULO 4	INVESTIGACIÓN CUALITATIVA
HERNÁN HENAO DELGADO y LUCELY VILLEGAS VILLEGAS INER Universidad de Antioquia	MÓDULO 5	ESTUDIOS DE LOCALIDADES
SANTIAGO CORREA URIBE Universidad de Antioquia ANTONIO PUERTA ZAPATA INER Universidad de Antioquia BERNARDO RESTREPO GÓMEZ Universidad de Antioquia	MÓDULO 6	INVESTIGACIÓN EVALUATIVA
BERNARDO RESTREPO GÓMEZ Universidad de Antioquia	MÓDULO 7	INVESTIGACIÓN EN EDUCACIÓN



# Contenido

<b>Presentación .....</b>	<b>9</b>
<b>Introducción .....</b>	<b>1</b>
<b>Estructura general del módulo .....</b>	<b>13</b>
Objetivo general .....	14
Objetivos específicos .....	14
Mapa conceptual del módulo .....	15

## PRIMERA UNIDAD

### Los conceptos de institución - evaluación e investigación

<b>evaluativa y aproximación a una tipología de evaluación .....</b>	<b>17</b>
Introducción .....	19
Objetivos de la unidad .....	20
Mapa conceptual .....	21
1. Concepto de Institución .....	23
2. Concepto de evaluación .....	27
3. Investigación y evaluación .....	29
4. Tipos de evaluación .....	37
<b>Autoevaluación No. 1 .....</b>	<b>46</b>
<b>Actividad de evaluación .....</b>	<b>47</b>
<b>Lecturas complementarias .....</b>	<b>48</b>
<b>Bibliografía .....</b>	<b>49</b>

## SEGUNDA UNIDAD

### Itinerario de la evaluación de programas

<b>como disciplina .....</b>	<b>51</b>
------------------------------	-----------

Mapa conceptual .....	52
Introducción .....	53
Objetivos de la unidad .....	54
Desarrollo de la unidad .....	54
Coda .....	72
<b>Lecturas complementarias</b> .....	73
<b>Autoevaluación No. 2</b> .....	74
<b>Actividades de Evaluación</b> .....	75
<b>Bibliografía</b> .....	76

## TERCERA UNIDAD

### **Paradigmas - enfoques - orientaciones**

#### **y modelos evaluativos** ..... 81

Mapa conceptual .....	82
Introducción .....	83
Objetivos de la unidad .....	84
1. Paradigmas de la investigación evaluativa .....	85
2. Enfoques de evaluación .....	87
3. Orientaciones predominantes en la investigación evaluativa .....	88
<b>Autoevaluación No. 3</b> .....	123
<b>Actividades de evaluación</b> .....	124
<b>Lecturas complementarias</b> .....	125
<b>Referencias bibliográficas</b> .....	126

## CUARTA UNIDAD

### **Diseño metodológico, implementación, análisis**

#### **e informes de la evaluación** ..... 129

Mapa conceptual .....	130
Introducción .....	131
Objetivos de la unidad .....	132
Etapas de la evaluación de programas .....	134
1. Especificación del marco de referencia .....	134
1.1. El programa de acción .....	134
1.2. La organización gestora .....	137
1.3. Los factores del contexto .....	142

2. La formulación del proyecto de evaluación .....	143
2.1. El objeto de la evaluación .....	143
2.2. El diseño de la evaluación. ....	144
2.3. El cronograma .....	151
3. Implementación del proyecto de evaluación .....	152
3.1. El equipo de evaluación .....	152
3.2. Ubicación de las unidades de análisis .....	155
3.3. Elaboración de instrumentos de recolección de información .....	155
3.4. Pruebas piloto .....	155
3.5. Preparación de auxiliares .....	156
4. Ejecución de la evaluación .....	156
4.1. Aplicación de los instrumentos de medición y registro .....	157
4.2. Procesamiento de la información .....	157
4.3. Análisis de los resultados .....	158
4.4. Conclusiones y recomendaciones .....	158
4.5. Informes de evaluación .....	159
<b>Autoevaluación No. 4.</b> .....	162
<b>Lecturas complementarias</b> .....	162
<b>Actividades de evaluación</b> .....	163
<b>Bibliografía</b> .....	164

## LECTURAS COMPLEMENTARIAS

Lectura complementaria N° 1 .....	169
Lectura complementaria N° 2 .....	211
Lectura complementaria N° 3 .....	230
Lectura complementaria N° 4 .....	240
Lectura complementaria N° 5 .....	273





# Presentación

En cumplimiento de su misión y de sus funciones, el Instituto Colombiano para el Fomento de la Educación Superior ha realizado diversas acciones que permitan a las instituciones de Educación Superior no solamente enfrentar los retos del desarrollo científico, técnico y tecnológico sino hacer del conocimiento y de la cultura, ejes de calidad y desarrollo. Esto implica necesariamente fomentar procesos de formación que además de articular y hacer coherentes los campos de conocimiento y sus campos específicos de prácticas, contribuyan a identificar problemas en diversos contextos y a encontrar o crear soluciones a partir de procesos de reflexión.

Dentro de este contexto, el ICFES promovió la creación del Programa de Especialización en teoría, métodos y técnicas de investigación social, ofrecido en la actualidad por cinco universidades públicas a saber: Universidad de Antioquia, Universidad del Valle, Universidad Industrial de Santander, Universidad de Cartagena y Universidad Pedagógica Nacional, las cuales atendieron el reto propuesto por el Instituto y después de siete años de iniciado el programa, en conjunto han graduado más de 300 especialistas.

La estructura temática de esta serie de Investigación Social la constituyen siete módulos orientados al desarrollo de competencias para la apropiación de metodologías y técnicas de investigación fundamentadas por las teorías que enmarcan las ciencias sociales. Los primeros cuatro módulos ofrecen elementos que permiten profundizar e indagar sobre el objeto y los procesos de conocimiento en las Ciencias Sociales, las principales escuelas que han surgido a través de los tiempos y las diferentes perspectivas de análisis que conducen a la comprensión de los hechos sociales. Los tres últimos módulos orientan la aplicación de estas ciencias en el campo de la investigación educativa, la evaluación y el estudio de localidades.

Dada la pertinencia y vigencia de esta serie y aprovechando el desarrollo de las tecnologías de información y comunicación que nos permite una socialización más amplia y ágil de sus contenidos, la presentamos hoy en formato CD ROM con la expectativa que esta estrategia de aprendizaje sea aprovechada por un mayor número de profesores y estudiantes de las instituciones de educación superior del país.

DANIEL BOGOYA M.  
Director General ICFES



# Introducción

La evaluación es una ciencia social de gran actualidad cuyo potencial no se discute, pero aún sigue en permanente proceso de construcción.

Como disciplina en construcción permite analizar su decurso y consolidación a través de la evolución histórica de sus conceptos y procesos, la constitución de una comunidad académica y profesional que cada vez se internacionaliza más con la aparición de conferencias internacionales, asociaciones, agencias, centros e instituciones especializadas en este campo, con una vasta producción intelectual y aportes metodológicos plasmados en un sinnúmero de publicaciones especializadas.

La investigación evaluativa se convierte, en la actualidad, en una importante fuente de conocimientos y directrices, en las diversas actividades e instituciones de las sociedades modernas porque indica el grado de eficiencia o deficiencia de los programas y señala el camino para su reformulación y valoración del éxito alcanzado por los esfuerzos realizados.

La investigación evaluativa se vale de los métodos y el instrumental de la investigación social; por lo tanto, su desarrollo sigue sus mismas evoluciones, lo cual le permite una aproximación permanente a criterios de cientificidad.

En consonancia con lo anterior, el énfasis del módulo se pone en las preocupaciones centrales de la investigación evaluativa, su evolución histórica, los enfoques evaluativos y la metodología de la evaluación.



## **Estructura general del módulo**

El módulo consta de cuatro unidades que en una secuencia lógica pretenden dar cuenta de los aspectos antes mencionados.

La primera unidad aborda la conceptualización básica de la investigación evaluativa a través del desarrollo de los conceptos de institución, programa, evaluación e investigación evaluativa, así como también de los propósitos de la evaluación y tipos de evaluación.

La segunda unidad da cuenta de la historia de la evaluación en el mundo y en Colombia. Inicia con un análisis de las etapas históricas del desarrollo de la evaluación de programas y concluye con una detallada presentación de los momentos de la evaluación en Colombia con sus proyectos más relevantes en los diferentes campos del quehacer evaluativo.

La tercera unidad parte del análisis de los paradigmas de la investigación y su incidencia en la investigación evaluativa para adentrarse en sus enfoques y caracterizar las orientaciones predominantes y los modelos subyacentes en cada una de ellas.

La cuarta unidad da cuenta del proceso metodológico de la investigación evaluativa o evaluación de programas. Se definen cuatro grandes momentos o etapas del proceso metodológico con sus correspondientes subrutinas. Estos momentos son: la formalización del marco de referencia, la formulación, la implementación y la ejecución del proyecto de evaluación.

Cada unidad parte de unos objetivos específicos y un mapa conceptual, continuando con el desarrollo de los contenidos, los cuales serán complementados con lecturas de afianzamiento y ejemplificación de los temas de la unidad. Presenta a continuación la autoevaluación sobre los contenidos y unas actividades de evaluación productiva que pretenden dar aplicación y dotar al estudiante de la experiencia requerida para el ejercicio de la investigación evaluativa. Para terminar, aparecen las referencias bibliográficas utilizadas en el texto de la unidad.

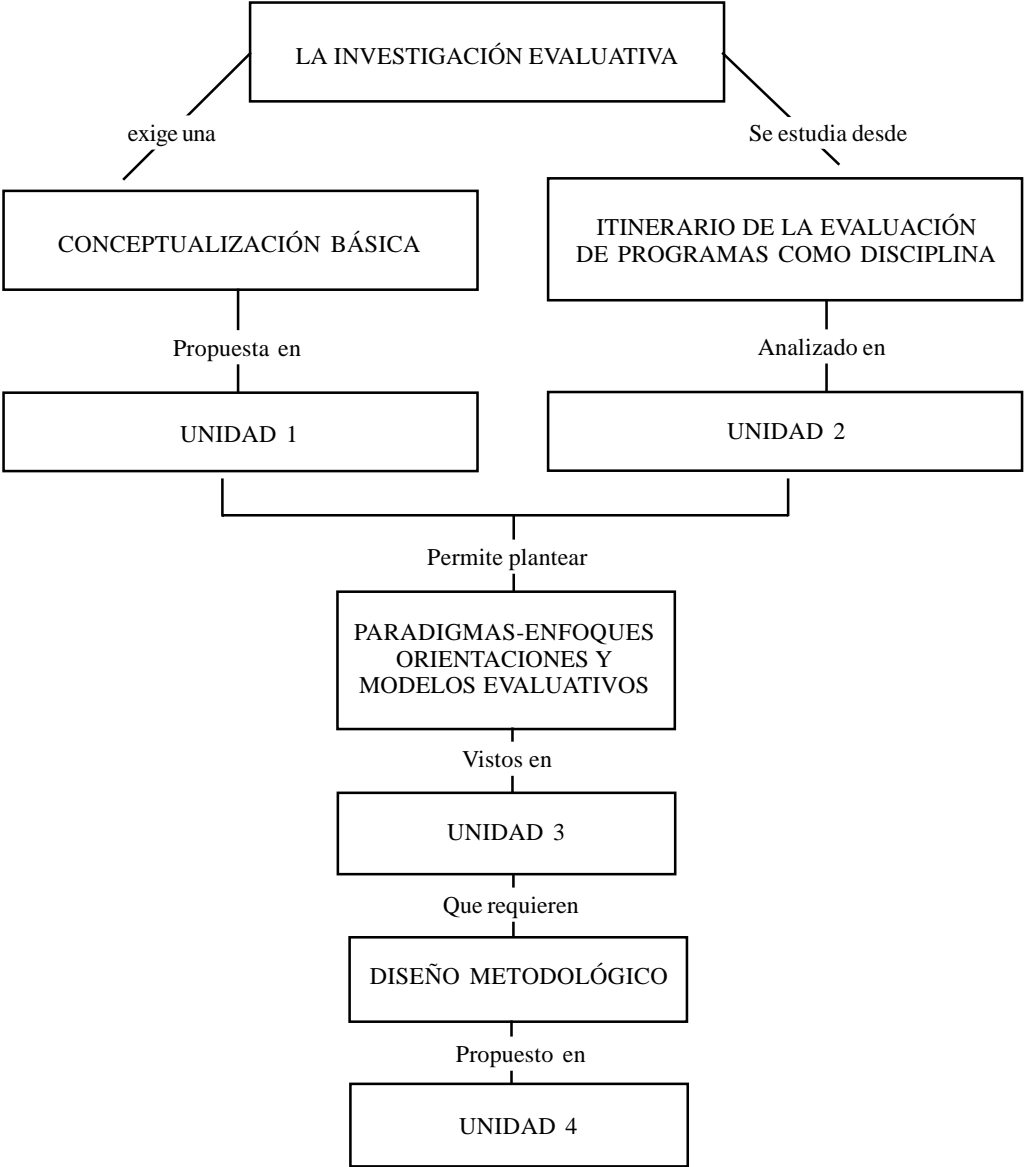
## OBJETIVO GENERAL

Orientar el desarrollo de una experiencia de aprendizaje teórico-práctico, que permita a profesionales de diferentes áreas que cursan la *Especialización teoría, métodos y técnicas de investigación social* lograr las competencias básicas para diseñar, desarrollar e interpretar procesos de investigación evaluativa de programas sociales.

## OBJETIVOS ESPECIFICOS

- Conceptualizar los términos de institución, programa, evaluación e investigación evaluativa y determinar sus alcances.
- Presentar un panorama del recorrido histórico de la investigación evaluativa en sus etapas de formación y consolidación como disciplina y delinear el derrotero de la evaluación de programas en Colombia y el mundo en los últimos 30 años, tanto en el ámbito investigativo como en el académico.
- Caracterizar paradigmas, enfoques, orientaciones y modelos de evaluación y sus posibles campos de aplicación.
- Desarrollar el proceso metodológico de la investigación evaluativa.
- Apoyar a los participantes en el desarrollo de proyectos de investigación evaluativa, dotándoles de herramientas conceptuales, metodológicas y técnicas, propias de este campo.

Mapa conceptual del módulo







**Los conceptos de  
institución-evaluación  
e investigación evaluativa  
y aproximación a una  
tipología de evaluación**

**Santiago Correa Uribe  
Profesor Universidad de Antioquia**

PRIMERA UNIDAD



## INTRODUCCIÓN

Esta primera unidad pretende hacer claridad sobre los conceptos de institución, evaluación e investigación evaluativa de tal manera que permita reconocer la evaluación como una ciencia social de gran actualidad que se vale de los métodos, técnicas e instrumental de la investigación social, pero que, dadas sus peculiares características y sus marcos de acción, presentan una complejidad diferente para lo cual no basta con el dominio de los métodos y técnicas antes mencionadas.

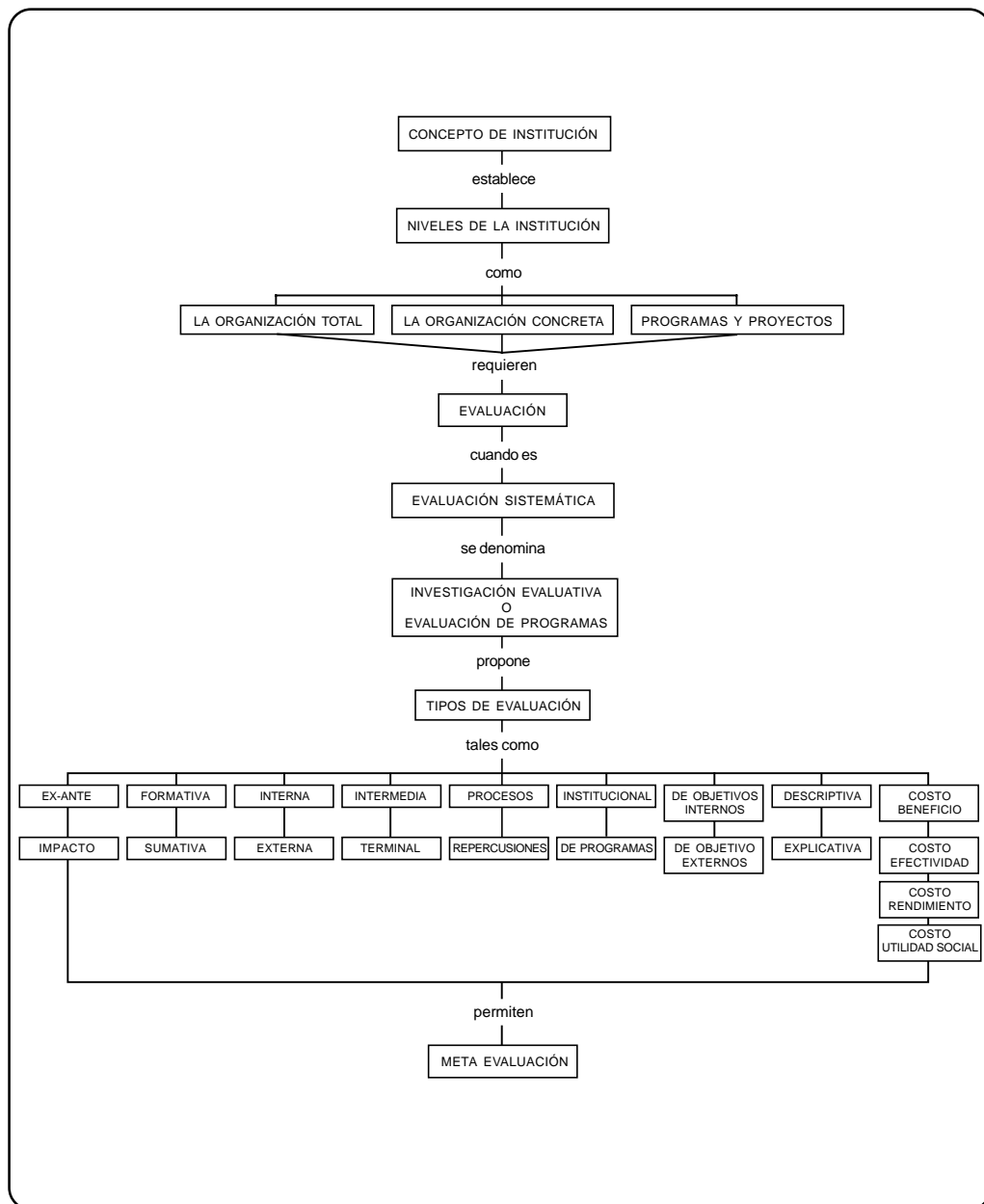
Define, además, esta unidad los propósitos de la evaluación y presenta diferentes tipos de evaluación referenciados en la literatura sobre el tema.

Por último, pretende resaltar los aportes hechos por la investigación evaluativa a la cualificación y mejoramiento de los programas sociales.

## OBJETIVOS DE LA UNIDAD

1. Hacer claridad sobre los conceptos de institución-evaluación e investigación evaluativa.
2. Reconocer los principios, criterios científicos, criterios prácticos-operativos y propósitos de la investigación evaluativa o evaluación sistemática.
3. Caracterizar los diferentes tipos de evaluación, referenciados en la literatura sobre el tema.
4. Resaltar los aportes hechos por la investigación evaluativa a la cualificación y mejoramiento de los programas sociales.

## Mapa conceptual





## I. CONCEPTO DE INSTITUCIÓN

Institución: concepto de la ciencia social, introducido por la sociología positivista y la antropología funcionalista que designa cualquier dispositivo (material, de organización, de intercambio) que, con carácter permanente, regula las relaciones de una sociedad en orden a atender las necesidades humanas (*Diccionario enciclopédico Grijalbo*, 1995).

La definición anterior plantea la complejidad del concepto institución, el cual puede asumirse como un complejo de normas institucionalizadas, como cada una de las organizaciones fundamentales de un estado o nación o como organismo que desempeña una función de interés público. En lo que si existe acuerdo es en afirmar que la Institución es el testimonio sensible de la verdad de que el hombre no está solo en el escenario de la vida social y que una institución sólo tiene fuerza cuando corresponde a la necesidad verdadera y actualmente sentida que hace establecerla.

Los sociólogos asumen la institución como un complejo de normas institucionalizadas. Operacionalizando este concepto, debe aclararse que la referencia a normas institucionalizadas es denotativa. Éstas, para cumplirse, deben ser internalizadas por quienes las practican. Las normas, además de ser aceptadas e internalizadas, son sancionadas por el grupo. El hecho de ser sancionadas posibilita que la norma sea guía de un grupo humano importante.

Es importante aclarar que el concepto de norma es igualmente extenso y complejo. La norma tiene el sentido de un patrón de medida con acepciones tanto en el mundo de lo material, como en el campo de la especulación intelectual y en el mundo de los valores. La norma, cuando es descubierta, promueve la regularidad, la predecibilidad. Sin embargo, no se hace referencia exclusiva a normas escritas o a disposiciones, se refiere además a modos activos de vida, a ambientes de desempeño. Un ejemplo de esta concepción sería el matrimonio, el cual, como institución, se compone de un conjunto de normas, algunas de las cuales no están escritas. Otros ejemplos podrían ser la familia, la justicia, la educación, etc., entendidas en su sentido más general.

Planteando el concepto de institución de una manera tan abstracta haría imposible la evaluación institucional, porque evaluar la justicia o la educación es una acción muy compleja que implica muchos niveles, diversas prácticas, algunas conscientes otras inconscientes, unas formales otras informales, pero esta dificultad se resuelve en el hecho de que toda institución se materializa en una organización concreta y esto es lo que la gente generalmente llama institución. En si, la educación nunca podría ser evaluada, pero ella como acción pauta cristaliza en una serie de organizaciones como la universidad, la escuela, el preescolar y dentro de éstas, a su vez, existen una serie de suborganizaciones que pueden ser evaluadas. Esto lleva a pensar en los alcances del término institución, lo cual incidirá en los alcances de la evaluación institucional.

En concordancia con lo anterior podrán distinguirse varios niveles de institución, a saber:

- Un primer nivel lo constituye la organización total; por ejemplo, la justicia colombiana, la salud, la educación, etc. Al evaluar éstas, se estarían evaluando todas las organizaciones en que se cristalizan cada una de las anteriores. Esta evaluación podría hacerse, pero sería de una gran magnitud y requeriría un equipo de expertos multidisciplinarios muy numeroso. Otro ejemplo sería el matrimonio colombiano como institución, aunque esta investigación sería más antropológica y social que institucional.
- Un segundo nivel es el de las organizaciones concretas dentro de todo el sector institucional, por ejemplo una universidad o una facultad, un hospital o una unidad intermedia de salud, etc. A este nivel, la evaluación es más factible, pues se facilitan tanto el diseño de la investigación, como la ejecución y el análisis. En este segundo nivel, la evaluación es total y deben considerarse aspectos del contexto, insumos, procesos, productos y su impacto social.

**En la organización hay que destacar:**

- Visión, misión y valores corporativos.
- Propósitos, metas y objetivos.
- Estructura organizacional.



- Funciones planeadas para cumplir los objetivos.
  - Recursos: físicos, económicos, humanos, financieros.
  - Funcionamiento de la entidad.
  - Integración cargos funciones.
  - Como se superan los conflictos y desviaciones.
  - Resultado de ella en el medio (aceptabilidad e impacto social de sus realizaciones).
- El tercer nivel hace referencia a los programas y proyectos. La evaluación de programas y proyectos trabaja con sectores más restringidos y en niveles mucho más concretos. El punto de referencia fundamental en esta modalidad de la evaluación son los objetivos de los programas y los proyectos.

En el lenguaje corriente, se entiende por programa una intención o un designio planificado, para cuya ejecución se requiere la aplicación de medios determinados con miras a la obtención de unos resultados o efectos predefinidos. La aplicación de recursos financieros, humanos, la estructura y procesos organizacionales tiene como finalidad obtener unos resultados más o menos concretamente definidos. En este sentido dinámico, el programa no es el mero acto de emplear recursos con fines determinados, sino más bien un proceso cuyo punto de partida consiste en la aplicación de los medios y su desarrollo se caracteriza por la secuencia de ciertos resultados previstos. Si definimos la totalidad de los medios y factores que habrán de transformarse en determinados efectos como “insumo” (*input*) y la totalidad de los efectos que en determinado momento se producen como consecuencia de este proceso de transformación como “producto” (*output*), entonces, puede designarse el programa, en un sentido puramente formal, como la transformación unitariamente organizada de un insumo en un determinado producto previamente definido, dentro de un sistema socioeconómico o en el ámbito de determinadas esferas o subsistemas del mismo.

Puerta plantea como condición de un programa su coherencia interna. “Cuando se habla de coherencia, se entiende que la naturaleza del programa armoniza con la necesidad o problema que se quiere resolver; que los objetivos son viables, las metas medibles, los recursos suficientes y adecuados a las estrategias propuestas y que la metodología utilizada corresponde y garantiza el cumplimiento de esos objetivos” (1997:77).

Es corriente encontrar en la literatura sobre evaluación de programas y proyectos diferenciación entre programas y proyectos económicos y programas y proyectos sociales aunque podría, en un momento dado, considerarse dudosa dicha delimitación ya que economía y sociedad no son subsistemas de un sistema superior del que provengan ni topológicamente separables. Todos los programas y proyectos de desarrollo generan cambios que pueden definirse tanto en términos económicos como sociológicos.

Hay hechos para cuyo análisis debemos recurrir preferentemente a los instrumentos y técnicas de las ciencias económicas, mientras otros resultan más asequibles a los métodos de las ciencias sociales y humanas. En la medida en que tales hechos se identifican como efectos de un proyecto estudiado, los denominaremos efectos económicos en el primer caso y sociales en el segundo. El criterio diferenciador va implícito en la metodología específica aplicada al caso.

Existen también diferencias, al respecto, a los objetivos perseguidos y, por tanto, a los tipos de programas. Una fábrica, al igual que una institución de salud, genera siempre efectos económicos como sociales. Sin embargo, en la primera predominan finalidades y características estructurales que se formulan generalmente en categorías económicas, mientras que en la segunda, los objetivos y características se definen por lo regular mediante el empleo de categorías referentes a beneficios sociales. En virtud de lo anterior denominamos como proyectos económicos a aquellos que persiguen, ante todo, fines económicos, y programas sociales a aquellos cuya finalidad principal consiste en beneficios expresados en términos sociológicos, sociopolíticos, culturales o psicosociales. Esta distinción es puramente operativa, independiente de toda consideración teórico-científica y tiene por objeto únicamente una delimitación pragmática de los factores económicos y sociales en los procesos de desarrollo.

## II. CONCEPTO DE EVALUACIÓN

El término evaluación es una palabra elástica que se emplea para abarcar muchas clases de juicios. Se habla de evaluación de la ejecución de la tarea, de evaluación de la calidad de la educación, de evaluación del clima organizacional, de evaluación del material didáctico, de evaluación del mercado potencial de un producto, de evaluación de sentido de pertenencia, de evaluación de un determinado programa, proyecto o institución, etc. Sin embargo, el término evaluación, a pesar de su creciente popularidad y empleo permanente, es pobremente definido y a menudo impropiaamente utilizado. Para la mayoría, su significado es tomado como garantía y muy pocos ensayos se han hecho en lo concerniente a formular una definición conceptualmente rigurosa o al análisis de los principios de su uso. El resultado es un gran desacuerdo con muchos otros términos como valoración, estimación, apreciación, juzgamiento, muchas veces utilizados como sinónimos de evaluación.

El concepto no se aplica, ni en la literatura científica de la especialidad ni en los concretos estudios de programas y proyectos, en un sentido inequívoco y unánime. Galtung, citado por Musto (1975), observa que la literatura del ramo emplea más de trescientas definiciones diferentes del concepto de evaluación. El examen de éstas revela una inevitable mixtura de definiciones conceptuales y operacionales. Las definiciones conceptuales en su mayoría no ofrecen una lógica formal y presentan una lista de características descriptivas de la evaluación como un proceso cognoscitivo y valorativo. Las definiciones operacionales concentran el propósito de la evaluación en los procedimientos involucrados en la conducción de un estudio evaluativo.

Por evaluación entienden algunos autores el control del éxito o fracaso de programas, proyectos o instituciones, en el sentido de identificar y medir resultados. Así, la evaluación es un esfuerzo por reconocer qué cambios se presentan durante y después de un programa de acción y qué parte de dichos cambios pueden atribuirse al programa. La enciclopedia sobre política del desarrollo plantea: "Bajo control de éxito o evaluación van comprendidas todas las medidas que se tomen para estudiar los efectos e implicaciones de un programa o un proyecto de desarrollo".

Otros autores, en cambio, no ven en la evaluación ningún fin científico en sí, sino solamente una ayuda para tomar decisiones administrativas racionales. Así, la evaluación es un

proceso que capacita al administrador para describir los efectos del programa y a partir de ello hacer los ajustes progresivos en orden a alcanzar las metas más efectivamente. Muchas otras definiciones coinciden en el concepto de finalidad como criterio de exploración de los resultados del proyecto.

La literatura de las ciencias económicas y del análisis de sistemas atribuye a la evaluación una función selectiva. Así, la evaluación es llegar, mediante determinados criterios o modelos de decisión, a poder pronunciarse acerca de la manera óptima de empleo de recursos en el curso de posibilidades alternativas de su aplicación. De acuerdo con este punto de vista, la evaluación no sería, en lo fundamental, otra cosa que la ex-ante valoración de las alternativas que se plantean para llevar a cabo un programa o un proyecto, mientras que el concepto del control del éxito se reservaría para denominar la valoración y el análisis de las fuentes de error de programas y proyectos ya realizados (ex-post evaluación), con lo que serviría la actuación evaluativa así definida como base de decisión para las inversiones.

*El Diccionario de la Real Academia de la Lengua, define evaluar como señalar el valor de una cosa y evaluación como sinónimo de valuación y valoración, pero seguidamente asigna al término evaluar apreciar el valor de las cosas no materiales y reserva el término a valuar y avalúo para apreciar el valor de las cosas materiales.*

Desde esta perspectiva, la evaluación queda designada a la determinación de la calidad de un proceso o actividad humana, sea ella económica, social, política o de cualquier otro orden. Es parte de un proceso y un proceso en sí misma, dirigido conscientemente a determinar el logro de ciertos propósitos fijados con anterioridad.

Restrepo (1974) plantea que la evaluación consiste en la apreciación de un objeto, persona, atributo, programa, institución o sistema, a partir de ciertos criterios y destaca el aspecto valorativo como elemento que objetiviza la escogencia de factores o indicadores relacionados con el propósito de lo que se evalúa. Ahora bien, como toda actividad humana, es social y la mayoría institucional, y cualquier sistema social está influenciado y determinado por variables circundantes, es en consideración de todas estas circunstancias como aparecen las conceptualizaciones teóricas y los modelos de evaluación.

Por diferentes que sean las maneras de concebir el fin y los métodos de la evaluación que estas definiciones implican, todas ellas podemos reducirlas a un común denominador: el de que todas se refieren a la investigación y la apreciación de la eficiencia, eficacia, efectividad y relevancia social de una institución, programa o proyecto, para lo cual se requiere disponer de información sobre los insumos, los procesos y los productos o resultados. Cuanto más válidas, confiables y representativas sean estas informaciones, mayor será la posibilidad de reorganizar los fines y los medios de un programa o proyecto, de tal manera que éste arroje resultados óptimos. Obtener y valorar estas informaciones es lo que denominamos evaluación.

### III. INVESTIGACIÓN Y EVALUACIÓN

“La evaluación es una ciencia social de gran actualidad cuyo ingente potencial no se discute, pero que aún no ha llegado a su rendimiento óptimo. Se vale de los métodos y el instrumental de la investigación social, pero no basta el conocimiento de estos para lograr el desempeño de las funciones que requiere el evaluador, pues su marco de acción es, sin duda, de una complejidad diferente”. (Carol H. Weiss, 1985).

Adentrarnos en el análisis de la relación evaluación e investigación nos obliga a examinar conceptualizaciones alternativas de evaluación. Stufflebeam y Shinkfield (1995) establecen tres grupos amplios de tipos de evaluación: Pseudoevaluaciones, Cuasievaluaciones y verdaderas evaluaciones.

Las pseudoevaluaciones incluyen, en primer, lugar las evaluaciones políticamente orientadas (o encubiertas), que son dirigidas o utilizadas subrepticamente para provocar un punto de vista determinado, positivo o negativo, sobre un objeto, independientemente de la valoración objetiva de su valor o mérito.

“Los charlatanes y los manipuladores utilizan a menudo la evaluación de una manera inadecuada para lograr sus objetivos. Puede que recopilen la información rigurosamente, pero no revelan las verdaderas conclusiones, o lo hacen de

modo selectivo, o incluso las falsifican, engañando tanto a la gente como a sí mismos". (Stufflebean y Shinkfield, 1995: 68).

La intención de este tipo de evaluación es conseguir asistencia para obtener, mantener o incrementar una esfera de influencia, poder o dinero. Los problemas que plantea son los que interesan al cliente, quien quiere que la información sea tan sólida técnicamente como sea posible, pero también que se le garantice la posibilidad de controlar su divulgación para crear una falsa imagen del valor de un objeto, programa o proyecto.

Otro tipo de pseudoevaluaciones son los estudios basados en las relaciones públicas, planificados, realizados y utilizados para satisfacer propósitos que tengan que ver con las relaciones públicas. Buscan información que pueda ser de ayuda para asegurar el apoyo del público. Los métodos que se usan más frecuentemente en este tipo de estudios son las inspecciones, las pruebas y la utilización de "asesores expertos". "La principal característica de la utilización de métodos dudosos por parte de este tipo de evaluadores es su propensión a crear una buena imagen del objeto de evaluación". (Stufflebean y Shinkfield, 1995).

Los estudios cuasievaluativos "aparecen especificados porque existe la evidencia de que a veces pueden ser utilizados para enjuiciar el valor o el mérito de un objeto, mientras que en otros casos su alcance es demasiado estrecho o solo tangencial a las cuestiones del mérito y el valor. Que la información obtenida sea suficiente para apoyar un juicio sobre el valor de algo es, definitivamente, una cuestión secundaria. Los estudios cuasievaluativos poseen otros usos muy legítimos aparte de su relación con la evaluación". (Stufflebean y Shinkfield, 1995: 71).

Pertenecen a este tipo de estudios los basados en objetivos que según el autor mencionado han sido, quizá, la principal fuerza que ha impulsado el movimiento para sistematizar el trabajo evaluativo y los estudios basados en la experimentación los cuales incluyen cuestiones y metodologías a veces relacionadas, pero a veces no, con el juicio de valores. La principal

ventaja de los estudios experimentales en el trabajo evaluativo es que proporcionan métodos sólidos para la determinación de relaciones causales relativamente inequívocas entre el programa y los resultados. Los problemas que plantea, no obstante, consisten en que el método, a menudo, no resulta factible a la hora de señalar los límites de la especificidad y proporciona una información mucho más restringida de la que es necesaria para evaluar programas de educación, salud y bienestar. Además, los estudios experimentales tienden a proporcionar una información final que no resulta útil para el proceso de desarrollo.

Por último aparecen las denominadas verdaderas evaluaciones, basadas en cuestiones de valor y mérito.

El método concreto de la evaluación es la investigación evaluativa, donde las herramientas de la investigación social se ponen al servicio del ideal consistente en hacer más preciso y objetivo el proceso de juzgar. En su forma de investigación, la evaluación establece criterios claros y específicos que garanticen el éxito del proceso, reúne sistemáticamente información, pruebas y testimonios de una muestra representativa de las audiencias que conforman el programa u objeto para evaluar, traduce dicha información a expresiones valorativas y las compara con los criterios inicialmente establecidos y finalmente saca conclusiones.

Suchman (1967) delineó claramente las diferencias existentes entre los objetivos y las facultades investigativas de la evaluación, oponiéndolas a las de la investigación básica o pura, las cuales hacen referencia a una “compleja mezcla de distintos valores, propósitos y recursos”.

“La investigación evaluativa es un tipo especial de investigación aplicada cuya meta, a diferencia de la investigación básica, no es el descubrimiento del conocimiento. Poniendo principalmente el énfasis en la utilidad, la investigación evaluativa debe proporcionar información para la planificación del programa, su realización y su desarrollo. La investigación evaluativa asume también las particulares características de la investigación aplicada, que permite que las predicciones se conviertan en un resultado de la investigación. Las recomendaciones que se hacen en los informes evaluativos son, por otra parte, ejemplos de predicción”. (Suchman, 1967: 119).

La investigación evaluativa contiene una amplia serie de variables sobre las que el evaluador tiene muy poco o ningún control y sus resultados son poco generalizables, pues son aplicables al programa que está siendo evaluado y a sus ramificaciones contextuales.

Existen también semejanzas importantes entre la evaluación (investigación evaluativa) y otras clases de investigación. Como otras investigaciones, la evaluación trata de descubrir, de comprender las relaciones entre las variables y establecer la relación causal. Los evaluadores utilizan toda gama de métodos de investigación para recabar información: entrevistas, cuestionarios, tests de conocimientos y destrezas, inventarios de actitudes, observación, análisis del contenido de documentos, registros y expedientes y exámenes de las evidencias físicas, entre otros. Los evaluadores ingeniosos suelen encontrar las maneras mas convenientes de explorar una amplia gama de efectos. La clase de esquema que se utilice para reunir los datos, dependerá de la clase de información que se necesita para dar respuestas a las preguntas específicas que plantea la evaluación.

### **Principios de la Evaluación**

1. **Propósito:** el propósito de evaluar es mejorar el funcionamiento del sistema. En el sistema educativo, esto significa mejorar su enclave en la sociedad y por tanto garantizar la calidad de los programas en sus diferentes niveles y modalidades.
2. **Función:** la evaluación tiene como patrón de funcionamiento la conciencia social. Si se carece de esta condición, la evaluación será restringida, unilateral y no procurará de una manera consciente el mejoramiento de la institución o el programa.
3. **Uso:** los resultados de la evaluación deben darse a conocer a todos los actores y ponerse a disposición del público.
4. **Fin:** la evaluación debe orientarse y organizarse teniendo en cuenta que el punto final de ésta es la toma de decisiones.



## **Criterios científicos de la evaluación**

- 1. Validez:** la evaluación debe reflejar una correspondencia inequívoca entre la información y el aspecto de la realidad que se evalúa, lo que implica que en el diseño de la evaluación:
  - a.** Se elimine o controle la interferencia de variables extrañas (validez interna).
  - b.** Se garantice que los instrumentos de medición midan lo que se quiere medir (todo lo que se quiere medir y sólo lo que se quiere medir), es decir, hayan congruencias entre los ítems que componen el instrumento y el objeto evaluado que se mide (validez de contenido de los instrumentos).
  - c.** Las muestras de informantes, al ser representativas de la población, seleccionadas al azar y controlando factores ambientales ajenos, permitan cierto grado de generalización de la información (validez externa).
- 2. Confiabilidad:** la evaluación debe ser consistente. Esto es, si los instrumentos se utilizan más de una vez para evaluar la misma calidad, los resultados obtenidos deben ser muy parecidos.
- 3. Objetividad:** la evaluación debe minimizar las apreciaciones personales en el momento de emitir juicios valorativos, empleando instrumentos que recojan hechos verificables, de modo que, si distintos evaluadores aplican el mismo instrumento, obtengan resultados similares.

## **Criterios práctico-operativos de la evaluación**

- 1. Integral:** de modo que comprenda todos los elementos e involucre todos los factores que intervienen en el programa o acción evaluada.
- 2. Participativa:** esto es que vincule activamente los diferentes actores que intervienen en el programa o acción que se va a evaluar.

3. **Permanente:** ha de realizarse en forma continua, asociada al desarrollo del programa.
4. **Acumulativa:** al considerar los resultados de evaluaciones previas, para determinar sus efectos sobre las sucesivas.
5. **Autocorrectiva:** esto es, contempla mecanismos de ajuste de estrategias e instrumentos para futuras etapas del proceso de evaluación permanente.
6. **Útil:** en la medida en que contemple los aspectos más importantes en relación con los objetos evaluados (relevante) y que sea oportuna, de manera que proporcione a los diferentes niveles decisorios la información necesaria en el momento requerido.
7. **Eficiente:** si obtiene un máximo rendimiento de los recursos disponibles (humanos, físicos, financieros).

## PROPÓSITOS DE LA EVALUACIÓN

Suchman (1967) asume como propósitos de la evaluación los expuestos por Bigman (1961).

1. Describir si los objetivos han sido alcanzados y de qué manera.
2. Determinar las razones de cada uno de los éxitos y fracasos.
3. Descubrir los principios que subyacen en un programa que ha tenido éxito.
4. Dirigir el curso de los experimentos mediante técnicas que aumenten su efectividad.
5. Sentar las bases de una futura investigación sobre las razones del relativo éxito de técnicas alternativas.
6. Redefinir los medios que hay que utilizar para alcanzar los objetivos, así como, incluso, las submetas, a la luz de los descubrimientos de la investigación.

Estos propósitos, en opinión de Suchman, plantean una relación intrínseca entre la planificación del programa, su desarrollo y la evaluación. Los medios para alcanzar dichos propósitos (la investigación evaluativa) deben proporcionar la información básica para planificar y, si es necesario, replanificar los programas. Así como la investigación tradicional suele conducir hacia una comprensión más amplia de los procesos básicos, la investigación evaluativa debe “conducir a una mayor comprensión de los procesos aplicados o administrativos”.

Weiss (1985) plantea como propósito de la investigación evaluativa “medir los efectos de un programa por comparación con las metas que se propuso alcanzar, a fin de contribuir a la toma de decisiones subsiguientes acerca del programa y para mejorar la programación futura” (p. 16). Aparecen aquí cuatro rasgos primordiales: la expresión “medir” hace referencia a la metodología de la evaluación que se emplea. La expresión “los efectos” hace hincapié en los resultados del programa. La “comparación” de los efectos con las metas subraya el uso de criterios explícitos para juzgar hasta qué punto el programa está marchando bien. La “contribución” a la toma de decisiones subsiguientes y al mejoramiento de la programación futura denota la finalidad social de la evaluación.

La evaluación de programas sociales directamente relacionada con la tarea de decisiones deberán, según, Weiss ser orientadas hacia uno o más de estos propósitos:

1. Continuar o discontinuar el programa.
2. Tratar de mejorar su funcionamiento.
3. Agregar o suprimir alguna de sus partes.
4. Establecer programas similares en otros lugares.
5. Redistribuir los recursos según los resultados de programas alternativos.
6. Aceptar o rechazar algún enfoque o teoría en la cual esté basado el programa.

Briones (1991) define como propósito principal de la evaluación: “proporcionar criterios para la toma de decisiones en relación con la estructura, el funcionamiento y el desarrollo del programa”, p. 17. Y a manera de ilustración presenta una lista de propósitos que pueden perseguirse en una investigación evaluativa:

1. Mejorar la efectividad cuantitativa del programa, es decir, el logro de sus objetivos.
2. Mejorar la calidad de los resultados, o sea, la efectividad cualitativa del programa.
3. Aumentar la eficiencia interna del programa, vale decir, la relación entre insumos y resultados.
4. Mejorar la eficiencia externa del programa o adecuación de resultados con objetivos externos.
5. Someter a verificación la teoría o hipótesis que relaciona los componentes del programa con los resultados buscados.
6. Redefinir la población objeto: reducirla, homogenizarla, etc.
7. Modificar la estructura, el funcionamiento y la metodología del programa.
8. Tratar de cumplir, si es pertinente, las expectativas de la población objeto del programa (sobre su funcionamiento, recursos, objetivos, etc.).
9. Controlar la aparición de posibles prácticas o resultados no deseados.
10. Tratar de homogenizar los efectos, consecuencias y repercusiones del programa en diversos subgrupos de la población de referencia del programa.

Plantea el autor en mención que, cualquiera que sea el caso, algunos de estos propósitos surgen de algún indicio o sospecha de que algo anda mal en el programa. Tales indicaciones, junto con otras fuentes de información, permiten definir los objetivos de la evaluación y, en definitiva, ayudar a preparar el diseño metodológico que se empleará.

A manera de síntesis se presenta el planteamiento de Crombach y sus colaboradores (1980) cuando puntualizan:

“El propósito de una evaluación es influir en el pensamiento social y la acción, durante la investigación o en los años subsecuentes. También es razonable esperar influencias a largo plazo” (p. 15).

Aparece en esta definición un nuevo propósito la evaluación de impacto, realizada para determinar los efectos residuales del proyecto.

#### IV. TIPOS DE EVALUACIÓN

Existe una amplia clasificación de los tipos de evaluación, diferenciados de acuerdo con múltiples criterios: el momento en que se realiza la evaluación, la naturaleza del programa que se quiere evaluar, la procedencia de los evaluadores, las categorías que se van a emplear y muchos otros que dificultan una tipología universalmente aceptada.

A continuación se definen los tipos de investigación mas conocidos, con el ánimo de que el investigador disponga de varias opciones metodológicas en el momento de emprender una evaluación.

#### EVALUACIÓN FORMATIVA Y EVALUACIÓN SUMATIVA

Es útil distinguir como lo hace Scriven (1967) entre una evaluación formativa y evaluación sumativa o acumulativa.

**Evaluación formativa:** llamada también evaluación intermedia, de proceso, configurativa, continua o concurrente, se caracteriza porque las mediciones de los efectos se hacen en diversos momentos del desarrollo del programa (el diagnóstico, la formulación, la implementación o la ejecución), de tal modo que los resultados o experiencias logradas hasta entonces son utilizados en modificaciones o reorientaciones del mismo.

En general, la evaluación formativa se aplica en la fase de ejecución de un programa o proyecto y es especialmente recomendable cuanto éste tiene etapas claramente definidas, con metas intermedias y a largo plazo, o en proyectos que admiten, a manera de retroalimentación, una modificación en la marcha (mejoramiento continuo). La evaluación formativa es parte integral de cualquier programa y debe estar presente en todas sus etapas, no solamente como siempre se ha pensado, para medir los efectos terminales del mismo. La evaluación formativa en general tiene un componente de participación de los diferentes actores del programa que se va a evaluar. Es entendida, desde los modelos

participativos y naturalistas o apreciativos, como un proceso educativo en el que aprenden tanto los evaluadores (técnicos) como los evaluados (actores sociales).

El Ministerio de Salud de nuestro país caracteriza la evaluación formativa así (1978:77):

1. Es un componente esencial de la planificación; mide y controla los desempeños y alcances logrados con la programación, dirección y ejecución de los servicios.
2. Es continua e integral.
3. Permite estudiar la eficiencia y eficacia de un servicio.
4. Implica una autoevaluación.
5. Cumple una función retroalimentadora dentro del desarrollo de un programa.
6. Brinda una información general del universo evaluado.
7. Ofrece registros sistemáticos durante su ejecución.
8. Minimiza el desfase entre identificación de problemas y toma de decisiones.
9. No representa costos adicionales mayores.

**Evaluación sumativa:** llamada también acumulativa, terminal, de punto, de resultado, final de producto o ex-post, se caracteriza porque se efectúa al término del programa y da cuenta, por lo tanto, sólo de los resultados finales.

Es, probablemente, el tipo de evaluación que más aplican las instituciones encargadas de prestar servicios sociales, “inclusive se asimila a lo que en la administración moderna se denomina informe o balance social” (Puerta, 1996:57).

En Colombia, según el Ministerio de Salud (1978), la evaluación sumativa o de punto:

1. Es un componente del proceso de evaluación, por tanto, es complemento de ese proceso.

2. Es circunstancial.
3. Se puede aplicar al estudio de factores de eficiencia.
4. Debe ser realizada por un personal diferente al que presta los servicios incluidos dentro de un programa.
5. Es complemento del proceso de retroalimentación de un programa.
6. Brinda una información específica acerca del universo de desarrollo de un programa.
7. Utiliza formularios no manejados como registros sistemáticos.
8. No ofrece identificación oportuna de problemas.
9. Representa un costo adicional relativamente alto.

La evaluación sumativa o ex-post debe evaluar:

- Los productos, o sea, los bienes y servicios producidos y prestados por el programa o proyecto.
- Los efectos, o sea, los resultados de la utilización de los productos.
- El impacto, o sea, los cambios que permanecen luego de terminado el programa o proyecto.

**Evaluación externa, evaluación interna y evaluación mixta:** esta clasificación hace referencia al tipo de personal encargado de llevar a efecto la evaluación.

**Evaluación externa:** es la evaluación realizada por agentes externos a la institución o programa. Generalmente, son profesionales con reconocido prestigio en el campo de la evaluación, o funcionarios de entidades especializadas tales como universidades, centros de asesoría y auditorías especiales.

Se dice que una evaluación efectuada por investigadores externos asegura mayor objetividad en cuanto a la apreciación del funcionamiento y resultados del programa.

Por otro lado es más probable que el evaluador externo se haya sometido a un entrenamiento especial en la evaluación de programas y que también tenga experiencia en otras situaciones que pueda aplicar en esta oportunidad. Los evaluadores externos adquieren en el ejercicio de su profesión una gran habilidad para determinar los asuntos claves que deben ser evaluados, los indicadores más adecuados y las mejores estrategias y diseños para la recolección, el procesamiento, el análisis y la interpretación de la información.

Pero también se hacen algunos cuestionamientos a la evaluación externa tales como: el evaluador externo no conoce el programa, por tanto invierte gran parte del contrato tratando de comprender sus aspectos formales, pero el ambiente informal y las intimidades propias de cada programa y cada institución nunca se comprenderán. No es fácil compatibilizar los puntos de vista del evaluador externo con los que tiene en los temas de evaluación el personal del programa. El evaluador externo hace que los costos se incrementen y no propicia una cultura de la evaluación y del empoderamiento y en algunos casos tiende a crear resistencias frente a los residentes del programa, quienes, a su vez, no consideran que tales evaluadores tengan autoridad suficiente para valorar lo que ellos están realizando.

**Evaluación interna:** es la evaluación realizada por el personal de la institución o del programa, generalmente grupos interdisciplinarios en los que están representados todos los estamentos y audiencias del programa.

Los encargados de la evaluación interna son responsables de analizar los problemas y ofrecer recomendaciones, pero también de corregir las dificultades e implementar las soluciones; de modo que la evaluación interna es, ante todo, una actividad administrativa (Love, 1991:24).

Los defensores de este tipo de evaluación afirman que los funcionarios tienen un conocimiento más profundo de la estructura y funcionamiento del programa y, por consiguiente, están en mejores condiciones de evaluarlo. Por el contrario, sus opositores afirman que el personal vinculado directamente al programa no podrá tomar distancia crítica frente al mismo, pues muchas veces el compromiso de estos funcionarios resulta más afectivo que racional o profesional, por lo cual no pueden ver las fallas del programa y, además, en ocasiones la defensa del programa tal cual se realiza les representa una estabilidad laboral.



Hamilton (1981) señala como ventajas de la evaluación interna las siguientes:

- Puede contemplarse como parte fundamental en la estructuración de una institución o programa.
- Permite que se acepten más fácilmente los resultados y que se tomen las medidas respectivas.
- La evaluación interna ofrece a los evaluadores mayores posibilidades de conocer la naturaleza del programa.

**Evaluación mixta:** es la evaluación realizada por un grupo conformado por evaluadores externos y personal del programa (evaluadores internos). Se trata de una alternativa que minimiza los inconvenientes o factores excluyentes de las evaluaciones interna y externa, para, en cambio, potenciar las ventajas de ambas. Un caso especial de esta modalidad lo constituye la investigación-acción, donde la evaluación del programa es una tarea en la que todos sus participantes se comprometen.

## **OTRAS TIPOLOGÍAS DE EVALUACIÓN**

En la literatura sobre evaluación de instituciones se encuentran, además de las clasificaciones anteriores, muchas otras que, de alguna manera, podrían encuadrarse dentro de éstas. Se definirán a continuación algunas de ellas.

### **Evaluación intermedia-evaluación terminal**

Briones (1991) identifica la evaluación intermedia como aquella que busca información durante el desarrollo del programa con el fin de detectar problemas de funcionamiento y buscarles soluciones adecuadas y oportunas. La evaluación terminal, denominada también evaluación ex-post, es el estudio que se realiza después de que el programa ha terminado, con la finalidad de establecer si se obtuvieron los resultados esperados y los factores que actuaron en una u otra dirección.

### **Evaluación de procesos-evaluación de repercusiones**

El mismo autor define la evaluación de procesos como aquella que se centra en el funcionamiento y otros aspectos dinámicos del programa, mientras que la evaluación de repercusiones se refiere a la determinación del logro de objetivos de un programa o a la evaluación de efectos externos.

### **Evaluación institucional-evaluación de programas**

La evaluación institucional se refiere a la evaluación de una institución en su nivel de organización, de tal modo que uno de sus focos principales de atención está constituido por las funciones que ella debe cumplir. Entre tales funciones se pueden definir objetivos o metas, pero como momentos puntuales dentro de esas actividades permanentes de la institución. (Briones, 1991).

La evaluación de programas se refiere a la evaluación de actividades más o menos coordinadas en torno a objetivos de duración variable. Estos últimos son estados o situaciones deseables y que pueden lograrse mediante la estrategia que el programa utiliza y respecto de los cuales se evalúan los resultados alcanzados o no. (Briones, 1991).

## **Evaluación de objetivos internos-evaluación de objetivos externos**

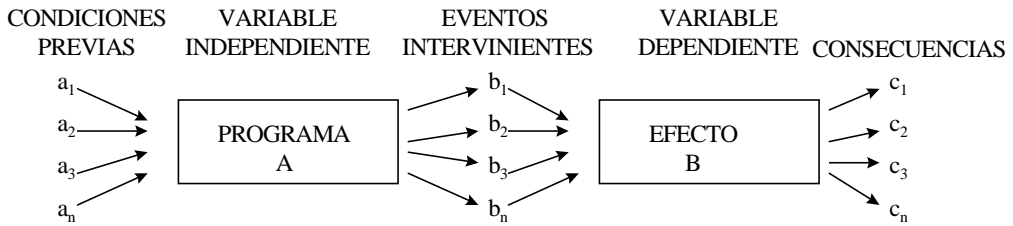
La evaluación de objetivos internos se refiere al proceso destinado a verificar el logro de los objetivos que han de alcanzarse en la población en la cual actúa directamente el programa (objetivos internos). La evaluación de objetivos externos hace referencia a la evaluación de situaciones o estados deseables de alcanzar fuera del ámbito del programa cuando éste ya ha finalizado. Por lo general, estos objetivos son de naturaleza económica, social y cultural, y se espera que se cumplan en la comunidad o en la sociedad global.

## **EVALUACIÓN DESCRIPTIVA Y EVALUACIÓN EXPLICATIVA**

**Evaluación descriptiva:** describe diferentes situaciones del programa (características, factores que intervienen y relaciones de asociación entre éstas). Se apoya en técnicas de análisis de información descriptiva tanto cuantitativas como cualitativas.

**Evaluación explicativa:** trata de establecer la relación causal entre los diferentes factores que intervienen en el programa y entre éste y el contexto. “Estas explicaciones, a su vez, están constituidas por generalizaciones incorporadas al cuerpo de las disciplinas en las cuales se apoya el programa (psicología, sociología, antropología, etc.), o bien se formulan hipótesis explicativas *ad hoc*, que nuevas evaluaciones podrán confirmar o desconfirmar” (Briones, 1991:28).

Al hacer la distinción entre evaluación descriptiva y evaluación explicativa, Suchman (1967:84) señala que en este segundo tipo es necesario examinar su fundamentación racional, la que esencialmente consiste en concebir una secuencia causal en la cual “nuestro programa viene a ser sólo una de muchas acciones o eventos posibles que pueden producir el efecto deseado. Además, nuestro programa no sólo influirá para que se produzca el efecto deseado, si no que también tendrá otros efectos” (ver diagrama).



## EVALUACIÓN COSTO-BENEFICIO, COSTO-EFECTIVIDAD Y COSTO-RENDIMIENTO

Puerta (199:63-64) define estos tres tipos de investigación así:

**Evaluación costo-beneficio:** se aplica para estimar el valor de bienes intangibles y para disponer de un elemento más de juicio, cuando se deben tomar decisiones que implican una asignación de recursos. (Ander Egg, 1984:73). Este tipo de evaluación puede aplicarse antes de que se de inicio al programa o una vez concluido el mismo.

En los proyectos sociales, la medición del beneficio se expresa mediante la utilización de uno o varios indicadores de impacto sobre la población objetivo. En cambio, en la evaluación financiera y en la evaluación económica, los beneficios se expresan en unidades económicas.

**Evaluación costo-efectividad:** esta evaluación contrasta los resultados o beneficios no monetarios con los costos monetarios (Abello, 1987). Determina la eficiencia económica de un proyecto al relacionar sus costos y productos con los efectos alcanzados.

**Evaluación costo-rendimiento:** de acuerdo con Musto (1975:99), la finalidad que persigue esta evaluación es la de hallar y determinar la diferencia entre el valor actual de los costos públicos y el valor actual de los rendimientos o utilidades de un proyecto, teniendo en consideración también aquellos costos y rendimientos que inciden sobre sujetos distintos de los directamente afectados por el proyecto.

El autor incluye en este tipo de evaluación el calculo de los rendimientos de lo que se considera el impacto del programa; que no son los resultados del programa, sino los efectos indirectos en cuanto a la población beneficiaria.

**Evaluación costo-utilidad social:** determina el costo de alguna actividad que afecta a la sociedad en conjunto y que no necesariamente son iguales a los costos que paga la

persona o empresa que realiza dicha actividad. Por consiguiente, los costos sociales están formados por los costos de oportunidad de los recursos usados en la actividad, junto con el valor de cualquier pérdida de bienestar o incremento de costos que esa actividad provoca para algún individuo o grupo social.

**EVALUACIÓN EX-ANTE:** es la evaluación que se aplica a la formulación del proyecto (evaluación social y económica del proyecto). Está orientada al análisis de la factibilidad del proyecto, coherencia interna e ingeniería del proyecto y eficiencia del mismo.

Los parámetros utilizados son (Pichardo, 1989:103-114):

- La viabilidad o posibilidad del proyecto en sus términos técnico, financiero, jurídico, sociocultural, institucional y político.
- La transcendencia social, la eficiencia del proyecto o su capacidad para producir impacto en los destinatarios, las organizaciones y el medio social.

**EVALUACIÓN DE IMPACTO:** esta evaluación se refiere al “efecto agregado del programa sobre la comunidad como un todo” (Quintero, 1995:6). Es decir, es aquella evaluación realizada para determinar, bien sea, los efectos residuales del proyecto en su totalidad los efectos de una sola de las actividades asociadas al mismo (García, 1992:6).

**META-EVALUACIÓN:** hace referencia a la evaluación de la evaluación. El término es introducido por Scriven en 1968 cuando escribe un artículo sobre la evaluación de los instrumentos de evaluación. Plantea que los evaluadores tienen la obligación profesional de que las evaluaciones propuestas o concluidas estén sujetas a una evaluación competente.

Las sugerencias metodológicas de Scriven, para la realización de metaevaluaciones, incluyen la utilización de un listado de control de indicadores de la evaluación.

Los resultados de la metaevaluación deben ser utilizados formativamente, pero también pueden incorporarse al informe o transmitidos al cliente y otras audiencias adecuadas.

## AUTO EVALUACIÓN 1

Intente dar respuesta a los interrogantes que a continuación se plantean recurriendo a la revisión de la unidad, a las fuentes bibliográficas referenciadas, incluida la lectura complementaria 1 y a la discusión con sus colegas.

1. ¿Qué diferencias encuentra entre la investigación social y la investigación evaluativa?
2. ¿Existe un método único para la investigación evaluativa? En caso afirmativo, ¿cuál es éste y en caso negativo es esto ventajoso o desventajoso?
3. ¿Hasta qué punto cree usted que la sociedad debe invertir en preparar y emplear especialistas en investigación evaluativa en lugar de simples evaluadores?
4. ¿Dónde, cómo y hasta qué punto debe proporcionársele preparación a ambos tipos de evaluadores?
5. ¿Se justifica la inversión económica en los procesos de evaluación institucional a sabiendas de que en última instancia la toma de decisiones es eminentemente política?

## ACTIVIDAD DE EVALUACIÓN

### UN CASO PARA EL ANÁLISIS

A continuación se presenta un caso en cuyo análisis usted aplicará los contenidos desarrollados en la unidad. Elabore un ensayo de máximo 3 páginas en el cual fije su posición frente a la propuesta del jefe.

El secretario de desarrollo comunitario de su municipio lo contrata para discutir los planes de evaluación de los programas a cargo de esta dependencia.

El funcionario le manifiesta que se ha definido como estrategia la autoevaluación, pues considera que todos los miembros del personal de la secretaría son responsables de evaluar su propio trabajo y advierte, además, que no se han empleado los servicios de ninguna agencia evaluadora ni de especialistas en evaluación. En lugar de ello, ha recordado a todos sus empleados su responsabilidad de evaluar su propio trabajo y les ha enseñado un método evaluativo con el fin de garantizar el éxito en el proceso. El secretario espera que este tipo de actividad evaluativa sea de utilidad para convencer al Concejo municipal para incrementar el presupuesto de la dependencia.

## LECTURA COMPLEMENTARIA:

“Introducción a la evaluación”. En: Stufflebeam, Daniel L. y Shinkield, Anthony J.  
*Evaluación sistemática. Guía teórica y práctica*. Barcelona: Ediciones Paidós, 1995.



## BIBLIOGRAFÍA

Ander-Egg, Ezequiel. *Técnicas de evaluación social*. Buenos Aires: Humanistas, 1984.

Abello, R. y Madarriaga, C. *Estrategias de evaluación de programas sociales*. Barranquilla: Uninorte, 1987.

Bigman (1961), citado por Suchman.

Briones, Guillermo. *Evaluación de Programas Sociales*. Editorial Trillas. México 1991.

Crombach, L. J. *Toward Reform of Program Evaluation*. San Francisco: Jossey-Bass, 1980.

García, José. *Evaluación de programas de planificación familiar*. Bogotá; Profamilia, 1992.

Ministerio de Salud. *Manual de evaluación de programas de salud*. Bogotá, 1978.

Musto, Stefan A. *Análisis de eficiencia-metodología de la evaluación de proyectos sociales de desarrollo*. Madrid: Editorial Tecnos, 1975.

Pichardo, Arlette. *Evaluación del impacto social*. San José, Costa Rica: Editorial Universidad de Costa Rica, 1989.

Puerta, Antonio. *Evaluación y sistematización de proyectos sociales*. Medellín: Imprenta Departamental de Antioquia, 1997.

Quintero, Víctor Manuel. *Evaluación de proyectos sociales: construcción de indicadores*. Cali: Fundación FES, 1995.

Restrepo, Bernardo. *Conferencias sobre evaluación*. Documento mimeografiado. Medellín: Facultad de Educación, Universidad de Antioquia, 1974.

Scriven, M. S. "The Methodology of Evaluation". En: *Perspectives of Curriculum Evaluation*. Chicago: Rand Mc Nally, 1967.

Stufflebeam, D. y Shinkfield, A. *Evaluación sistemática, guía teórica y práctica*. Barcelona: Ediciones Paidós, 1995.

Suchman, E. A. *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. New York: Russell Sage Foundation.

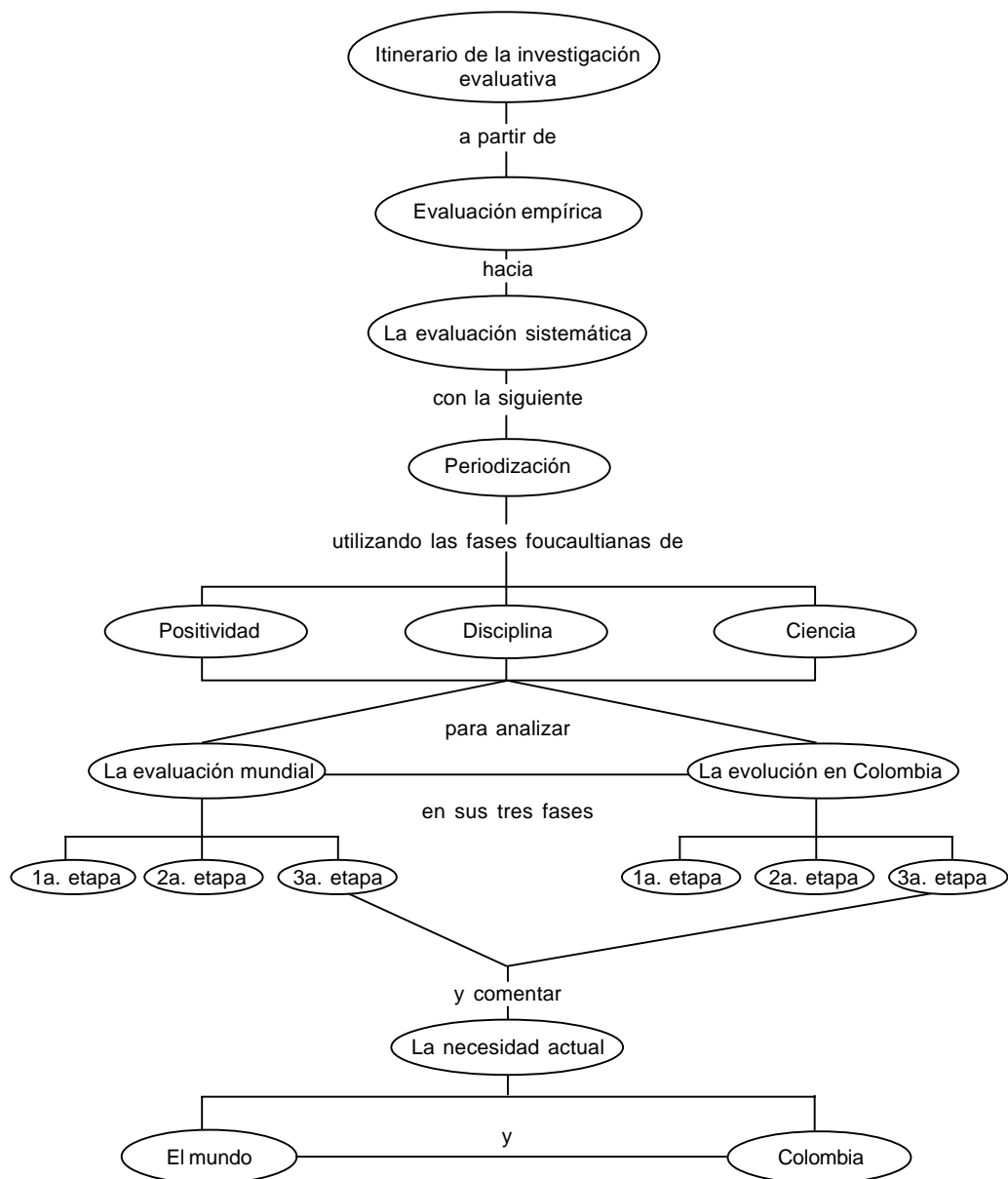
Weiss, Carol. *Investigación evaluativa*. México: Editorial Trillas, 1985.

# **Itinerario de la evaluación de programas como disciplina**

**Bernardo Restrepo Gómez**  
**Profesor Universidad de Antioquia**

SEGUNDA UNIDAD

## Mapa conceptual



## INTRODUCCIÓN

Esta unidad sobre la historia de la evaluación como disciplina, que comprende la evolución de la investigación evaluativa, pretende dar al evaluador de instituciones y programas sociales una visión del itinerario seguido por este campo de acción investigativa a partir de la década de 1930 cuando comenzó la sistematización de esta actividad, que, a decir verdad, ha existido desde tiempos inmemorables. El valor y mérito de una política, de una institución, de un programa concreto, en efecto, han sido observados y medidos siempre por los estados con el propósito de tomar decisiones sobre la continuidad de unos y otros. Sólo que se hacía con base en la experiencia de quien apreciaba la situación, sin instrumentos de ayuda capaces de refinar los resultados. El nacimiento de una disciplina académica, encargada de producir conocimientos y tecnología sobre la acción evaluativa, suele ubicarse en la década de 1930, concretamente en los trabajos de Ralph Tyler sobre evaluación "como una área que debía ir más allá del examen" y de lo académico para incluir lo contextual de los programas educativos y de su rendimiento.

Desde el acuñamiento del término "evaluación", hecho debido a Tyler (1934), la disciplina se ha ido configurando y en torno a ella se ha ido consolidando una comunidad académica encargada de cultivar sus temáticas, corrientes u orientaciones, metodología y aplicaciones. Esta unidad se ha pensado como vehículo para pasear a los lectores por los escenarios históricos que la evaluación de programas ha recorrido desde 1934 hasta finales del siglo XX.

La unidad consta de tres partes: el itinerario de la evaluación en el mundo, el itinerario de la evaluación en Colombia y la necesidad y vigencia de la evaluación hoy en los distintos campos del sector social.

### 3. Objetivos

#### 3.1 Objetivo general

La unidad busca dar al estudiante un panorama del recorrido histórico de la investigación evaluativa en sus etapas de formación y consolidación como disciplina. Adicionalmente, pretende escudriñar eventos evaluativos que han impulsado el desarrollo de la evaluación institucional y de programas en Colombia en los últimos 30 años.

#### 3.2 Objetivos específicos

La unidad se centra en los objetivos de:

- 3.2.1. Identificar etapas de desarrollo de la evaluación de programas desde el modelo tyleriano de la década de 1930 hasta los modelos participativos, particularmente el de empoderamiento (*Empowerment*) de Fetterman de la década de los 90 y su aplicación en diversos campos del sector social.
- 3.2.2. Señalar hechos y características que permiten periodizar la historia de la evaluación: orientaciones o enfoques, métodos, modelos, organización de una comunidad académica y expresiones colectivas de ésta.
- 3.2.3. Delinear el derrotero de la evaluación de programas en Colombia en los últimos 30 años, tanto en el ámbito investigativo como en el académico.
- 3.2.4. Discutir la vigencia de la evaluación institucional y de programas en la actualidad en los sectores público y privado.

### 4. Desarrollo de la unidad

Evaluar es someter algo a valoración o tratar de juzgar el valor y mérito de algo, como la ha definido Scriven (Stufflebeam, 1997). Esto puede hacerse empíricamente acudiendo al sentido común, o puede hacerse sistemáticamente con enfoques y métodos compartidos por comunidades académicas y científicas. Los Estados en sus distintos entes territoriales, las instituciones internacionales y multilaterales y, en general, quien patrocina políticas,

programas o proyectos de cualquiera índole exigen desde hace unas cuatro décadas que la evaluación de políticas y programas sociales, científicos y productivos se adelante a través del segundo tipo de evaluación.

Para entender el surgimiento y evolución de la evaluación institucional y de programas sociales como una disciplina con objeto y método de trabajo, es necesario recorrer el derrotero de la formalización de la evaluación. Para hacerlo acudiremos al planteamiento Foucaultiano sobre la historia de las ciencias, esto es, a las etapas de positividad, disciplina y ciencia. Este recorrido lo haremos primero en el ámbito universal y luego en el ámbito colombiano.

#### **4.1 Itinerario de la evaluación en el mundo**

Hoy en día, la evaluación de programas tiene gran desarrollo en los EE.UU., Europa (Inglaterra, Francia, Italia, Hungría, países escandinavos...), Australia, Israel, África y América Latina. Su origen se ubica en los EE.UU. e Inglaterra (Worthen, Sanders y Fitzpatrick, 1997).

##### **4.1.1 *Los inicios de la evaluación***

Apoyándonos, pues, en el esquema foucaultiano sobre la formación de las ciencias (Foucault, 1972), podríamos decir que la evaluación de programas sociales, como toda disciplina, ha seguido un curso progresivo de lo simple a lo complejo, de propósitos particulares a una visión más integral y de un enfoque único a una variedad creciente de enfoques u orientaciones y de múltiples modelos alternativos dentro de cada orientación. Los campos que desde más temprano se han preocupado por evaluar su accionar son la educación y la salud. En educación, que por su razón de ser siempre ha incluido la actividad evaluativa o de comprobación como componente esencial, la evaluación de programas comenzó a figurar como una “positividad”, o sea, un campo con cierto objeto más allá de la evaluación del aprendizaje académico y abarcando otros factores, a partir de los trabajos de Ralph Tyler quien desde 1934 y cada vez más intensamente planteó la necesidad de transformar el sistema de evaluación incluyendo aspectos filosóficos, culturales y escolares al constatar logros de objetivos. Tyler acuñó el término “evaluación” para expresar una apreciación del rendimiento estudiantil más holística, más allá de la era del “examen” que incluyese la consideración de las circunstancias explicativas de los

resultados de dicho rendimiento. Con sus trabajos iniciales al respecto se inició una individualización de la evaluación educativa en cuanto investigación evaluativa. Esta etapa se extendió hasta finales de la década de 1950.

Hasta finales de la década de 1950, la orientación tyleriana de una evaluación basada en comprobación del logro de objetivos campeó en el ámbito no solo de la educación, sino de otros programas sociales y los métodos utilizados para detectar productos fueron los métodos cuantitativos. La década de 1960 trajo nuevos desarrollos en el campo evaluativo impulsados principalmente por hechos teóricos y políticos. En el campo teórico hay que mencionar la influencia creciente de la psicología cognoscitivista, en particular los trabajos de Piaget, Bruner y Ausubel, al insistir en el desarrollo y evaluación de las funciones intelectuales primarias, el aprendizaje por descubrimiento y el aprendizaje significativo, énfasis que demandan una evaluación más orientada a procesos que a resultados comparados con objetivos. Este cambio en la teoría del aprendizaje puntualiza, en el campo de la evaluación institucional y de programas, un viraje de la valoración de productos hacia la comprensión y mejoramiento de procesos, concepción que se evidenciará en el florecimiento de modelos cualitativos a partir de la década de 1960. Los métodos cualitativos entran a compartir con los cuantitativos la implementación de la evaluación, como se verá luego.

Desde el punto de vista político, el sentimiento de culpa interiorizado en Estados Unidos sobre la baja calidad de su educación, puesta de presente ante el triunfo espacial soviético que lanzó el primer Sputnik en 1959, intensificó la conciencia de la necesidad de evaluar todo proyecto educativo antes de renovar fondos federales a los distritos educativos y de exigir evidencias evaluativas en planes y propuestas innovadoras. Tal necesidad se materializó en una norma legal, el Título I del Acta de Educación Secundaria de 1964 del mismo país, que incluía requisitos específicos de evaluación de programas como criterio para renovar fondos ofrecidos a los distritos educativos para acelerar el mejoramiento de la educación (Madauss et al., 1983). La inexistencia de procedimientos apropiados para el seguimiento y evaluación de proyectos sociales despertó el interés por la elaboración y refinamiento de modelos y procedimientos evaluativos diferentes a los subyacentes en el examen tradicional, esto es, distintos al simple cuestionario basado en objetivos.

Se ensayó, entonces, el método experimental a través de diseños varios, experimentales y cuasiexperimentales, aplicados a programas de educación, salud y bienestar social



(Suchman, 1967; Campbell, 1969), impulsados para evaluar proyectos en varias áreas sociales desde varias décadas atrás hasta hoy día por el Banco Mundial (Grossman, 1994; Newman et al., 1994), diseños preocupados principalmente, sobre todo en las décadas de 1970, 1980 y 1990, con las repercusiones sociales o el impacto real de la inversión en programas sociales. Pero se hizo obvio pronto que este enfoque, si bien lograba evidencias bastante certeras sobre efectos debidos a determinados tratamientos o intervenciones transformadoras, dejaba por fuera elementos no verificables experimentalmente, pero de necesaria apreciación cuando de mejorar procesos se trata. Se acudió, entonces, al diseño de modelos más cualitativos como el estudio de casos, el modelo de discrepancia entre resultados y objetivos (Provus, 1969), la evaluación de expertos, el modelo de evaluación formativa y sumativa, más orientada al cliente (Scriven, 1967), el modelo de semblanza de Stake, más preocupado con procesos y con la visión de todos los interesados en los programas (1967, 1968), y otros, con lo cual el avance de la evaluación como disciplina fue notorio en la década de 1960. Muchos de los teóricos consolidados hoy en la investigación evaluativa aparecen ya en esta segunda etapa del itinerario de la evaluación y plantean modelos con respecto al saber y práctica que manejan, normas y criterios de verificación y pretensiones de sistematicidad, sin proponer aún reglas universales.

Paralelamente al desarrollo de la evaluación en educación, la inquietud por evaluar resultados y procesos se desarrolló también en el sector salud, comenzando por la acreditación de las facultades de medicina de Estados Unidos y Canadá desde los inicios del siglo XIX, en el área de condiciones de sectores urbanos deprivados, en la administración y eficacia de las escuelas públicas, en la investigación de corrupción de los gobiernos locales (Worthen, et al., 1997) y en la administración de la investigación.

Muy conocido es el libro sobre *Evaluative Research*, de Edward Suchman, profesor de Sociología y Práctica de la Salud Pública en la Universidad de Pittsburgh (1967), en el cual expone formas de evaluación de programas de salud, pero aplicables a programas de educación y bienestar. Esta obra puede considerarse como puente entre las etapas dos y tres de la historia de la evaluación de programas. El mismo Suchman, en la parte final del capítulo sobre “El crecimiento y *status* actual de la evaluación”, sostiene que una de las razones del estancamiento de la disciplina y aún del abuso que se comete en su nombre es la “ausencia de modelos y estándares para conducir los estudios evaluativos” y que “este informe (libro) abra al menos la puerta a posterior debate y análisis de las

funciones, objetivos y métodos de la investigación evaluativa” (Ibídem). Suchman concebía la evaluación como un experimento social, como aplicación de los diseños clásicos controlados a la prueba de efectividad de los programas de acción social y al uso del conocimiento resultante como guía para la acción práctica. A partir de esta crítica sobre el estancamiento de la evaluación y de su propuesta comienza una etapa de gran dinámica que llega hasta nuestros días.

Finalmente, a partir de la década de 1970 se inició una tercera etapa que enrutó la evaluación por senderos de madurez y de cientificidad. Difícil es trazar la línea divisoria entre la segunda y tercera etapas de la historia reciente de la investigación evaluativa. En ambas emergen modelos alternativos dentro de los diferentes enfoques u orientaciones. Quizás el hecho más visible en la última etapa es la consolidación de una comunidad académica identificada y fuerte a través de asociaciones de evaluación, a través de la conformación de redes internacionales y de foros y conferencias también internacionales para intercambiar temas de interés, problemas, métodos, indicadores y estándares orientadores.

Esto es lo que ha ocurrido desde 1970. Se promueven grupos profesionales cultores de este campo, se crean asociaciones profesionales en Estados Unidos, Canadá, Australia y Europa, como la American Evaluation Association, la Canadian Evaluation Society, La Sociedad Europea de Evaluación (creada en 1994), la Australasian Evaluation Society (Worthen, Sanders, and Fitzpatrick, 1997), la African Evaluation Association, la United Kingdom Evaluation Society, y la Associazione Italiana di Valutazione; se crean publicaciones periódicas especializadas, como *The American Journal of Evaluation* (antes *Evaluation Practice*) y *New Directions For Evaluation*, de la Asociación Americana de Evaluación, de carácter general, *Educational Evaluation and Policy Analysis*, de la Asociación Americana de Investigación en Educación, más específica, *Canadian Journal of Program Evaluation*, y *Evaluation* (de la Asociación Europea) con 80 números las americanas y 50 la canadiense hasta 1998; se establecen redes electrónicas de evaluadores e investigadores, como XC-Eval y Evaltalk, con grupos de discusión (On-Line) en internet; se llevan a cabo Cyberconferencias como la Primera Cyberconferencia Internacional de Evaluación, celebrada en 1998 con la participación de 25 países, entre ellos Colombia; se organizan posgrados en evaluación de programas y se refinan modelos evaluativos, apareciendo, además, nuevas propuestas con gran despliegue y acogida mundial, como es el caso de los modelos de evaluación de contexto, insumos, proceso

y producto (Stufflebeam et al., 1971), que empezó siendo un modelo de evaluación de programas en educación, pero que en sucesivas actualizaciones se proyectó como un modelo para evaluar la educación continuada en medicina, para evaluar programas de desarrollo comunitario y de servicio social (Stufflebeam, 1997); los modelos de oposición o evaluación cuasijudicial del Wolf (1979), Owens y Levine (1976); el de Análisis de necesidades de Kaufman (1983) y el modelo participativo de empoderamiento de Fetterman (1994, 1995, 1996, 1998), más orientado a capacitar a los usuarios de los programas para hacer su seguimiento y tomar decisiones sobre los mismos, poniendo énfasis, entonces, más en el contexto y en su transformación, que en el producto y en los insumos. Todos estos modelos han sido pensados y aportados para diferentes áreas en el sector social. Se concertaron también estándares internacionales para regular la evaluación, se llevaron a cabo conferencias internacionales con amplia participación de varios continentes (el 9 de enero de 1999 se llevó a cabo en Chicago un panel para discutir la creación de una Relación Colaborativa entre los profesionales de evaluación en todo el mundo, con la asistencia de la Asociación Americana de Evaluación, la Sociedad Australiana de Evaluación, la Sociedad de Evaluación del Reino Unido, la Asociación Africana de Evaluación y la Asociación Italiana de Evaluación) y se hizo el esfuerzo por fundamentar epistemológicamente los enfoques evaluativos existentes, tanto los experimentales clásicos y sus afines, más orientados a evaluar productos, como los más cualitativos, empeñados en apreciaciones más relacionadas con los procesos de implementación de los programas y las necesidades de los interesados.

Los objetivos de la evaluación en la primera etapa de la historia de esta disciplina fueron los objetivos alcanzados en el programa de instrucción, inicialmente, extendidos luego a otras áreas del sector social, es decir, comparar resultados frente a objetivos propuestos. En la segunda etapa, junto con los productos, se resaltó la importancia de los procesos y más recientemente, durante la tercera etapa, se incluyeron el contexto, los esfuerzos, insumos o estrategias generales de implementación, y los procesos y resultados organizacionales, internos y de impacto en la sociedad. Se pasó así, poco a poco, de una evaluación meramente académica o de aprendizajes, a la evaluación de programas y de instituciones.

A través de estas siete décadas de formación de la disciplina, en sus etapas de positividad, de 1930 a 1950, de disciplina en ascenso, de 1950 a 1970, y de pretensión científica, de 1970 hasta nuestros días, tres preocupaciones han guiado el desarrollo de esta formación:

la finalidad o propósito, los métodos o procedimientos para llevar a cabo la evaluación y el planteamiento epistemológico, esto es, la concepción del conocimiento que se privilegia para definir métodos de acercamiento a la realidad del objeto evaluado.

## **4.2 La evaluación en Colombia: el tránsito de la evaluación de expertos a evaluaciones orientadas por modelos**

### **4.2.1 *La práctica evaluativa***

Un primer momento de la evaluación educativa proviene de la evaluación de colegios y universidades por parte del Ministerio de Educación, las secretarías departamentales del ramo y el ICFES. La evaluación de escuelas y colegios tuvo siempre un carácter institucional, o sea, de todos los subsistemas de la escuela. Se adoptaron múltiples guías evaluativas en distintas épocas para dotar a inspectores y supervisores de instrumentos objetivos que refinaran su trabajo. A nivel de Educación Superior, la práctica se orientó a evaluar programas profesionales y el modelo más utilizado fue el de expertos, ya que se acudía a evaluadores externos y se confiaba en su pericia. El ICFES, no obstante, tuvo también pautas o guías para valorar el funcionamiento de los programas universitarios, tecnológicos y técnicos, más de carácter descriptivo y principalmente en los aspectos administrativos, hasta comienzos de la década de 1980, cuando la reforma universitaria puso de presente la necesidad de una evaluación más completa y sistematizada. En los primeros años de esta década aparecieron varios documentos y se construyó un modelo que dio cuenta de una evaluación integral (ICFES, 1985).

Otro esfuerzo evaluativo de esta primera etapa lo constituyeron las pruebas de estado o pruebas ICFES, aplicadas a los estudiantes que terminan bachillerato, con el fin de comprobar niveles mínimos de aptitudes y conocimientos de quienes aspiran a ingresar a la educación superior. La aplicación de estas pruebas se inició en 1968 con el nombre de exámenes nacionales que en 1980 pasaron a ser exámenes de estado (ICFES, 1998).

En otros ámbitos distintos al educativo, como los de salud, agricultura, bienestar social y desarrollo comunitario, se llevan a cabo evaluaciones diagnósticas y de efectividad de reformas e innovaciones, casi todas de carácter descriptivo.

#### 4.2.2 *Praxis evaluativa*

A la aplicación de pautas de evaluación sin mucho acompañamiento teórico, siguió una combinación de teoría y práctica. Los organismos internacionales comenzaron a exigir a los gobiernos e instituciones públicas y sin ánimo de lucro la inclusión de componentes evaluativos para el seguimiento de los programas y la rendición de cuentas no solo a las entidades financiadoras, sino también a la comunidad. La exigencia pasa por la elaboración de modelos sistemáticos que detecten la efectividad de los programas, modelos que van desde los diseños experimentales y cuasiexperimentales hasta la adopción o adaptación de modelos empírico analíticos y cualitativos contruidos y sustentados por los expertos en evaluación. El ICOLPE llevó a cabo seminarios y eventos de preparación de personal para evaluar diversos objetos, instituciones y programas. Es así como en 1971 y 1972 elaboró *Criterios y escalas de evaluación de textos* (ICOLPE, 1972) y entre 1973 y 1975 adelantó seminarios sobre evaluación institucional y de programas. El ICFES, por su parte, entró en la década de 1970 a manejar los modelos de evaluación reconocidos, como los de Stufflebeam y otros, para fundamentar su actividad evaluadora, produciendo en esta década y en los primeros cinco años de la de 1980 varios documentos sistemáticos sobre evaluación de programas (Cano y Kairuz, 1978; Tréllez, 1983; Botero, 1983a, 1983b; Subdirección Académica del ICFES, 1983; Tréllez, 1984; ICFES, 1985). Universidades como las de Antioquia, Valle y Javeriana introdujeron en sus currículos de postgrado en administración educativa la evaluación institucional y de programas. Stufflebeam fue invitado por la Universidad Javeriana a participar en un seminario sobre evaluación, en Bogotá, y el Departamento de Educación de la OEA llevó a cabo otro Seminario sobre Tecnología Educativa, Sistemas y Evaluación a mediados de la década de 1970. Estos eventos pusieron en boga el modelo CIPP en el país.

#### 4.2.3 *Los grandes proyectos*

A la práctica más o menos fundamentada siguió una etapa de integración teórico-práctica que tiene su momento de auge con el advenimiento de grandes proyectos educativos de las décadas de 1970 y 1980, particularmente los proyectos de los INEM, la Educación a Distancia y el Proyecto Principal de la UNESCO, en especial el Plan de Universalización de la Educación Básica, el Programa de Escuela Nueva y el Programa de Experimentación y Renovación Curricular, todos ellos con apoyo internacional que incluía, como requisito, el seguimiento y evaluación de tales proyectos.

En otros frentes se dio la Evaluación del Programa Experimental de Bibliotecas Público-Escolares de Colcultura (Isaza y Valencia, 1979); la evaluación del programa de Hogares Populares del Instituto Colombiano de Bienestar Familiar; la evaluación de los programas de Desarrollo Rural Integrado, DRI (DNP, 1979; Ministerio de Agricultura, 1989; Forero, 1990), con un enfoque descriptivo de tipo socioeconómico, sin claridad desde la perspectiva de las orientaciones teóricas de la evaluación; la evaluación del Instituto de Seguros Sociales, para la cual el doctor Ricardo Galán propuso un modelo de evaluación en 1978. El mismo doctor Galán, experto en salud, profesor universitario e investigador, ha adelantado, desde la década de 1970, propuestas de evaluación en el campo de la salud (Galán y otros, 1978).

### **– La evaluación de los INEM**

Varias evaluaciones se han hecho de los Institutos de Enseñanza Media Diversificada, INEM. La primera, más curricular, data de 1973-75, paralela al Seminario de Evaluación Institucional, adelantado por el Instituto Colombiano de Pedagogía, ICOLPE. Este instituto desarrolló y publicó documentos y modelos relacionados con la evaluación institucional (ICOLPE, 1973, 1974, 1975).

Una segunda evaluación de los INEM tuvo lugar en 1981-82 (Vélez y Troughón, 1983) y se orientó a detectar resultados laborales y académicos de los egresados de estos institutos. El estudio fue patrocinado por el Ministerio de Educación, el BID y el Instituto SER de investigación. El BID realizó luego varias evaluaciones a fondo del experimento INEM, no solo en Colombia, sino también en Brasil y en Tanzania. Psacharopoulos (1984a) efectuó evaluaciones de los INEM en Colombia y Tanzania, y Psacharopoulos, Vélez y Zabala (1984b) efectuaron otra evaluación en Colombia para comparar los niveles de empleo, ingresos y habilidades cognitivas de estudiantes egresados de estos institutos frente a los mismos atributos de egresados de instituciones de educación media tradicional. Los resultados fueron poco halagüeños. Posteriores evaluaciones arrojaron resultados similares y subrayaron la deficiente capacidad de los INEM para actualizar su infraestructura tecnológica y, aunque un estudio nacional arrojó resultados contrarios en lo referente al desempeño laboral esperado de los INEM (Puig, 1987), el plan de desarrollo del gobierno de César Gaviria desestimuló, con base en estas evaluaciones, la modalidad de educación media diversificada y sugirió dejar al SENA la función de formación vocacional que antes se había encargado a los INEM.

### – La evaluación de las escuelas normales

En 1977, el Ministerio de Educación, con el auspicio del PNUD y UNESCO, llevó a cabo una evaluación nacional de las escuelas normales (Kurmen y Briones, 1978) con los objetivos de evaluar los insumos, funcionamiento y calidad de éstas, así como su administración y estimar la situación de oferta y demanda de profesores de las escuelas primarias del país, con miras a establecer los principales problemas de las normales y hacer recomendaciones relativas a mejorar el funcionamiento, la calidad y adecuar el número de establecimientos a las necesidades del país. En la parte final del estudio se hizo una clasificación o jerarquización de las normales, y se encontró que en todo el país sólo el 11,3% de las 209 escuelas normales quedaba en la primera categoría, es decir, en la categoría de excelentes, un 12,9% en la segunda categoría, un 26,3% en la tercera, un 33,0% en la cuarta y un 16,5% en la quinta categoría, esto es, en la de más baja calidad. Se recomendó no abrir más normales y cerrar progresivamente hasta el año 1980 las que habían quedado en categorías tercera, cuarta y quinta, o sea, el 75% de estas instituciones formadoras de maestros.

### – La evaluación del Proyecto de Educación Abierta y a Distancia

Aunque la evaluación sistemática había tenido lugar en las Universidades de Antioquia y Valle, con reconocidos enfoques y modelos evaluativos a comienzos de la década de 1980, fue durante el Proyecto PNUD-Unesco-ICFES, entre 1984 y 1987, cuando la evaluación se abordó como un programa nacional. El Proyecto en mención realizó varios seminarios durante los años aludidos, en diferentes sitios del país, consiguió y puso a disposición de las universidades del país la literatura corriente en el mundo sobre evaluación y autoevaluación (Riascos, 1984), desarrolló documentos sobre evaluación y los distribuyó profusamente (Restrepo y otros, 1985; Ortiz y otros, 1986), invitó expertos internacionales a compartir sus experiencias y a servir como metaevaluadores del proyecto de evaluación de la Universidad Abierta y a Distancia de Colombia (Lafourcade, 1986), y, en fin, diseñó modelos de evaluación y los aplicó al estudio de costos, medios, rendimiento, administración y calidad de los programas de educación a distancia servidos en el país.

Sin lugar a dudas, este proyecto contribuyó ampliamente a construir la cultura de la autoevaluación de la educación superior. El ICFES compiló los informes y documentos producidos directamente por el Proyecto o contratados por éste con universidades del país.

– **La evaluación de la calidad de la educación básica primaria, de la renovación curricular y de la Escuela Nueva**

En 1986, La UNESCO, en convenio de cooperación con el Banco Mundial, realizó un importante estudio evaluativo de los aspectos de la calidad de la enseñanza básica primaria de Colombia, poniendo énfasis en la evaluación del proyecto adelantado desde el gobierno de Alfonso López con el nombre de renovación curricular, en el programa de Escuela Nueva, en los aspectos administrativos, en la capacitación del personal docente y en el mismo componente evaluativo de la renovación curricular (UNESCO, 1986). Aunque todo el documento evalúa y recomienda, el capítulo V del informe está dedicado a evaluar el componente evaluativo de la Renovación Curricular en sus etapas de experimentación y generalización desde 1977 hasta 1985 y a proponer una acción integral para mejorar la evaluación en el interior del MEN, acción que comprendería la formación de una masa crítica de evaluadores capaces de realizar “evaluaciones diagnósticas, hacer seguimientos y evaluar el impacto de las políticas implementadas en el sector educacional; la adopción e instalación de la evaluación como una práctica corriente en la programación del MEN...; y la valoración de la evaluación por parte de los cuadros decisionales...” (UNESCO, *Ibídem*).

El programa de Escuela Nueva, por lo demás, ha sido objeto de varias evaluaciones realizadas por el mismo Ministerio de Educación y por el Instituto Ser de investigaciones.

– **La evaluación del Proyecto Principal de la Unesco para América Latina y El Caribe**

El Proyecto Principal abarcó de los años 1982 a 1999 y se fijó tres objetivos básicos relacionados con la escolarización para 1999 de todos los niños en edad escolar: eliminar el analfabetismo antes del fin del siglo, ampliar el servicio educativo para los adultos y mejorar la calidad y eficiencia de los sistemas educativos. El Proyecto fijó como máxima entidad de seguimiento y evaluación las Reuniones del Comité Intergubernamental del Proyecto Principal en la Esfera de la Educación en la Región de América Latina y El Caribe, PROMEDLAC. La segunda de estas reuniones se llevó a cabo en Bogotá en 1987. A su vez, cada país integró una Comisión Nacional del Proyecto Principal para hacer la evaluación de logros de los objetivos del Proyecto y presentó ante las reuniones de PROMEDLAC informes sobre los avances y problemas, así como recomendaciones para el mejor cumplimiento de los compromisos adquiridos frente al Proyecto (MEN,



1987). Este gran proyecto, por sus implicaciones y compromiso internacional, fue otro mojón en el desarrollo de la evaluación en el país, particularmente en el área educativa. El programa de alfabetización CAMINA y el Plan de Universalización de la Enseñanza Básica Primaria, materializaciones del Proyecto Principal, fueron objeto de evaluación sistemática.

### – La enseñanza de la evaluación

Paralelamente a la exigencia evaluativa de estos proyectos, la universidad colombiana se vio exigida con respecto a la preparación de recurso humano para esta labor. Fue así como en la década de 1970 universidades como las de Antioquía, Javeriana y Valle incluyeron, en sus currículos de posgrado en administración, planeación e investigación, cátedras de evaluación institucional y de programas. Abrieron también la línea de investigación evaluativa en sus tesis de grado. Posteriormente, la discontinuidad de estos posgrados disminuyó el cultivo de esta línea que casi desapareció en estas universidades.

#### 4.2.4 *Evaluación de programas educativos más particulares*

En la década de 1970 se llevó a cabo la evaluación de Acción Cultural Popular por parte de un equipo de evaluadores de la Universidad Estatal de la Florida (Morgan, 1980), quienes propusieron un modelo para evaluación de sistemas de comunicación educativa.

El SENA, por su parte, realizó en 1988 y 1989 una evaluación integral, de carácter nacional, de su programa FAD de formación a distancia, utilizando para ello el modelo de evaluación responsiva de Stake, que intenta detectar no solo productos e impacto de los programas, sino comprender paralelamente los procesos experimentados por los mismos (Restrepo, Correa y Flórez, 1989).

### 4.3 Necesidad de la evaluación hoy

¿Qué vigencia mantiene la investigación evaluativa en el presente? ¿Cómo se está desarrollando la tercera etapa de formación de la evaluación de programas como disciplina? Echemos un vistazo a esta situación en el mundo y en nuestro país para escudriñar desarrollos y retos.

#### 4.3.1 *La tendencia mundial*

La modernización del Estado trae aparejada la necesidad de evaluar permanentemente la efectividad, eficiencia y productividad de las políticas, instituciones, programas, proyectos y recursos que la implementación de la modernidad implica. El progreso de la evaluación debe tanto al seguimiento de la calidad de la educación, política universal en las últimas décadas, como al seguimiento de los programas de guerra contra la pobreza, utilizando el PPBS, Sistema de Planeación, Programación y Presupuesto, adaptado de la Compañía Ford por Robert McNamara, secretario de Defensa del Presidente Kennedy y recomendado para todas las agencias federales (Worthen, te. al., 1997).

Más tarde, en 1970, McNamara, al frente del Banco Mundial, creó un grupo de evaluación como parte del Departamento de Planeación y Presupuesto. En 1975, este departamento fue elevado a Departamento de Evaluación de Operaciones (OEA) con la misión de evaluar todas las actividades del Banco Mundial. En 1994, dicho departamento evaluó 350 operaciones de desarrollo en variedad de sectores: agricultura, cultura, energía, finanzas, transporte, provisión de agua, desarrollo urbano, educación, salud, evaluación que abarca políticas, programas y proyectos en numerosos países (Piccioto, 1997). El OED sintetizó los hallazgos de todas sus evaluaciones en su *Anual Review of Evaluation Results* (Ibídem).

En las décadas de 1960 y 1970, el enfoque predominante de evaluación seguido por el Banco fue el orientado a resultados, a productividad: *input-output*, tasas de retorno, costo-beneficio, para determinar asignación de recursos. Se siguió, en cierta forma, la orientación tyleriana de evaluación de logro de objetivos. En la década de 1980 el enfoque anterior se mantuvo, pero combinado con el enfoque de beneficios indirectos de los programas, de la pertinencia o utilidad social de estos frente a las prioridades de los países y de su aporte al propósito de reformas y de construcción de capacidad nacional, regional y local para emprender desarrollo sostenible. Ya la importancia no recae tan exclusivamente en la efectividad y eficiencia de cada programa individual, sino que se tiene en cuenta su contribución al desarrollo nacional. Éste es criterio vital en la evaluación de programas del Banco (Piciotto, 1997).

En la era de las ONG, la evaluación de programas y proyectos sociales, bien sea en el área educativa, en la de desarrollo comunitario (Giraldo, 1993), formación de líderes, programas de vivienda social, salud, nutrición, derechos de los niños, evaluación de

violación de los derechos humanos, evaluación de la ayuda extranjera, programas de mejoramiento de acueductos y alcantarillado y, por supuesto, proyectos de desarrollo social y económico. La evaluación ha irrumpido como un componente esencial en la financiación y refinanciación de tales programas. El Banco Mundial, UNICEF, UNESCO y otros organismos internacionales demandan la evaluación de todos los programas relacionados con política social que promueven o financian (Grossman, 1994).

La Oficina Regional de UNICEF para América Latina y El Caribe ha desarrollado, a partir de un Taller sobre Orientación de Programas, realizado en Nueva York en 1997, ha desarrollado líneas estratégicas regionales relacionadas con evaluación, tales como: fortalecimiento de la cultura de la evaluación, mejoramiento de la planeación y ejecución integradas de seguimiento y evaluación, fortalecimiento de la capacidad técnica, desarrollo de metodologías para evaluar áreas emergentes relacionadas con la infancia y fortalecimiento de la evaluación democrática (Segone, 1998).

Puede decirse que una de las tendencias de la evaluación en la actualidad es su vinculación a debates sobre políticas de estado, lo cual ha puesto sobre el tapete la ingerencia de la política en la evaluación. Chelimsky lo ha anotado (1997): la práctica de la evaluación se ha extendido por los cuatro puntos cardinales del globo. Instituciones internacionales como los Programas Ambientales y de Desarrollo de las Naciones Unidas, la Comisión Europea, el Banco Mundial, las Fundaciones Kellog y Rockefeller han incrementado notablemente su uso de la evaluación para detectar resultados de sus proyectos. Pero la importancia de tales proyectos hace vulnerable la evaluación a las presiones políticas, a la presión de múltiples grupos que luchan por obtener fondos para sus programas.

Chelimsky (1997) señala seis necesidades que los evaluadores deben atender como resultado de las implicaciones provenientes del ambiente político de la evaluación en la actualidad: la producción de informes defendibles y dotados de credibilidad, el valor para luchar por independencia de la presión política, la diseminación de hallazgos críticos en la evaluación, el desarrollo de vínculos con otras disciplinas y con la investigación básica, la consecución de mejor acceso a los datos y la incorporación del realismo político en la capacitación de los evaluadores.

La Fundación Kellog, segunda más grande del mundo después de la Fundación Ford, ha diseñado un manual de evaluación para los programas subvencionados por la misma. La

presentación del manual refuerza la convicción de que “la evaluación tiene que ser un apoyo comprensible y sensato de los proyectos, más que una finalidad en sí misma” (Kellogg Foundation, 1995). La Fundación subvenciona proyectos de salud, desarrollo comunitario, formación de líderes, educación...

En Suecia y Noruega, la evaluación ha sido utilizada como estrategia para hacer seguimiento a reformas políticas y evaluar el desarrollo industrial (Huang, 1998; Karlsson, 1998), mientras que en Dinamarca se la ha utilizado como estrategia de modernización dentro del Estado benefactor (Andersen, 1998). En Francia, la evaluación ha seguido este mismo derrotero, es decir, se orienta más a evaluar políticas estatales que programas concretos (Ambrois, 1998). En 1990, un decreto del 22 de enero creó un mecanismo interministerial de evaluación, mecanismo que fue ratificado y reforzado en 1993 con una circular acerca de la necesidad de elaborar estudios de impacto de las políticas oficiales (Ibídem, 1998). Se supone que las más importantes políticas nacionales deben ser evaluadas y sus resultados publicados.

En educación, la evaluación se ha hecho presente de muchas maneras. En la presente década, con singular fuerza, el movimiento de evaluación y acreditación de la educación superior, que ya existía en los países desarrollados (Estados Unidos—donde la acreditación de universidades data desde finales del siglo XIX— (Worthen, et al., 1997), Reino Unido, Francia, Alemania, Dinamarca, Países Bajos), ha incursionado en la mayoría de los estados. En España, por ejemplo, Se desarrolló, entre 1992 y 1994, el Programa Experimental de Evaluación de la Calidad de la Universidad, organizado por 17 universidades. Este programa se conectó en 1994 con el Proyecto Piloto de la Unión Europea y como consecuencia de ambos proyectos se logró, en 1995, el Plan Nacional de Evaluación de la Calidad, por propuesta del Consejo de Universidades, plan que instauró con carácter regular y sistemático la evaluación con miras al mejoramiento de la calidad (De Luxán Meléndez, 1998). En 1998, España se atuvo más al modelo europeo dentro de su integración a los países del área.

Cuba practica una evaluación institucional y de programas de sus centros de educación superior basada en el cumplimiento de objetivos institucionales y de los objetivos de los planes y programas establecidos para la formación de los profesionales. La evaluación es un elemento clave para el perfeccionamiento del trabajo, la calidad y la pertinencia (Alemany, 1998).

Chile está experimentando últimamente una creciente diversificación del sistema de educación superior y ante este hecho se considera que uno de los retos del momento es “definir criterios de evaluación adecuada para los diversos programas que se ofrecen en el país” (Lemaitre, 1999), a fin de complementar los sistemas de autoevaluación diseñados por algunas universidades tradicionales y orientados al licenciamiento de las instituciones de educación superior, esto es, a la instalación y consolidación institucional para lograr la autorización de funcionamiento autónomo. El país dispone también de evaluadores con amplia experiencia en “diversos tipos de evaluación (institucional, de gestión, financiera, de programas, de examinación)” (Lemaitre, Ibídem). Se trabaja en el diseño de un sistema nacional de acreditación voluntaria de programas, que parte de un proceso de autoevaluación de la institución interesada. En la actualidad, la acreditación es obligatoria, pero sólo para las universidades privadas (Ibídem).

En Colombia, el movimiento de evaluación de la educación superior tuvo desarrollos fuertes en la década de 1980 con el advenimiento de la Educación Abierta y a Distancia. En la Ley 30 de 1992 se dio gran importancia a la evaluación y acreditación como correlatos de la autonomía universitaria y en los decretos reglamentarios se perfeccionó este mandato legal con la creación del Sistema Nacional de Acreditación.

Los métodos para el manejo de la calidad total han sido impulsados en Japón y Estados Unidos como métodos para el mejoramiento de la calidad y como procesos de evaluación formativa y sumativa (Kanji and Asher, 1996).

Otro campo de gran desarrollo en las últimas dos décadas es el de la evaluación de la investigación, incluyendo la internacionalización de este proceso, es decir, el sometimiento de los proyectos e informes de investigación al juicio de la comunidad científica internacional.

Para terminar, la necesidad y vigencia de la evaluación hoy es bueno plantear las tres perspectivas actuales de la evaluación según Chelimsky. Para esta autora, las perspectivas o tendencias son la de rendición de cuentas, la perspectiva de conocimiento y la perspectiva de desarrollo (1997). La de **rendición de cuentas** preconiza la medición de resultados para justificar gastos e inversiones. Su utilidad se centra en el apoyo a las reformas. La perspectiva de **conocimiento** busca comprender problemas públicos, políticas y programas; desarrollar nuevos métodos y someter a crítica y ajuste los

anteriores. Su utilidad tiende más hacia la investigación. La perspectiva de **desarrollo**, por su parte, se orienta a fortalecer las instituciones, programas y proyectos, o sea, a construir capacidad organizacional.

#### 4.3.2 *La tendencia en Colombia*

La modernización de la educación en Colombia está comprendida principalmente en las Leyes 30 de 1992, sobre Educación Superior, y 115 de 1994 o Ley General de Educación. Tales leyes desarrollaron principios y políticas de la Constitución de 1991, como la descentralización, la autonomía, la acreditación, la evaluación de docentes y la autogestión, entre otros. Para monitorear la actuación de estos aspectos innovadores y ejercer las medidas de control necesarias para ajustar su desempeño, es menester diseñar y aplicar sistemas apropiados de evaluación.

##### 4.3.2.1 *Ley 30 de 1992: Evaluación de la Educación Superior*

La Ley 30 de 1992 creó instituciones inexistentes antes en el país y propulsoras de la calidad y eficiencia de la Educación Superior, como el Sistema Nacional de Acreditación, el Sistema de Información de la Educación Superior y el Sistema de Universidades Estatales, SUE. La acreditación, en particular, tiene que ver con la evaluación.

#### – **La acreditación**

Esta práctica, orientada a que las universidades e instituciones de educación superior rindan cuentas de su calidad y cumplimiento de objetivos a la comunidad, a través de un reconocimiento del Estado, se basa en la autoevaluación de las mismas instituciones y en evaluación externa adelantada por pares académicos de reconocida solvencia. Es una institucionalización de la evaluación que supone la preparación de evaluadores en todas las entidades de educación superior.

#### – **El crecimiento de la Educación Superior**

En las últimas décadas se ha presenciado un crecimiento desmesurado de la cobertura de la educación superior, crecimiento especialmente grande a partir de la Ley 30 que apuntaló la autonomía universitaria y facilitó la creación de programas de pregrado y

especialización. La Ley abrió también la compuerta para la transformación del carácter académico de las instituciones de educación superior, facilitando la transformación de nivel a nivel de estas instituciones, pero exigiendo un proceso de evaluación que comprende un diagnóstico socioeconómico de necesidades, un estudio de factibilidad, rediseños curriculares, estudio de evolución de los aspectos académicos, de la estructura administrativa y de los recursos de todo tipo de la institución solicitante. En la base de todo este proceso subyace un serio estudio evaluativo de la entidad y de su proyección futura. Nuevos campos se abren, así, a la evaluación de programas.

### – Evaluación de la calidad de la educación básica y media

El auge que en el mundo ha tomado la evaluación de la calidad de la enseñanza básica y media a través de pruebas como el TIMSS (Third International Mathematics and Science Study) promovió en Colombia la evaluación de la calidad en estos niveles de enseñanza desde 1992, primero en matemáticas y lenguaje en los grados tercero y quinto y luego en otras áreas en estos grados, así como en séptimo y noveno. El Informe Saber, como se conocen los informes de esta evaluación permanente, cubre no solo aspectos cognoscitivos, conocimientos y habilidades, sino también factores socioeconómicos asociados con el rendimiento.

#### 4.3.2.2 Ley 115 de 1994: Evaluación de la Educación Preescolar, Básica y Media

La Ley 115 de 1994 plantea necesidades concretas en evaluación de programas y desempeño profesional. ¿Qué componentes de esta Ley demandan conocimiento sobre evaluación y promueven desarrollos significativos en el área? La descentralización, la autonomía curricular, la acreditación, la evaluación de docentes, de directivos y de la institución como un todo.

Los PEI son un instrumento de descentralización pedagógica y de gestión, incorporados en la Ley 115 como medio para implementar la descentralización y promover la innovación en la institución escolar. Pero el cumplimiento de estos propósitos no se logra con la sola reglamentación contenida en el Decreto 1860. Es menester evaluar su funcionamiento y resultados en los primeros años de aplicación con miras a realizar ajustes y seguir evaluando su efectividad. El Ministerio de Educación contrató con universidades y otros organismos la primera evaluación de los PEI en distintos municipios del país a finales de 1997 y comienzos de 1998.

La Ley 115 en sus artículos 80, 81, 82, 83 y 84 ordena crear un sistema de evaluación de la calidad de la educación, evaluar cada seis años la idoneidad académica y la actualización pedagógica de los docentes, evaluar a los directivos docentes del país, oficiales y privados, y llevar a cabo la evaluación institucional anual de las instituciones educativas con el fin de mejorar su calidad. Este mandato es un acicate para promover el desarrollo de la evaluación.

En 1998, el MEN expidió el Decreto 272 que obliga la acreditación previa de los programas de formación de docentes, medida que compromete a las instituciones respectivas a evaluar su accionar, su currículo y sus insumos como paso previo para la mencionada acreditación.

#### 4.3.2.3 *Proyectos de mejoramiento de la calidad en las regiones. Créditos internacionales*

Finalmente, los organismos internacionales como el Banco Mundial, el BID y otros, están condicionando los créditos para la educación y otras áreas a la inclusión en los proyectos de subsistemas de evaluación y seguimiento de sus procesos y productos, así como de sus impactos. Así se ha exigido en recientes proyectos para el mejoramiento de la calidad de la enseñanza básica en el Departamento de Antioquia y en la ciudad de San Juan de Pasto.

4.3.2.4 *La especialización en teoría y métodos de la investigación social*, auspiciada por ICFES y ASCUN desde 1997 y ofrecida en varias universidades colombianas, tiene como una de sus especialidades o énfasis la investigación evaluativa.

## 5. Coda

Por todo lo analizado, es obvio que la cultura de la evaluación ha ido penetrando en Colombia y se hace indispensable que su práctica esté acompañada de construcción teórica, de preparación de recurso humano capaz de adelantar seriamente estudios evaluativos que repercutan en aprendizaje organizacional y en el mejoramiento de la calidad, la eficiencia y la productividad de instituciones y programas.



## Lecturas complementarias

Forero, J. *Evaluación general del programa de desarrollo Rural integrado, DRI, en Colombia. Algunos elementos para una evaluación descentralizada de proyectos*. Bogotá: Universidad Javeriana y Ministerio de Agricultura, 1990, pp. 198-210.

Restrepo, B. *La autoevaluación como proceso previo a la acreditación de programas de pregrado en la Universidad de Antioquia. Proyecto general*. Medellín: Vicerrectoría de Docencia, Universidad de Antioquia, 1998.

## 6. Autoevaluación

1. ¿Cuáles fueron las orientaciones de los modelos más conocidos de la primera y última etapas del desarrollo de la investigación evaluativa y qué paradigmas metodológicos utilizaron uno y otro?
2. ¿Qué hechos caracterizaron la tercera etapa de la historia de la evaluación de programas? Comente al menos tres de ellos.
3. ¿Qué fases en el desarrollo de la evaluación de instituciones y programas pueden observarse en Colombia en los últimos 40 años y cuál es su naturaleza?
4. ¿Qué vigencia tiene hoy la evaluación de programas sociales en Colombia y cómo se materializa esta vigencia en el sector educativo, por ejemplo?

## 7. Actividades de evaluación

1. Realice un mapa conceptual a partir del concepto de vigencia actual de la evaluación de programas sociales en Colombia y en el mundo. Asegúrese de incluir la presencia política en la evaluación y el papel que debe hacer el evaluador frente a ésta.
2. Elabore una contrastación entre las aplicaciones dadas a la evaluación por los países escandinavos y Francia, por un lado, y por Estados Unidos y el Banco Mundial, por otro.

## BIBLIOGRAFÍA

Alemaný, S. "Evaluación institucional en las universidades cubanas: premisas y objetivos de la acreditación". Ponencia presentada en el Seminario-Taller sobre Educación Superior y Acreditación en los Países del Convenio Andrés Bello. Medellín: U de A-SECAB, 1998.

Andersen, L. *Evaluation as a Strategy of Modernization. New Directions for Evaluation*. No. 77, 1998, pp. 39-54.

Botero, M. *La práctica evaluativa del ICFES ante los desarrollos de la teoría sobre esta área*. Bogotá: Subdirección Académica; División de Formación Avanzada, ICFES, 1983a.

———. *Fundamentos teóricos de la práctica evaluativa del ICFES*. Bogotá: Subdirección Académica; División de Formación Avanzada, ICFES, 1983b.

Cadenas, J. M. La educación superior en Venezuela. Ponencia presentada en el Seminario-Taller sobre Educación Superior. Medellín: UdeA-SECAB, 1998.

Cano, F. y Kairuz, V. *Manual de Evaluación de Instituciones y programas de Educación Superior*. Bogotá: ICFES, 1978.

Campbell, D. T. "Reform as Experiments". *American Psychologist*, No. 4, 24, 1969.

Chelimsky, E. *The political environment of evaluation and what it means for the development of the field. Evaluation for the 21st Century*. London: Sage Publications; 1997, pp.53-68.

De Luxán Meléndez, J. M. "La evaluación de la universidad en España". Ponencia presentada en el Seminario-Taller sobre Educación Superior y Acreditación en los Países del Convenio Andrés Bello. Medellín: U de A-SECAB, 1998.

DNP. *Metodología de evaluación de impacto del programa de desarrollo rural integrado*, DRI. Informe Final. Bogotá: DNP, 1979.

Fetterman, D. "Empowerment Evaluation". *Evaluation Practice*, N. 15, 1994.

\_\_\_\_\_. In Response. *Evaluation Practice*, N. 16.

Fetterman, D., Kaftrian, S. and Wandersman, A. *Empowerment Evaluation: Knowledge and tool for Self-Assessment and Accountability*. Thousand Oaks, CA: Sage Publications, 1996.

Forero, J. *Evaluación general del programa de desarrollo rural integrado, DRI, en Colombia*. Bogotá: Universidad Javeriana, 1990.

Foucault, M. *La arqueología del saber*. México: Siglo XXI Editores, 1972.

Fundación W. K. Kellogg. *Manual de evaluación*. Versión en Español. Buenos Aires: Crisol, Proyectos Sociales, 1999.

Galán, R. y otros. *Modelo de evaluación para el I.S.S.* Bogotá: Ministerio de Salud, Superintendencia de Salud, 1978.

Giraldo, F. E. Evaluación de impacto social. Un modelo de evaluación de procesos cualitativos basados en la experiencia de trabajo comunitario de Futuro para la Niñez. Medellín: Futuro para la Niñez. Litoimpresos, 1993.

Grossman, J. B. "Evaluating Social Policies: Principles and U.S. Experience". *The World Bank Research Observer*, Vol. 9, No. 2, 1994.

HAUG, P. "Linking Evaluation and Reform Strategies". *New Directions for Evaluation*. No. 77, 1998, pp. 5-20.

ICFES. *Autoevaluación institucional*. Bogotá: Subdirección Académica, 1983.

ICFES. *La evaluación académica en el ICFES*. Bogotá: Subdirección Académica, 1985.

ICFES. Exámenes de Estado. Home Page. SNP, 1998.

ICOLPE. Lecturas sobre definición del concepto de evaluación; metodología de evaluación institucional; definición de evaluación educacional. Bogotá: Centro de Documentación e Información Pedagógica, CENDIP, 1973, 1974, 1975.

Isaza, Y. y Valencia de V. M. *Evaluación del Programa Experimental de Bibliotecas Público-Escolares de Colcultura*. Medellín: Universidad de Antioquia; Escuela Interamericana de Bibliotecología, 1979.

Kanji, G. P. and Asher, M. "100 Methods for Total Quality Management. Newbury Park, CA: Sage". *Reviewed by David Coulton, in American Journal of Evaluation*. Vol 19, No. 2, Summer, 1998, pp. 273-275.

Karlsson, O. "Socratic Dialogue in the Swedish Political Context". *New Directions for Evaluation*. No. 77, 1998, pp. 21-38.

Kaufman, R. y otros. "La relación entre la evaluación de necesidades y el desarrollo, implementación y evaluación de necesidades". *Revista de Tecnología Educativa*, Vol. 18, No.1, 1983.

Kurmen, C. y Briones, G. *Evaluación de las escuelas normales de Colombia*. Bogotá: Ministerio de Educación Nacional. Oficina Sectorial de Planeación Educativa. Programa MINED-PNUD-UNESCO. Proyecto Col 76/003, 1978.

Lafourcade, P. D. *Consultoría en investigación evaluativa: aportes para esclarecer los planteamientos contemplados en las funciones asignadas*. Bogotá: ICFES. Proyecto PNUD-UNESCO-ICFES, 1986.

Lemaitre, M. J. Una experiencia de aprendizaje y el diseño de un nuevo sistema. El caso de Chile y el Consejo Superior de Educación. Caracas: IESALC/UNESCO, 1999.

Levine, M. *Experiences in adapting the jury trial to the problem of educational program evaluation*. Buffalo: State University of New York, 1976.

MEN. Segunda Reunión del Comité Regional Intergubernamental del Proyecto Principal en la Esfera de la Educación en América Latina y El Caribe: Informe Nacional del Proyecto Principal. Bogotá: Editorial del Ministerio de Educación Nacional, 1987.

Newman, J. et al. "Using Randomized Control Designs in Evaluating Social Sector in Developing Countries". *The World Bank Research Observer*, vol. 19, No. 2, 1994.

Ministerio de Agricultura. *Proyecto sistema de control, seguimiento y evaluación del Fondo DRI*. Documento preliminar de discusión. Bogotá: Unidad de Planeación, 1989.

Ortiz, A. y otros. *Estudios de caso sobre los procedimientos administrativos utilizados para el desarrollo del proceso autoevaluativo en el SED*. Bogotá: Proyecto PNUD-UNESCO-ICFES, 1986.

Psacharopoulos, G. *Evaluation of Curriculum Diversification in Colombia and Tanzania*. Washington, D.C.: IRDB, Discuss, R.P.O 672-45, 1984a.

Psacharopoulos, G., Vélez, E. y Zabala, A. *Una evaluación de la educación media diversificada en Colombia*. Bogotá: BIRD, MEN, Instituto SER, 1984b.

Picciotto, R. *Evaluation in the World Bank: Antecedents, Instruments, and concepts. Evaluation for the 21st Century*. London: Sage Publications, 1997, pp. 201-214.

Puig, J. *La escolarización de minorías populares en la escuela comprensiva Colombiana*. Medellín: Universidad de Antioquia-Copiyepes, 1987.

Restrepo, B. y otros. *Una propuesta flexible de autoevaluación para los programas de educación a distancia*. Bogotá: Proyecto PNUD-UNESCO-ICFES, 1985.

Restrepo, B., Flórez, R. y Correa, S. *Situación actual y perspectivas de la formación a distancia en el SENA*. Medellín: Universidad de Antioquia, Centro De Investigaciones Educativas, 1989.

Restrepo, B. *Itinerario de la evaluación como disciplina*. Ponencia presentada en el Seminario Internacional de Evaluación en Educación. Santa Fe de Bogotá: IDEP, 1997.

Riascos, B. *Bibliografía comentada sobre evaluación, investigación evaluativa y evaluación de la educación superior a distancia* (segunda versión). Bogotá: ICFES, Subdirección de Fomento, 1984.

Segone, M. *Democratic Evaluation*. Santa Fe de Bogotá: UNICEF Regional Office for Latin America and the Caribbean, 1998.

Stake, R. "The Countenance of Educational Evaluation". *Techers College Record* 68: 523-540, 1967.

Stufflebeam, D. et al. *Educational Evaluation and Decision Making*. Bloomington, Indiana: Phi, Delta, Kappa, in, 1971.

Stufflebeam, D. *The CIPP Model for Evaluation. Draft of Chapter 6. of a new book on evaluation*. Kalamazoo: The Evaluation Center, Western Michigan University, 1997.

Suchman, E. *Evaluative Research*. New York: Russell Sage Foundation, 1967.

Tréllez, E. *Procesos de la evaluación académica en el ICFES*. Bogotá: División de Formación Universitaria, ICFES, 1983.

———. *Fundamento y método de la evaluación académica*. Bogotá: Subdirección Académica; División de Formación Universitaria, ICFES, 1984.

Tyler, R. W. *Characteristics of a satisfactory diagnosis in Educational Diagnosis*. Edited by G.M. Whipple. NSSE Yearbook, 1935, pp. 95-112.

UNESCO, *Colombia: aspectos de la calidad de la enseñanza básica primaria*. París: UNESCO, 1986.

Vélez, E. y Troughón, P. L. “Logros educativos del bachiller colombiano: el caso de la cohorte de 1978”. *Revista Colombiana de Educación*, No. 12, 1983, pp. 77-99.

Villarroel, C. *La evaluación institucional en las universidades venezolanas*. Ponencia presentada en el Seminario-Taller sobre Educación Superior y Acreditación. Medellín: U de A-SECAB, 1998.

Wolf, R. L. “The Use of judicial Evaluation Methods in the Formation of Educational Policy”. *Educational Evaluation and Policy Analysis*, 1, 19-28, 1979.

Worthen, J., Sanders, J., and Fitzpatrick, J. *Program Evaluation*. New York: Longman Publishers, 1997.

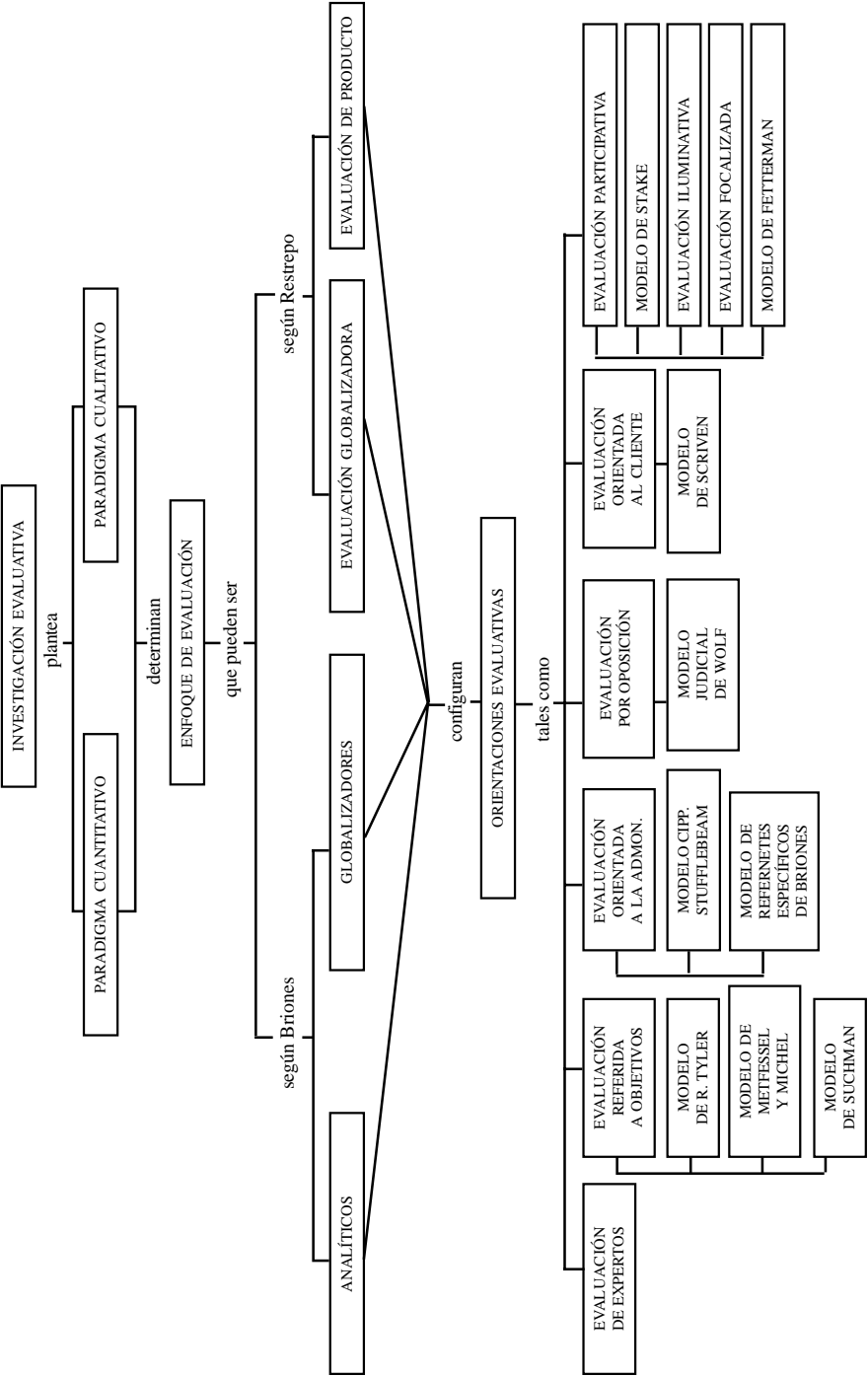


# **Paradigmas - enfoques - orientaciones y modelos evaluativos**

**Santiago Correa Uribe**  
**Profesor Universidad de Antioquia**

TERCERA UNIDAD

Mapa conceptual



## INTRODUCCIÓN

Esta unidad pretende abordar los grandes paradigmas de la investigación evaluativa, los enfoques que de dichos paradigmas se derivan, las orientaciones o grandes propósitos que guían las propuestas evaluativas y, por último, los modelos más representativos dentro de las orientaciones aquí definidas.

Se inicia con una revisión a los paradigmas cualitativo y cuantitativo y sus implicaciones en la investigación evaluativa, como condición para comprender los diferentes enfoques o también denominadas escuelas de evaluación.

Continúa la unidad con una aproximación a las orientaciones predominantes en la investigación evaluativa: evaluación referida a objetivos, evaluación de expertos, evaluación orientada a la administración, evaluación por oposición o evaluación contrapuesta, evaluación orientada al cliente o centrada en el consumidor y evaluación participativa.

Dentro de cada una de estas orientaciones se presentan los modelos más representativos con sus características principales.

## OBJETIVOS DE LA UNIDAD

1. Diferenciar los paradigmas cualitativos y los paradigmas cuantitativos en la Investigación evaluativa y analizar sus compatibilidades y posibilidades de complementariedad.
2. Identificar los enfoques predominantes en la investigación evaluativa.
3. Caracterizar las orientaciones predominantes en la investigación evaluativa y los modelos subyacentes en cada una de ellas.
4. Analizar las particularidades de los diferentes modelos de evaluación y sus posibles campos de aplicación.

## Tercera unidad

### PARADIGMAS-ENFOQUES-ORIENTACIONES Y MODELOS EVALUATIVOS

#### I. PARADIGMAS DE LA INVESTIGACIÓN EVALUATIVA

La polémica sobre los métodos de investigación en las ciencias sociales no es nueva. Estudiosos del tema ubican sus orígenes en los filósofos presocráticos, cuyas nociones y tomas de posición son esenciales en la historia de las relaciones entre lo “cuantitativo” y lo “cualitativo”, sin embargo se puede situar en la Grecia clásica el lugar y el momento en el que se inicia de una forma más clara la polémica entre lo cuantitativo y lo cualitativo siendo Platon y Aristóteles las dos figuras y concepciones emblemáticas del planteamiento inicial de la polémica. “Polémica que originalmente estaba centrada en torno a la confrontación entre las visiones mas **formalista** y mas **sustantivista** defendidas por ambos filósofos. De esta forma, ambos autores/concepciones fundaron las bases filosóficas en los que, posteriormente, se iban a enmarcar las distintas aproximaciones y desarrollos que, desde hoy, podemos denominar e inscribir como perspectivas CUALITATIVAS y CUANTITATIVAS” (Conde, 1994:54).

Las relaciones entre las perspectivas CUALITATIVAS y CUANTITATIVAS, en la investigación social, se abordan desde diferentes puntos de vista: epistemológicamente, metodológicamente y técnicamente. Recientes debates consideran la confrontación no solo como un desacuerdo respecto a las ventajas y desventajas relativas de cada uno de los métodos (cuantitativos y cualitativos), sino como una confrontación básica entre paradigmas metodológicos. “Según esta concepción, cada tipo de método se halla ligado a una perspectiva paradigmática distintiva y única y son estas dos perspectivas las que se encuentran en conflicto”. (Reichardt y Cook, 1986).

Tal como lo expone Rist citado por Reichardt: “En definitiva, la cuestión no estriba per se en unas estrategias de investigación: lo que sucede más bien es que la adhesión a un paradigma, y su oposición a otro, predispone a cada uno a concebir el mundo y los

acontecimientos que en él se desarrollan de modos profundamente diferentes”. Rist, (1977:43)

El concepto de paradigma procede de Kuhn (1962-1970), basándose en sus trabajos, Patton citado por Reichardt, define un paradigma como:

“Una visión del mundo, una perspectiva general, un modo de desmenuzar la complejidad del mundo real. Como tales, los paradigmas se hallan profundamente fijados en la socialización de adictos y profesionales; los paradigmas les dicen lo que es importante, legítimo y responsable. Los paradigmas son también normativos; señalan al profesional lo que ha de hacer sin necesidad de prolongadas consideraciones existenciales o epistemológicas”. (Patton,1978:203)

Lo expuesto anteriormente permite inferir que en la consolidación de un determinado paradigma confluyan entre otros factores los siguientes:

- La naturaleza del conocimiento y de la realidad.
- La naturaleza de las relaciones entre el investigador y el conocimiento que generen.
- El modo de construir el conocimiento.

Estos factores aparecen tratados en la Unidad 1 del módulo de *Investigación cualitativa*, por lo cual no profundizaremos en su análisis.

En una aproximación a la caracterización de los dos paradigmas se podría decir:

El paradigma cuantitativo posee una concepción global positivista, donde el experimento y el tratamiento estadístico de la información se convierten en las principales herramientas de las que dispone el investigador. El marco de observación y de análisis está orientado por la formulación de hipótesis y preguntas que pretenden anticipar el comportamiento de la realidad objeto de estudio (hipotético deductivo). Es particularista al pretender identificar las variables presentes en dicho objeto de estudio, las cuales son sometidas a pruebas empíricas bajo condiciones de control riguroso con la pretensión de garantizar objetividad en sus resultados. Está orientada a los resultados, los cuales se pretenden explicar en términos de principios y leyes. Propio de las ciencias naturales.

El paradigma cualitativo propone una concepción global fenomenológica, inductiva, estructuralista, subjetiva, orientada al proceso y propia de la antropología social.

Esta confrontación paradigmática ha impactado igualmente a la investigación evaluativa, donde después de una larga tradición de corte experimental emerge con fuerza la perspectiva cualitativa, sin que esto implique descartar, ni siquiera infravalorar, la perspectiva cuantitativa. Se pretende más bien defender la idea de que existen diferentes opciones metodológicas para abordar este tipo de investigación.

Tal vez lo más novedoso, en el momento actual, resida en el interés por buscar las compatibilidades y la complementariedad entre las dos tendencias que posibiliten el trabajo conjunto y que proponen un nuevo panorama de investigación denominado por algunos teóricos INVESTIGACIÓN TOTAL.

## **II. ENFOQUES DE EVALUACIÓN**

Existen diferentes posiciones en torno a lo que se podría denominar ESCUELAS o ENFOQUES de evaluación. Así, Restrepo (1974) identifica dos grandes escuelas de evaluación: la que tiene más en cuenta los resultados y la congruencia de estos con los objetivos, que se denomina EVALUACIÓN DE PRODUCTO y aquella que presta atención al contexto, a los insumos, a los procesos y a los productos, que se denomina EVALUACIÓN GLOBALIZADORA.

Por su parte, BRIONES (1991) clasifica los enfoques de evaluación de los programas sociales en dos grupos: analíticos y globales.

Los enfoques ANALÍTICOS se caracterizan por abordar con mayor o menor desagregación los componentes estructurales del programa, los cuales a su vez están conformados por características o variables con las cuales es posible determinar relaciones específicas. Hacia estos componentes, características o variables se orientan los procesos de recolección, selección y análisis de la información. Los enfoques analíticos también admiten evaluaciones parciales de sólo algunos componentes atendiendo a las necesidades del programa.

Desde la perspectiva epistemológica, los enfoques analíticos son de corte racionalista con posiciones realistas u objetivistas, aunque pueden presentar matices naturalistas y utilizan tanto métodos cuantitativos como cualitativos de investigación.

Los enfoques GLOBALES pretenden hacer una evaluación holística y totalizadora de tal manera que se pueda establecer el significado de las acciones y actividades que se desarrollan dentro del programa. Este enfoque no hace referencia a factores causales ni a variables particulares que pudieran ser sometidos a tratamientos estadísticos.

Desde la perspectiva epistemológica, los enfoques globales son en su mayoría de corte naturalista, subjetivista e idealista y emplean preferentemente métodos cualitativos de investigación.

En la práctica de evaluación de programas, con frecuencia se combinan enfoques y procedimientos de uno o más modelos y, con base en ellos, los investigadores definen su diseño específico de evaluación. Lo anterior, tal como lo plantea Briones (1985), no sugiere un eclecticismo arbitrario, sino un esfuerzo teórico y metodológico en la búsqueda de una propuesta que responda a necesidades y requerimientos del programa que se va a evaluar, así como a las características y limitaciones que existen en la realidad.

### **III. ORIENTACIONES PREDOMINANTES EN LA INVESTIGACIÓN EVALUATIVA**

#### **A. EVALUACIÓN REFERIDA A OBJETIVOS**

La evaluación, referida a objetivos tiene sus orígenes en educación, donde Ralph W. Tyler desarrolló el primer método sistemático de evaluación en el cual procuró hacer énfasis en una amplia gama de objetos por evaluar como los currículos y las ayudas didácticas, además de la evaluación centrada en el estudiante, para lo cual destacó la necesidad de establecer, clasificar y definir los objetivos en términos de rendimiento, como etapa inicial o punto de partida de todo estudio evaluativo. La evaluación se convirtió en un proceso para determinar la congruencia entre los objetivos y las operaciones.

Tyler definió así la evaluación:

“El proceso de evaluación es esencialmente el proceso de determinar hasta qué punto los objetivos educativos han sido actualmente alcanzados mediante los programas de currículos y enseñanza. De cualquier manera, desde el momento en que los objetivos educativos son esencialmente cambios producidos en los seres humanos, es decir, ya que los objetivos alcanzados producen ciertos cambios deseables en los modelos de



comportamiento del estudiante, entonces, la evaluación es el proceso que determina el nivel alcanzado". (Tyler, 1950:69)

El concepto tyleriano de evaluación ha sido, a menudo, aplicado con una rigidez y estrechez de miras que Tyler nunca propugnó. Mientras proponía estrategias y técnicas prácticas para la recolección y análisis de la información requerida para el proceso de evaluación, la atención terminó enfatizando en los datos, que fueron considerados válidos sólo cuando el programa había concluido.

A pesar de los cuestionamientos a la evaluación asumida como un proceso terminal y del limitado alcance asignado hoy a los objetivos, el trabajo de Tyler permanece como uno de los grandes hitos en la historia de la evaluación concebida como ciencia. Más aun, destacados estudiosos de la evaluación educacional, como Michael y Metfessel, Provus y Hammond, han llevado más allá el trabajo de Tyler, sin alejarse de su filosofía básica y de su tecnología.

## **1. MODELO DE EVALUACIÓN DE RALPH TYLER**

### **a. ENFOQUE: ANALÍTICO RACIONALISTA**

### **b. CONCEPTO DE EVALUACIÓN**

Proceso para determinar hasta qué punto los objetivos propuestos por el programa han sido actualmente alcanzados, mediante los programas definidos para tal fin.

### **c. PROPÓSITOS**

Proporcionar al cliente o a quienes toman decisiones la información requerida, basada necesariamente en la coincidencia entre los objetivos del programa y sus resultados reales.

### **d. CARACTERÍSTICAS**

La ventaja de la definición de evaluación basada en la congruencia es que procede de una base lógica organizada a partir del desarrollo de un programa. Tyler consideraba que la evaluación debía ser una etapa esencial de este desarrollo, además consideraba que

debían existir una serie de etapas lógicas que afianzaran el proceso de evaluación real. Estas etapas consolidan un método que mida el alcance del cambio. Antes de que esto suceda deben ser determinadas con claridad y definidas las metas y objetivos. Dicho en términos de comportamiento, deben comprobarse los objetivos después de que se hayan establecido las condiciones apropiadas para su comprobación.

e. *Metodología*

De acuerdo con Tyler, el procedimiento para evaluar un programa es el siguiente:

1. Establecer las metas y objetivos.
2. Ordenar los objetivos en amplias clasificaciones.
3. Definir los objetivos en términos de comportamiento.
4. Establecer situaciones y condiciones según las cuales puede ser demostrada la consecuencia de los objetivos.
5. Explicar los propósitos de la estrategia al personal más importante, en las situaciones más adecuadas.
6. Escoger o desarrollar las adecuadas medidas técnicas.
7. Recopilar los datos de trabajo.
8. Comparar los datos con los objetivos de comportamiento.

## **2. MODELO DE EVALUACIÓN DE METFESSEL Y MICHAEL**

- a. ENFOQUE: ANALÍTICO RACIONALISTA
- b. CONCEPTO DE EVALUACIÓN

Por ser una ampliación del método tyleriano el concepto de evaluación es congruente con éste:

Proceso para determinar hasta qué punto los objetivos propuestos por el programa han sido alcanzados, mediante los programas definidos para tal fin.

c. PROPÓSITOS

Hacer más perfectas las decisiones a partir de la valoración de los resultados y ayudar al personal básico de las escuelas (profesores, administradores y especialistas) a evaluar la consecución de los objetivos en los programas escolares.

d. CARACTERÍSTICAS

Metfessel y Michael (1967) desarrollaron una lista de los múltiples criterios de valoración que podrían ser utilizados para la evaluación de programas. Especificando toda la gama de posibles criterios para la valoración de los resultados, estos autores van bastante más allá que Tyler. Sin embargo, su paradigma evaluativo es básicamente similar al método tyleriano.

Este modelo proporciona detalles acerca de los individuos que deberían ser los principales responsables de la evaluación en cada una de las etapas del proceso.

Propone las valoraciones basadas en múltiples criterios (etapa 4 de la metodología propuesta), lo que representa una excelente contribución a la creación de instrumentos, porque incluye una lista de las diversas fuentes y tipos de valoraciones necesarias para alcanzar una comprensión acabada y completa de los diferentes aspectos de un programa, ampliando el alcance de los juicios sobre un programa.

e. METODOLOGÍA

El modelo propone ocho etapas principales del proceso evaluativo para seguir:

- 1) Implicar a todos los miembros de la comunidad escolar como participantes o ayudantes de la evaluación.

- 2) Hacer una lista coherente de todas las metas y objetivos específicos, clasificados en un orden jerárquico, desde los resultados más específicamente deseados hasta los más generales.
  - a) Poniendo primero las metas que más ampliamente abarcan las intenciones teóricas del programa.
  - b) Luego, los objetivos específicos en términos operacionales, para permitir una valoración objetiva siempre que sea posible, y
  - c) Desarrollando finalmente criterios de enjuiciamiento que permitan la definición de los resultados más significantes y relevantes, el establecimiento de prioridades realistas en términos de necesidad social, de la facilidad del alumno, de la retroalimentación alumno-maestro y de la disponibilidad de personal y de recursos materiales.
- 3) Traducir los objetivos específicos de rendimiento a una forma inteligible que pueda ponerse en práctica para facilitar el aprendizaje en el ambiente escolar.
- 4) Seleccionar o crear una variedad de instrumentos que suministren valores de los que se pueda colegir la efectividad de los programas a la hora de conseguir los objetivos que se habían propuesto.
- 5) Llevar a cabo observaciones periódicas, a través de la utilización de instrumentos variados, para calibrar el alcance del cambio en el rendimiento que resulta válido respecto a los objetivos seleccionados.
- 6) Analizar los datos proporcionados por la valoración del cambio a través de la utilización de métodos estadísticos apropiados.
- 7) Interpretar los datos relativos a los objetivos específicos según ciertas normas de enjuiciamiento y ciertos valores que se consideren apropiados para los niveles a los que se desee llegar en el trabajo. De esta manera, se podrá llegar a conclusiones que proporcionen información acerca de la dirección de la evolución, el progreso de los estudiantes y de la efectividad de todo el programa.

- 8) Hacer recomendaciones que proporcionen una base para futuros perfeccionamientos, modificaciones y revisiones de toda la gama de metas y objetivos específicos con el propósito de mejorar el programa; proporcionar continua información, basada en recomendaciones, a todos los individuos implicados en el programa y prepararse para cuando vuelva a recomenzar el ciclo del proceso evaluativo, una vez que los resultados ya hayan sido presentados a las audiencias correspondientes.

### **3. MODELO DE EVALUACIÓN DE EDWARD A. SUCHMAN: EL MÉTODO CIENTÍFICO DE LA EVALUACIÓN**

Edward Suchman, científico investigador que creía firmemente en la necesidad de basar las conclusiones en evidencias científicas, consideraba que la evaluación debía disponer de la lógica del método científico.

La contribución de Schuman al campo de la evaluación lo constituye su obra *Evaluative Research: Principles and Practice in Public Service and Social Action Programs* (1967). En ella deja claro su punto de vista de que un evaluador debe utilizar todo tipo de técnicas investigativas que sean útiles y apropiadas a las circunstancias y necesidades de un estudio evaluativo determinado.

Sus estudios evaluativos, especialmente referidos a la salud pública, le hicieron reconocer que la evaluación científica tiene algunas limitaciones prácticas y aunque consideraba que el estudio ideal debería estar enmarcado en los modelos experimentales clásicos, afirmaba que en la realidad los proyectos investigativos de evaluación utilizan comúnmente alguna variación o adaptación de dicho modelo.

Siguiendo el esquema hasta ahora propuesto, presentamos las características de esta orientación .

a. ENFOQUE: ANALÍTICO-RACIONALISTA

b. CONCEPTO DE EVALUACIÓN

La evaluación es un proceso científico, por lo tanto, los mismos procedimientos que se utilizan para descubrir el conocimiento podrían utilizarse para evaluar el nivel de éxito

obtenido en la aplicación de este conocimiento. La investigación evaluativa es investigación aplicada y su propósito es determinar hasta qué punto un programa específico ha conseguido el resultado esperado. Los resultados siempre serán utilizados por el administrador para tomar decisiones del futuro del programa.

#### c. PROPÓSITOS

- 1) Describir si los objetivos han sido alcanzados y de qué manera.
- 2) Determinar las razones de cada uno de los éxitos y fracasos.
- 3) Descubrir los principios que subyacen en un programa que ha tenido éxito.
- 4) Dirigir el curso de los experimentos mediante técnicas que aumenten su efectividad.
- 5) Sentar las bases de una futura investigación sobre las razones del relativo éxito de técnicas alternativas.
- 6) Redefinir los medios que hay que utilizar para alcanzar los objetivos, así como las submetas, a la luz de los descubrimientos de la investigación.

#### d. CARACTERÍSTICAS

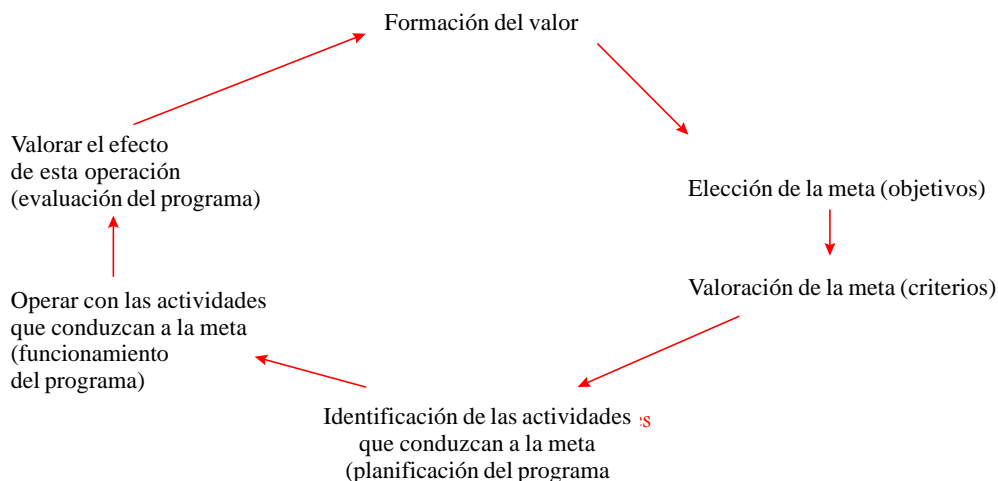
Una precondition de cualquier estudio evaluativo, según Suchman, es la presencia de alguna actividad cuyos objetivos tengan algún tipo de valor. Este autor define el valor como "cualquier aspecto de una situación, actividad u objeto que tengan un interés particular, como ser bueno, malo, deseable, indeseable, etc. Los valores, por lo tanto, pueden estar constituidos como modos de organizar la actividad humana basados en principios que determinan tanto las metas como el perfeccionamiento de los programas, así como el medio de alcanzar esas metas.

#### e. METODOLOGÍA

Suchman consideraba que el proceso evaluativo parte de y regresa a la formación de los valores, siguiendo los siguientes pasos:

- 1) Identificar un valor concreto (ya sea explícito o implícito).
- 2) Definir la meta: selección entre todas las metas alternativas posibles.
- 3) Seleccionar los criterios para valorar la consecución de la meta.
- 4) Identificar las actividades que conduzcan a la meta: planificación del programa.
- 5) Operar con las actividades que conduzcan a la meta: funcionamiento del programa.
- 6) Valorar el efecto de esta operación: evaluación del programa.
- 7) Formación del valor: finalmente, basándose en esta evaluación, se emite un juicio acerca de si esta actividad dirigida hacia la meta ha resultado útil.

Fig. 1. El proceso evaluativo.



## b. LA EVALUACIÓN DE EXPERTOS

La evaluación de expertos, o también denominado método del juicio profesional, asume el proceso de evaluación igual al juicio profesional de un experto o de un grupo de expertos.

El experto o grupo de expertos es generalmente un profesional o grupo de profesionales de reconocido prestigio en el campo de la evaluación y en el objeto por evaluar. Se les solicita que estudien el sistema y emitan un juicio acerca de cómo está funcionando.

Las ventajas de conformar un equipo de expertos que estudien la marcha de la institución o programa y rindan luego un informe sobre lo que creen está bien o mal, lo que puede mejorarse y hacia dónde debe enrutarse el desarrollo de lo evaluado, residen en la facilidad de implementar tal tipo de evaluación y en plantear la posibilidad de tener en cuenta todo tipo de variables.

Los teóricos de la evaluación formulan ventajas y desventajas a este método. Analicemos algunas de éstas.

#### Ventajas:

- Es fácil de implementar, no requiere de la construcción de instrumentos complejos y el tiempo de la evaluación es más corto.
- Toma potencialmente en consideración todas las variables.
- Pondera la experiencia de los expertos, la tiene en cuenta y le da posibilidad de que desempeñe un papel importante en la evaluación.
- Utiliza toda la habilidad y experiencia disponibles.
- No media tiempo muerto entre la recopilación de datos y el análisis de los mismos.

#### Desventajas:

- Le falta refinamiento. Dictados principalmente por la oportunidad.
- Su fiabilidad y objetividad son cuestionables.
- Tanto los datos como los criterios son ambiguos y subjetivos.
- No se puede comprobar su validez.



Limitaciones en cuanto a la generalización de los resultados de la investigación.

### c. EVALUACIÓN ORIENTADA A LA ADMINISTRACIÓN

La evaluación orientada a la administración, también denominada enfoque de la decisión orientada, subraya que la evaluación debe ser utilizada tanto para el perfeccionamiento de un programa, como para enjuiciar su valor. Las decisiones que se van a tomar, ya sea por un administrador, un consejo político, un personal de programación o un electorado, proporcionan organizadores previos para los estudios de orientación de la decisión. Básicamente, la intención de estos estudios es proporcionar conocimientos y bases valorativas para tomar y justificar decisiones.

Los principales interrogantes que plantea este tipo de estudios son, entre otros:

- ¿Qué es lo que hay que conseguir?
- ¿Cuál debe ser la planificación de una empresa determinada?
- ¿Cómo debe llevarse a cabo el plan?
- ¿Cuándo y cómo debe ser modificada la empresa?

Crombach (1963) fue el primero que introdujo a los evaluadores en la idea de que la evaluación debía ser reorientada y pasar a ser una ayuda para que los administradores tomaran decisiones más acertadas acerca de cómo orientar los programas o instituciones. Más tarde Stufflebeam (1966, 1967) introdujo una conceptualización de la evaluación que se basaba en la idea de que ésta debía ayudar a los educadores a tomar y justificar decisiones que pudieran satisfacer más las necesidades de los estudiantes. Posteriormente, otros autores como Alkin (1969), Reinhard (1972), Taylor (1974), Webster (1975) y Guba (1978) han contribuido al desarrollo del concepto de la evaluación centrada en las decisiones y orientada a la administración.

La principal ventaja de esta orientación es que anima a los grupos de profesionales, así como a las organizaciones, a utilizar la evaluación continua y sistemáticamente para planificar y proporcionar servicios que satisfagan las necesidades de sus clientes. También

proporciona una base lógica que ayuda a los profesionales a responder por las decisiones tomadas en el curso de un servicio.

## **1. MODELO DE STUFFLEBEAM-MODELO C.I.P.P. (CONTEXTO-INSUMOS-PROCESOS-PRODUCTOS)**

### **a. ENFOQUE: ANALÍTICO-RACIONALISTA**

### **b. CONCEPTO DE EVALUACIÓN**

La evaluación es el proceso de identificar, obtener y proporcionar información útil y descriptiva acerca del valor y el mérito de las metas, la planificación, la realización y el impacto de un objeto determinado, con el fin de servir de guía para la toma de decisiones, solucionar problemas de responsabilidad y promover la comprensión de los fenómenos implicados.

### **c. PROPÓSITO**

Proporcionar información útil para tomar decisiones con respecto a un programa, ya sean estas de planeación como resultado de la evaluación de contexto; de estructuración como resultado de la evaluación de insumos; de implementación como resultado de la evaluación de proceso o de reciclaje como resultado de la evaluación de producto.

### **d. CARACTERÍSTICAS**

Establecer cuatro tipos de evaluación:

*Evaluación de contexto:* busca definir el contexto institucional, identificar la población objeto del estudio y valorar sus necesidades, identificar las oportunidades de satisfacer las necesidades, diagnosticar los problemas que subyacen en las necesidades y juzgar si los objetivos propuestos son lo suficientemente coherentes con las necesidades valoradas.

*Evaluación de insumos:* busca identificar y valorar la capacidad del sistema, las estrategias del programa y sus alternativas, la planificación de procedimientos para llevar a cabo las estrategias, los presupuestos y los programas.

*Evaluación de procesos:* busca identificar o pronosticar, durante el proceso, los defectos de la planificación del procedimiento o de su realización, proporcionar información para las definiciones preprogramadas y descubrir y juzgar las actividades y aspectos del procedimiento.

*Evaluación del producto:* busca recopilar descripciones y juicios acerca de los resultados y relacionarlos con los objetivos y la información proporcionada por el contexto, los insumos y por los procesos e interpretar su valor y su mérito.

Propone y utiliza la evaluación formativa y sumativa para cada uno de estos tipos de evaluación, así como el empleo de una amplia gama de métodos y técnicas de recolección de información como el análisis de sistemas, la inspección (observación directa), la revisión de documentos, las audiciones, las entrevistas, los test diagnósticos y la técnica Delphi entre otras.

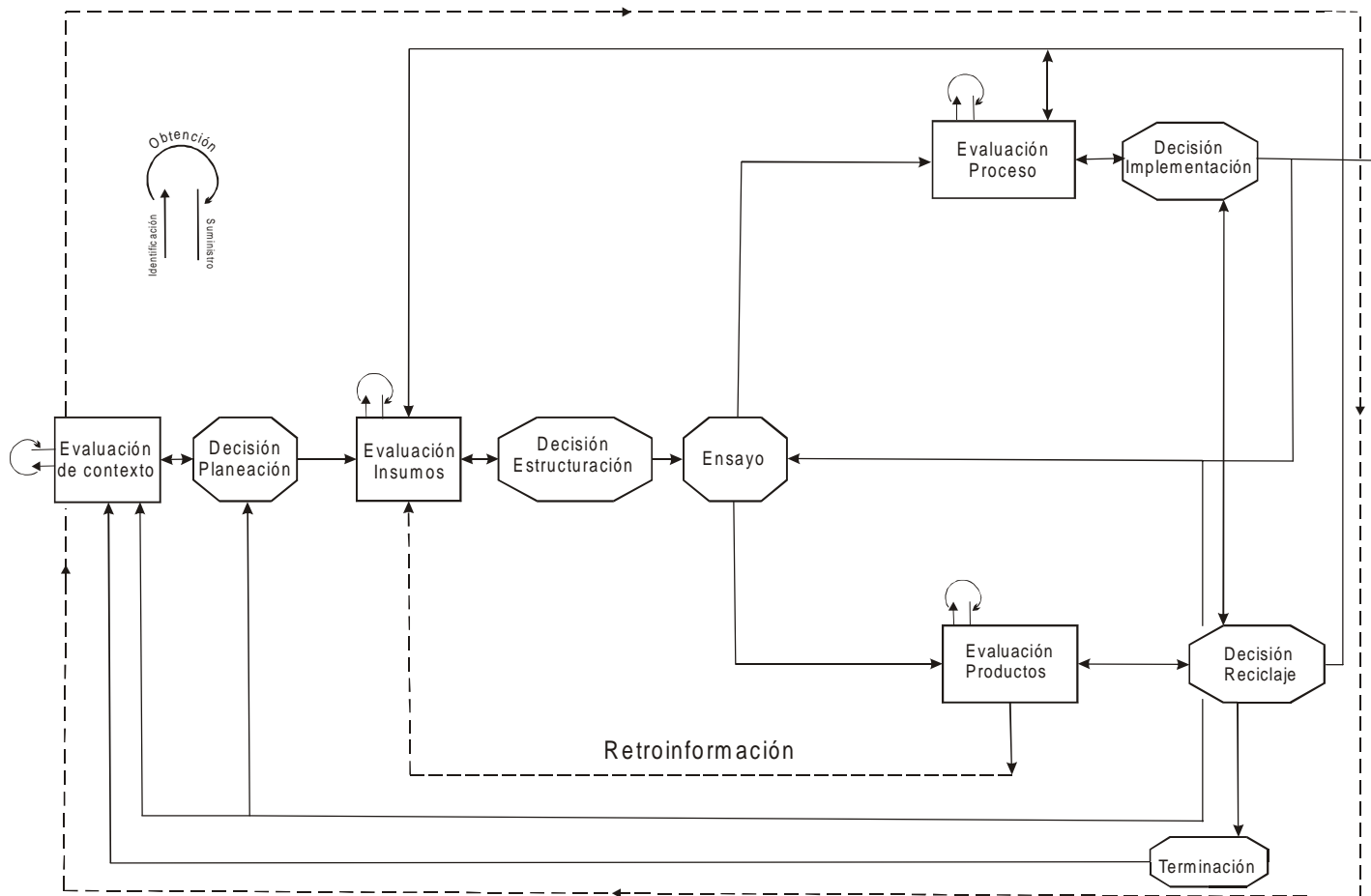
#### e. METODOLOGÍA

Plantea cuatro niveles o etapas:

- 1) **Delineación:** en la que se definen los objetivos de la evaluación, se formulan preguntas, se delimitan las decisiones por tomar se identifican impedimentos.
- 2) **Obtención de información:** se diseñan o adaptan instrumentos, se determinan las muestras, se aplican los instrumentos y se realiza el procesamiento estadístico de los resultados.
- 3) **Análisis de la información:** se elabora un diseño de análisis que garantice la mayor efectividad en la interpretación de los datos obtenidos y se procede al análisis de la información.
- 4) **Aplicación de la información:** en la que se comunica la información a la audiencia y se le asesora para la utilización y aplicación de dicha información.

A continuación se presenta el modelo globalizador simplificado (C.I.P.P) de Stufflebeam, gráfica 1.

GRÁFICA 1: MODELO GLOBALIZADOR SIMPLIFICADO



Tomado de: Stufflebeam y Col. Educational Evaluation. Decision Making. 1972.

## 2. MODELO DE REFERENTES ESPECÍFICOS DE BRIONES

### a. ENFOQUE: RACIONALISTA-ANALÍTICO

### b. CONCEPTOS DE EVALUACIÓN

Investigación aplicada que consiste en describir, medir, comparar y juzgar las propiedades que definen el estado de los objetos de evaluación que se seleccionan de un programa.

### c. PROPÓSITOS

Proporcionar información útil para tomar decisiones con respecto a un programa, bien sea referidas a sus objetivos, a su estructura, a su funcionamiento, a sus resultados o a los usuarios.

### d. CARACTERÍSTICAS

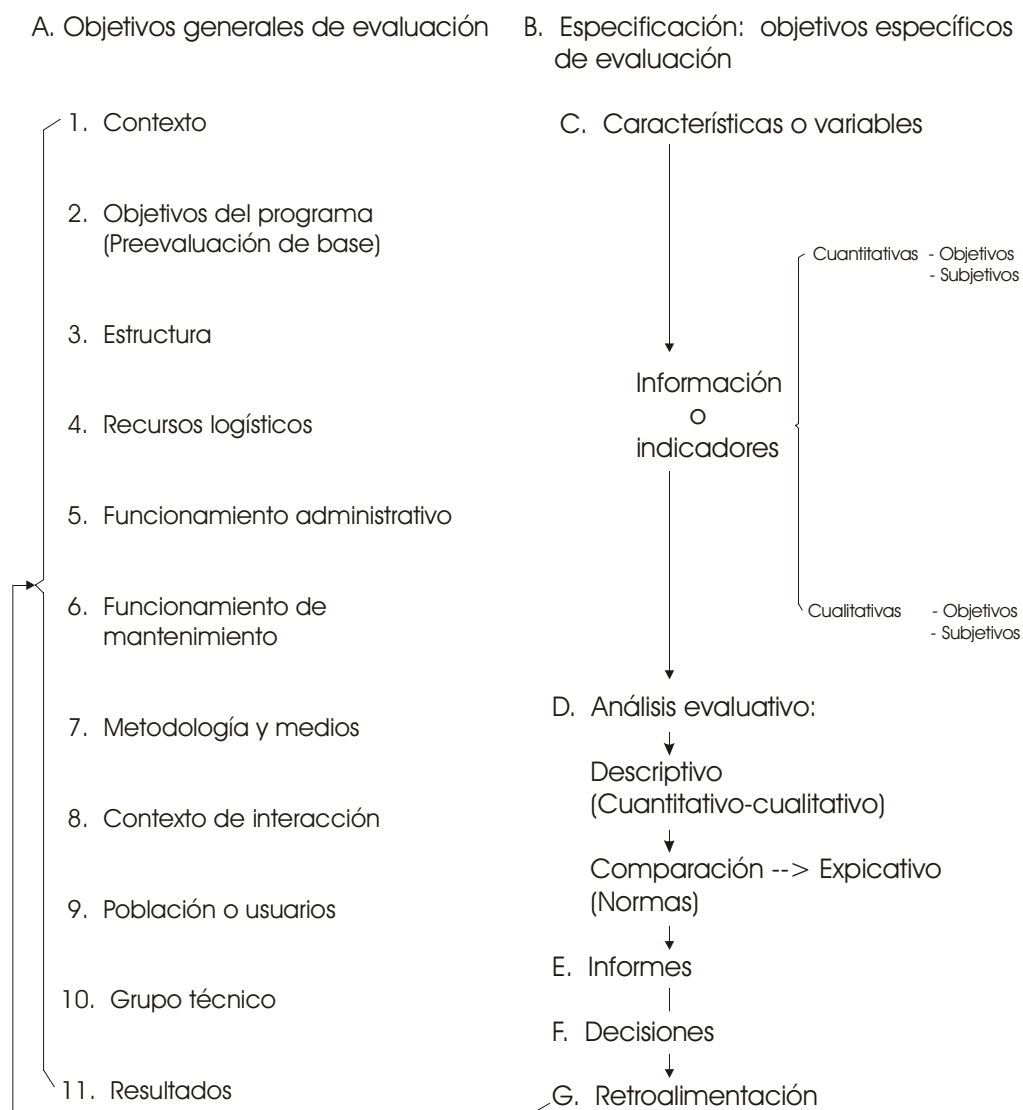
Plantea la necesidad de tener claridad respecto a cuáles son los referentes específicos de la evaluación que reciben el nombre de objetos de evaluación, y que se clasifican en generales y específicos. Tales objetos poseen propiedades o atributos que constituyen las variables de evaluación, se estudian con base en criterios definidos previamente, llamados referentes de comparación. Las variables se definen en términos de indicadores, que pueden ser de naturaleza objetiva o subjetiva, cuantitativos o cualitativos. El modelo distingue entre objetivos finales e instrumentales, entre resultados previstos-no previstos, logros-consecuencias y entre procesos de mantenimiento y desarrollo del programa y la metodología.

### e. METODOLOGÍA

Utiliza, tanto en la recolección de información como en el procesamiento de datos, técnicas cuantitativas y cualitativas, teniendo como criterios, para ello, los objetivos de la evaluación, los objetos y variables de la evaluación y el diseño metodológico básico (experimental o no experimental).

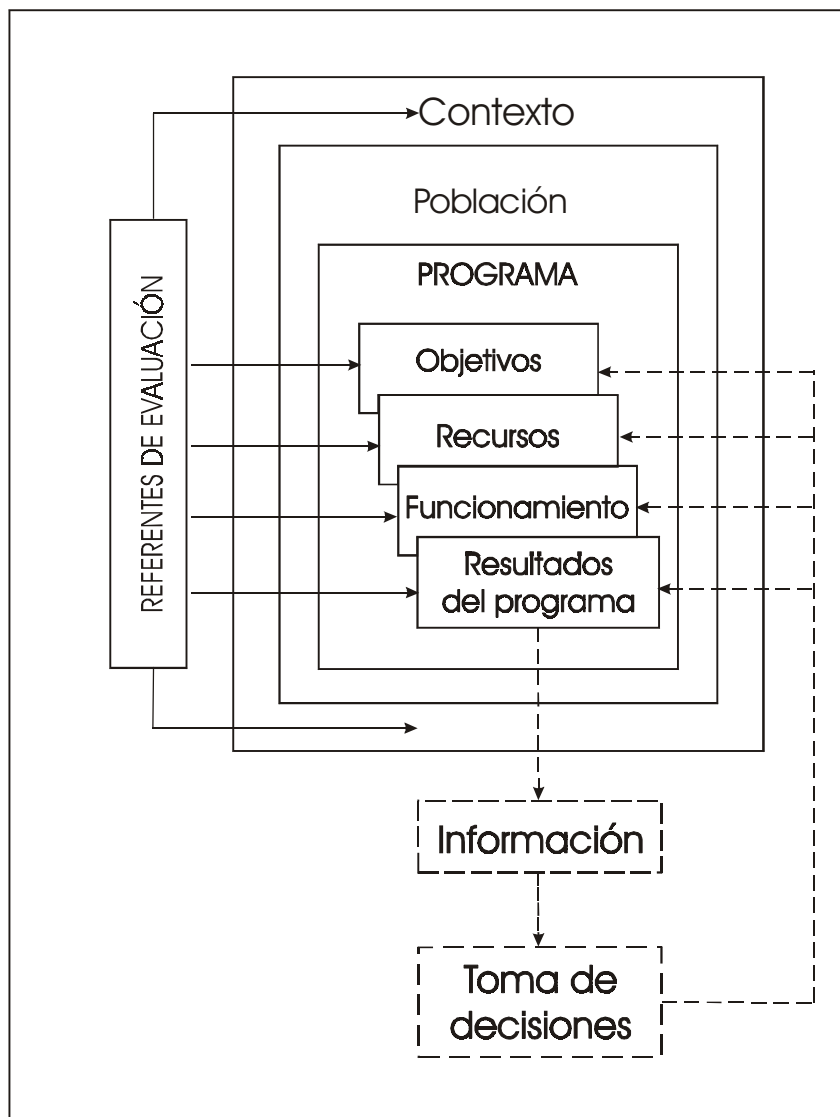
A continuación se presenta el modelo general y el modelo operativo de Briones, gráficas 2, 3 y 4.

## GRÁFICA 2: MODELO DE REFERENTES ESPECÍFICOS



Tomado de Briones, G. Evaluación de programas sociales. Santiago, 1985.

GRÁFICA 3: MODELO DE REFERENTES ESPECÍFICOS DE BRIONES  
-ESTRUCTURA SUSTANTIVA-



Tomado de Briones, G. Evaluación educacional. Vol. 4, Bogotá, 1985.

## **D. EVALUACIÓN POR OPOSICIÓN O EVALUACIÓN CONTRAPUESTA**

La evaluación por oposición o evaluación contrapuesta aparece en la literatura de la evaluación a principios de la década de 1970. El método de evaluación contrapuesta o por oposición pretende examinar a fondo e iluminar los aspectos más importantes del programa que se evalúa. Los aspectos explorados pueden ir más allá de los objetivos originalmente propuestos para el programa.

El modelo de evaluación por oposición sigue un proceso dialéctico en el cual dos grupos de evaluadores investigan los aspectos positivos y los aspectos negativos de un programa con el fin de clarificar honesta y abiertamente sus principales problemas con la intervención de un juez o grupo de jueces, de allí que también se denomine evaluación JUDICIAL.

Este método aporta a la evaluación las técnicas de un tribunal, donde no se trata de presentar una decisión de tipo culpable o inocente, sino donde la esencia del concepto de juicio es el establecimiento de un acuerdo en el contexto de un desacuerdo con miras a perfeccionar la toma de decisiones.

En un contexto legal, los procedimientos de contraposición se desarrollan como una audiencia administrativa destinada a juzgar los méritos de un caso en el que están implicadas dos partes opuestas.

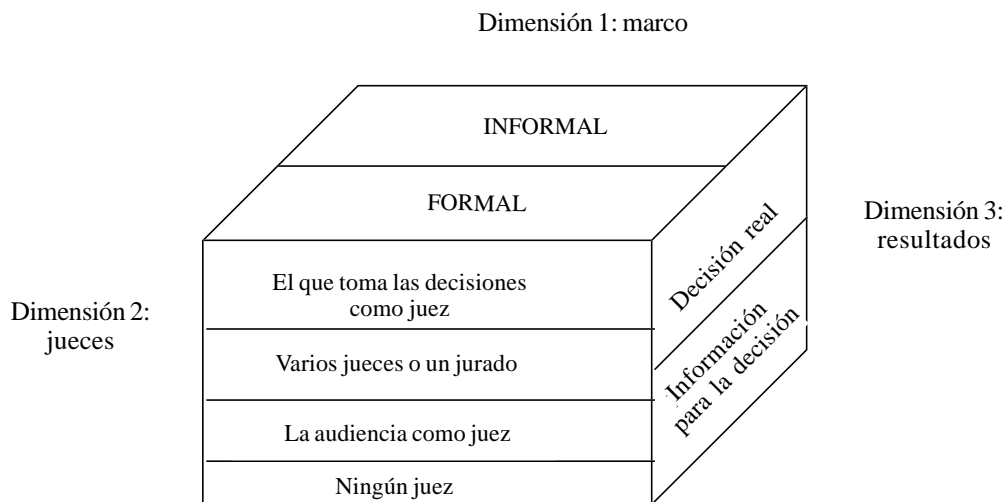
Owens (1973: 296-297) ha identificado las siguientes características de los procedimientos de contraposición como las adecuadas para la toma de decisiones:

- Las reglas establecidas para manejar los procesos de contraposición son totalmente flexibles.
- Las complejas reglas de la evidencia son sustituidas por una evaluación libre de la evidencia realizada por el director de la audiencia.
- Puede pedirse a ambas partes, antes del juicio, que informen al director de la audiencia sobre todos los hechos relevantes, de las pruebas y de los nombres de los testigos.
- Debe proporcionarse una copia de los cargos al director de la audiencia y al defensor, antes de que empiece el juicio, teniendo el defensor la opción de admitir ciertos cargos por anticipado y recusar otros.



- Los testigos declararan libremente y podrán ser interrogados.
- A menudo, los expertos son llamados a testificar, incluso antes del juicio.
- Las entrevistas previas al juicio, entre el director de la audiencia y las dos partes, suelen convertir el juicio más en una búsqueda de hechos relevantes que en una batalla de talentos.
- Además de las dos partes implicadas, puede permitirse la participación de otros grupos interesados.

El modelo de evaluación por oposición o evaluación contrapuesta puede funcionar conjuntamente con otros modelos de evaluación, o funcionar conjuntamente con otros modelos de evaluación.



Cubo de Owens: *Aplicaciones y dimensiones de los procedimientos contrapuestos*. Tomado de Stufflebeam y Shinkfield (1995).

## 1. MODELO JUDICIAL DE WOLF

### a. ENFOQUE: GLOBAL NATURALISTA

### b. CONCEPTO DE EVALUACIÓN

Audiencia administrativa destinada a juzgar los méritos de un caso en el que están implicadas dos partes opuestas. Permite que un amplio espectro de personas interesadas en el programa expresen sus puntos de vista. Los juicios no se basan en las metas iniciales del programa, sino más bien en la combinación entre éstas y los problemas más importantes planteados por quienes participan en las audiencias.

### c. PROPÓSITOS

Proporcionar a quienes toman las decisiones **evidencias** sólidas para actuar en consecuencia. Lo más importante es perfeccionar la toma de decisiones.

### d. CARACTERÍSTICAS

El modelo adapta y modifica conceptos tanto de los juicios con jurado como de las audiencias administrativas. La metodología legal (judicial) ofrece un útil sistema de procedimientos destinados a valorar las pruebas y a producir deducciones alternativas a partir del análisis de los datos, antes de emitir juicios. Los evaluadores pueden centrarse en un grupo de problemas, apoyarse en los testimonios de diferentes grupos y personas, utilizando dos equipos evaluativos, lo que permite el desarrollo de un punto de vista equilibrado y estructurar las deliberaciones de quienes toman las decisiones.

### e. METODOLOGÍA

Wolf ha definido cuatro etapas para la realización del modelo judicial, a saber:

- 1) *La etapa que propone los problemas:* inicialmente, se identifica una amplia gama de problemas relativos a la evaluación del programa, desde el punto de vista de todo el personal implicado que resulte posible. Estos problemas deben coincidir con los objetivos iniciales del programa. A través de entrevistas

y otros medios, deben identificarse los problemas que se consideren más importantes.

- 2) *La etapa que selecciona los problemas:* durante esta etapa, se reduce el número de problemas hasta dejarlo en una cantidad apta para una audiencia. El personal del programa y otras personas relacionadas con él pueden ayudar a establecer un orden prioritario de problemas. Un grupo especial de investigación, formado por representantes de los grupos más importantes del programa, decide luego la importancia de cada problema, hace las modificaciones necesarias y lo pone todo por escrito.
- 3) *La etapa de preparación de los argumentos:* los dos equipos evaluativos preparan entonces los argumentos formales. Se desarrollan, a partir de cada problema, los aspectos específicos del enfrentamiento. Los argumentos desarrollados proceden a menudo de los datos recopilados acerca de las virtudes y defectos del programa. Las afirmaciones que apoya un aspecto u otro del problema pueden proceder de testigos implicados en el programa o afectados por él. Además, los informes evaluativos más importantes (por ejemplo, acerca de las primeras evaluaciones formativas), los documentos y los informes formales son estudiados para su posible utilización en los argumentos finales.
- 4) *La etapa de la audiencia:* esta etapa tiene dos partes. En la sesión previa a la audiencia, ambos equipos examinan sus principales argumentos y, junto con el presidente de la audiencia, desarrollan las reglas y procedimientos de la audiencia. Aspectos como el número de testigos que deben ser convocados, el alcance de los interrogatorios y los criterios para determinar la admisibilidad de la evidencia (por ejemplo, la relevancia) deben surgir de un acuerdo mutuo. Finalmente, se hace un borrador de las cuestiones más específicas que sirve como guía para las deliberaciones.

La audiencia propiamente dicha, que puede ocurrir tras meses de preparación a través de las tres primeras etapas, puede constar de más de una sesión, dependiendo de la naturaleza de la evaluación. Tras la presentación de los argumentos y los interrogatorios, el jurado delibera. Si el propio jurado no es el que toma las decisiones, recomienda qué decisiones deben tomarse.

Wolf ha afirmado que el modelo judicial “ofrece un foro para la comunicación y la difusión, tanto como para la evaluación”. Ciertamente, el proceso puede clarificar las alternativas y las consecuencias de una decisión, que casi nunca pueden conocerse antes de la audiencia.

## **E. EVALUACIÓN ORIENTADA AL CLIENTE O CENTRADA EN EL CONSUMIDOR**

Los exigentes requisitos federales para la evaluación que se impusieron en la educación americana en la década de 1970 estimularon el desarrollo de algunos nuevos métodos evaluativos, entre los cuales aparece el método evaluativo centrado en el cliente.

Este método plantea que el evaluador debe trabajar con y para ayudar a aquellas personas que desarrollan e imparten servicios. Potencialmente, contempla al evaluador al servicio de una amplia gama de clientes tales como administradores, funcionarios, contribuyentes, legisladores, patrocinadores financieros, líderes comunitarios y el público en general. Los distintos clientes tienen distintas prioridades y el personal del programa debe, a menudo, prestarles mayor atención.

Michael Scriven es un filósofo que ha contribuido ampliamente al desarrollo de la profesión evaluativa. Ha criticado duramente tanto las clásicas como las más recientes conceptualizaciones de la evaluación. Desarrolló su concepción de la evaluación sobre una postura filosófica básica y ha desarrollado conceptos para ayudar a articular y aplicar su concepción de la evaluación.

“El verdadero papel del evaluador –según Scriven–, es el de un sustituto informado del consumidor, provisto de técnicas para obtener una información pertinente y exacta, y con un punto de vista profundamente razonado acerca de la ética y el bien común, el evaluador debe ayudar a los profesionales a proporcionar productos y servicios de alta calidad y gran utilidad para los consumidores. Y lo que es aun más importante, el evaluador debe ayudar a los consumidores a identificar y valorar los bienes y servicios alternativos. Scriven ha sido muy crítico con las ideologías evaluativas que se centran en alcanzar los objetivos de los participantes en lugar de satisfacer las necesidades de los consumidores. Por ello ha propuesto ciertos conceptos y métodos destinados a desplazar la evaluación desde los objetivos hasta las necesidades” (Stufflebeam y Shinkfield, 1995).

## **1. MODELO DE EVALUACIÓN DE MICHAEL SCRIVEN**

### **a. ENFOQUE: ANALÍTICO-RACIONALISTA**

### **b. CONCEPTO DE EVALUACIÓN**

La evaluación es la determinación sistemática y objetiva del valor o el mérito de un objeto, basándose en pruebas acumuladas procedentes de su comparación con otros objetos distintos a la hora de satisfacer las necesidades del consumidor.

### **c. PROPÓSITOS**

Buscar los resultados reales obtenidos por el programa, sin determinar por anticipado sus objetivos para no limitar la identificación de sus efectos posibles, entre ellos, los resultados no anticipados y los efectos imprevistos.

### **d. CARACTERÍSTICAS**

Scriven introduce los conceptos de evaluación formativa y evaluación sumativa. La evaluación formativa se realiza durante el proceso y sirve como retroalimentación y guía para el mejoramiento del programa durante su marcha. La evaluación sumativa es un juicio valorativo sobre lo hecho. Se concentra en la evaluación de los resultados y efectos del programa, aunque estudia los diversos elementos del programa con un enfoque analítico. Busca establecer relaciones de causalidad entre variables dependientes e independientes.

### **e. METODOLOGÍA**

Se siguen los siguientes pasos:

1. Caracterización de la naturaleza del programa.
2. Determinación de los fines de la evaluación.
3. Indicadores de las relaciones que pueden darse entre las variables dependientes e independientes del programa.

4. Verificación amplia de todas las consecuencias del programa.
5. Determinación de los criterios de valor y presupuestos filosóficos del programa.
6. Estimación de los costos.
7. Identificación de acciones alternativas.
8. Identificación de los elementos del programa y sus posibles impactos.
9. Conclusión sobre el valor del programa.

## **F. EVALUACIÓN PARTICIPATIVA**

La evaluación participativa es la forma que asume la investigación participativa, al evaluar programas de acción. Por lo tanto, sus características básicas corresponden a las de dicho tipo de investigación. Sin embargo, podría decirse que lo participativo es más una modalidad y que cualquiera de los enfoques y modelos que hasta ahora se han analizado podría presentar tales características con las limitaciones que las diferentes maneras y niveles de participación plantean.

La evaluación participativa, en su sentido más amplio, puede comprender todas las estrategias en las que los usuarios de un programa (clientes internos y externos) participan activamente en la toma de decisiones y en la ejecución de algunas o todas las fases del proceso investigativo (proceso evaluativo).

“Desde luego no existe ninguna investigación empírica, en la que no haya algún grado de participación de sus actores: desde la voluntad de ser observados hasta responder cuestionarios o someter a entrevistas en profundidad. Sin embargo, en la evaluación participativa, la idea es que las personas que pertenecen al programa tomen en sus manos las tareas que implican un estudio de este tipo, con claras ventajas para ellos en lo que toca a lograr los propósitos buscados: aprendizaje grupal e individual, autorealización, etc. y, para el programa, mejor determinación de los problemas, propuesta de soluciones realistas, mayor compromiso de los usuarios, tanto con los objetivos de la evaluación como con las soluciones” (Briones, 1995: 59-60).

De manera muy sintética, la evaluación participativa es una combinación de INVESTIGACIÓN; EDUCACIÓN-APRENDIZAJE Y ACCIÓN. Ella tiene como objetivo conocer y analizar una realidad (programa o institución) en sus momentos constitutivos:

1. Los procesos y los problemas.
2. La percepción que las personas tienen de ellos.
3. Las experiencias vivenciales dentro de dicha realidad concreta con el fin de emprender acciones tendientes a transformar dicha realidad.

Tal como lo plantea Briones (1995), no existen fórmulas preestablecidas para llevar a cabo evaluaciones participativas. Ellas deberán encontrarse en cada caso de evaluación, para lo cual recomienda formularse, entre otras, las siguientes preguntas:

1. ¿Es posible conciliar los propósitos, objetivos y diseños de investigación, tiempo de realización, etc., entre agencias de financiamiento (privadas o estatales), grupos técnicos y usuarios de un programa?
2. ¿Están todos los miembros de un programa, interesados en participar en la evaluación? Ese interés, de existir, ¿se traducirá en un trabajo real, efectuado con responsabilidad?
3. ¿Disponen las personas interesadas del tiempo suficiente para realizar el estudio?
4. ¿Es la población del programa que ha de evaluarse demasiado grande de tal modo que impide la participación de todos? Si fuera así, ¿quiénes participarán y de qué forma se haría la selección?
5. ¿Es posible formular un calendario de actividades para el grupo de participantes que asegure el logro de resultados y la formulación de recomendaciones en un plazo que se considera adecuado?
6. ¿Tiene el grupo un nivel cultural suficiente que le permita comprender, de manera más o menos semejante, los propósitos y la dinámica de la evaluación? Si hay diferencias marcadas, ¿existe o no el riesgo de que algunos miembros del programa controlen su contenido y su marcha?

7. ¿A qué nivel de información metodológica pueden llegar los participantes?, ¿tiene el conocimiento técnico necesario para recoger, analizar e interpretar datos y para escribir informes?
8. Si las respuestas planteadas en el numeral anterior indican la necesidad de recurrir a uno o más especialistas, ¿cuáles serán las relaciones de estos con los miembros del programa?, ¿cómo se evitará que el estudio se oriente hacia la competencia profesional?
9. Si –como es natural– hay un reparto de las tareas, ¿cómo se logrará mantener una motivación en los diferentes grupos por todas las actividades que se están cumpliendo?, ¿cómo se hará la distribución de esas tareas y de qué modo se supervisarán?
10. ¿Quiénes se harán cargo del costo de la evaluación?
11. ¿Cómo se diseminarán los resultados de la evaluación dentro del grupo y fuera de él?
12. ¿Cómo participarán los miembros del programa en la ejecución de las recomendaciones con las cuales termine la evaluación?

Concluye el autor con la siguiente reflexión:

“Esta breve lista de preguntas permite detectar los diversos problemas que se le plantean a la evaluación participativa. Reconocidos los beneficios que indudablemente tiene este tipo de investigación aplicada, las situaciones concretas del contexto definirán los niveles y modalidades de la participación de las personas; y asimismo los niveles técnicos y las formas de participación. La evaluación participativa, al igual que otros esquemas, es un proyecto de intenciones, muchas de las cuales se quedarán en el camino en diversos momentos de su desarrollo” (Briones, 1995: 61). El líder de esta nueva orientación es Robert Stake que exige un método pluralista, flexible, interactivo, holístico, subjetivo y orientado hacia el servicio.

El juicio racional de la evaluación es una decisión que dice cuánta atención hay que prestar a las normas de cada grupo de referencia (clientes) a la hora de decidir si se emprende alguna acción administrativa. Stake recomendaba a los evaluadores trabajar



continuamente con las distintas audiencias y responder a sus necesidades evaluativas. Ellas son los clientes en el sentido de que son precisamente ellas, las audiencias, las que apoyan, administran o directamente operan con los programas que se están estudiando y buscan el consejo y los planes del evaluador para comprenderlos y mejorarlos.

## **1. MODELO DE EVALUACIÓN DE ROBERT STAKE (EVALUACIÓN SENSITIVA O EVALUACIÓN RESPONDENTE)**

### **a. ENFOQUE: GLOBAL-NATURALISTA**

### **b. CONCEPTO DE EVALUACIÓN**

Evaluación sensitiva o concertada, según la cual ésta debe responder a lo que el cliente realmente necesita. Evaluación orientada a la naturaleza compleja y dinámica de los programas educativos, atendiendo sus diversos aspectos y componentes.

### **c. PROPÓSITOS**

Proporcionar al cliente o a los que adoptan decisiones la información que necesitan, en términos de descripciones y juicios sobre diversos aspectos y componentes del programa.

### **d. CARACTERÍSTICAS**

El modelo parte de la distinción entre antecedentes, transacciones y logros, tanto propuestos como observados, y utiliza los conceptos relacionales claves de contingencia y congruencia. Debe existir contingencia lógica y empírica entre los antecedentes, las transacciones y los logros propuestos y observados, respectivamente, y debe existir congruencia entre lo propuesto y observado.

Los aspectos principales atribuidos a la concepción de Stake son los siguientes (Stufflebeam y Shinkfield, 1995:238).

- Las evaluaciones deben ayudar a las audiencias a observar y mejorar lo que están haciendo (de allí la denominación de evaluación centrada en el cliente).

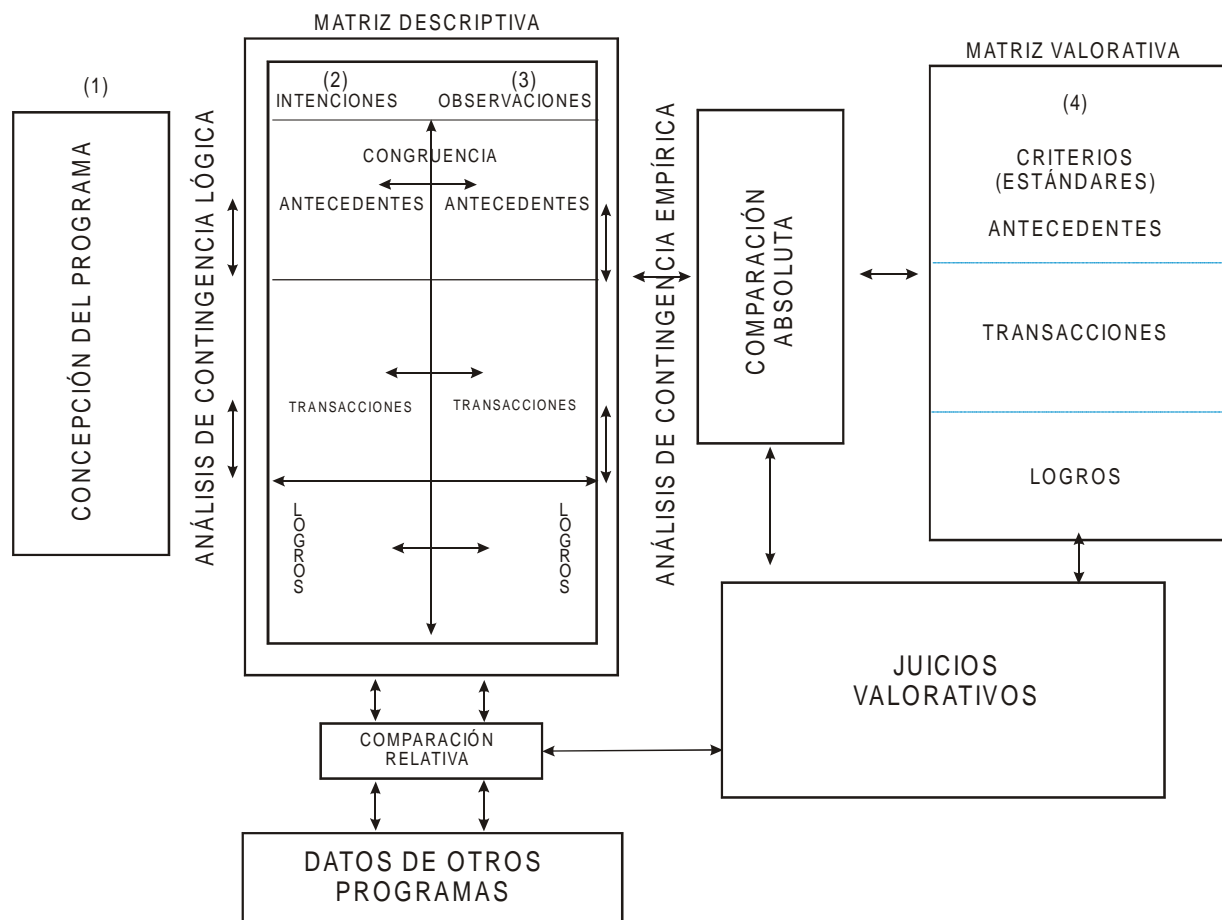
- Los evaluadores deben describir programas con relación tanto a los antecedentes y las operaciones como a los resultados.
- Los efectos secundarios y los logros accidentales deben ser tan estudiados como los resultados buscados.
- Los evaluadores deben evitar la presentación de conclusiones finales resumidas, pero en su lugar deben recopilar, analizar y reflejar los juicios de una amplia gama de gente interesada en el objeto de la evaluación.
- Los experimentos y los test regularizados son a menudo inadecuados o insuficientes para satisfacer los propósitos de una evaluación y frecuentemente hay que sustituirlos o complementarlos con una variedad de métodos incluyendo los subjetivos.

#### e. METODOLOGÍA

Se prepara un plan de observaciones y transacciones que se ejecutan llevando un diario de campo, a partir de lo cual se elaboran narraciones, descripciones, gráficos, etc. Se averigua qué es valioso para los usuarios y se reúnen juicios de diversas personas cuyos puntos de vista difieren. Se presenta la información obtenida a las personas encargadas del programa y a los mismos usuarios para que se pronuncien sobre la precisión e importancia de los hallazgos.

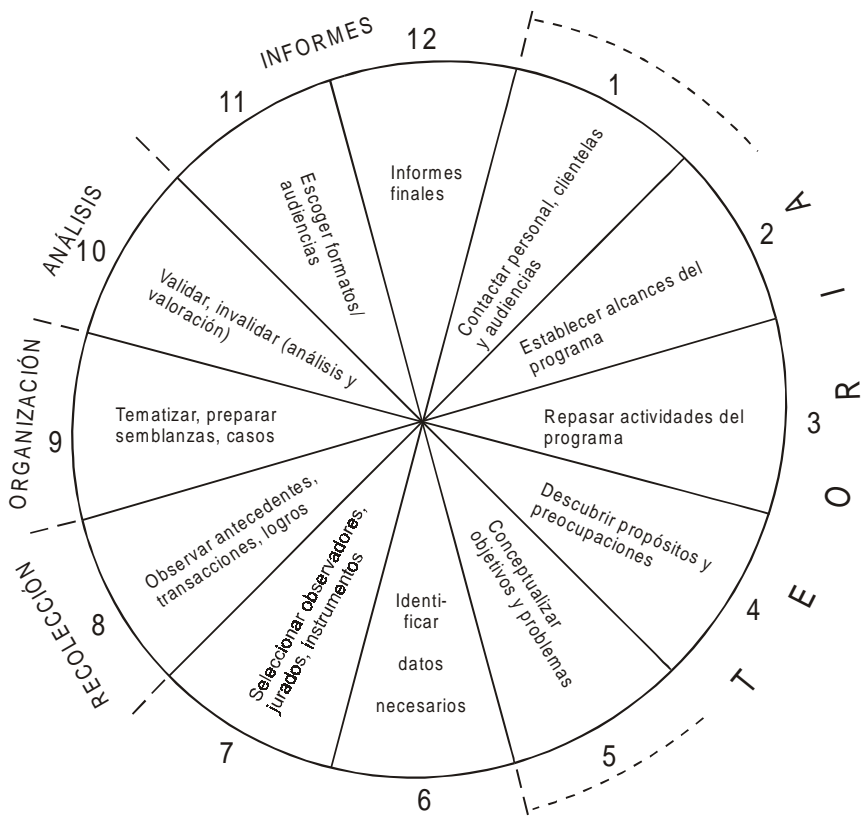
A continuación se presenta el modelo orientador (estructura sustantiva) gráfica 5, y la estructura operativa gráfica No. 6, del modelo de R. Stake.

GRÁFICA 5: MODELO ORIENTADOR DE LA EVALUACIÓN (R.STAKE). ESTRUCTURA SUSTANTIVA



Tomado de: Investigación evaluativa sobre situación actual y prospectiva de la formación abierta y a distancia en el SENA. Restrepo, Correa y Flórez 1989.

GRÁFICA 6: ESTURCTURA OPERATIVA DEL MODELO STAKE



Tomado de Investigación evaluativa sobre situación actual y prospectiva de la formación abierta y a distancia en el SENA. Resprepo, Correa y Flórez, 1989.

## **2. LA EVALUACIÓN ILUMINATIVA**

La evaluación iluminativa no es un método regularizado, sino una estrategia de investigación general. Pretende ser adaptable para descubrir las dimensiones, metas y técnicas de evaluación, y ecléctica para proporcionar al evaluador una serie de tácticas investigativas. Las decisiones acerca de la estrategia que debe ser utilizada en cada caso concreto surgen del programa que debe investigarse. No intenta manipular, controlar o eliminar variables situacionales, sino abarcar toda la complejidad de la escena: “su tarea principal es desenmarañarla, aislar sus características más significativas, esbozar los ciclos causa-efecto y comprender las relaciones entre las creencias y las prácticas y entre los modelos de organización y los modelos de los individuos” (Parlett y Hamilton, 1977).

### **MODELO DE EVALUACIÓN ILUMINATIVA DE PARLETT Y HAMILTON**

#### **a. ENFOQUE: GLOBAL NATURALISTA**

#### **b. CONCEPTO DE EVALUACIÓN**

Evaluación interpretativa, en tanto que recoge las interpretaciones de la realidad que hacen las personas que participan en el programa. Hermenéutica, ya que interpreta los símbolos del lenguaje que utilizan los participantes. E iluminativa, por cuanto busca clarificar los hechos y comprender las situaciones que se presentan en el programa, desde distintas perspectivas.

#### **c. PROPÓSITOS**

Descubrir y documentar lo que supone participar en el programa en cualquier rol, discernir y discutir los rasgos más importantes del programa, sus concomitantes recurrentes y sus procesos críticos: cómo opera, cómo es influenciado por situaciones externas e internas, sus ventajas e inconvenientes, etc.

#### **d. CARACTERÍSTICAS**

Se caracteriza por el uso del paradigma investigativo de la antropología cultural que reconoce como natural el conflicto y la contradicción que se dan en todo grupo humano.

La evaluación misma tiene como destinatarios principales los participantes del programa, quienes, además, participan activamente en ella, puesto que hacerlo les permite reflexionar sobre su situación.

Parlett y Hamilton afirman que existen tres etapas características de una evaluación iluminativa:

- 1) La fase de observación, en la cual se investiga toda la amplia gama de variables que afectan al resultado del programa o la innovación.
- 2) La etapa de investigación, en la cual el interés se desplaza desde el reconocimiento (de la fase de observación) hacia la selección y planteamiento de cuestiones de una manera coherente y ecuánime, con el fin de realizar una lista sistemática y selectiva de los aspectos más importantes del programa en su contexto.
- 3) La etapa de la explicación, en la que los principios generales subyacentes a la organización del programa son expuestos a la luz del día y se delinear los modelos causa-efecto en sus operaciones.

#### e. METODOLOGÍA

Se realiza en varias etapas:

1. Identificación de interlocutores que puedan enseñar problemas y características importantes.
2. Observación y entrevistas etnográficas.
3. Profundización de los aspectos encontrados en la primera etapa, mediante entrevistas en profundidad, historias de vida, etc.
4. Validación de los resultados del programa mediante la triangulación metodológica.
5. Redacción del informe final en un lenguaje accesible a los usuarios del programa.

### **3. MODELO DE EVALUACIÓN FOCALIZADA DE PATTON**

#### **a. ENFOQUE: GLOBAL NATURALISTA**

#### **b. CONCEPTO DE EVALUACIÓN**

Evaluación focalizada, según la cual se seleccionan o focalizan los problemas y aspectos más relevantes del programa, para luego tomar decisiones con respecto al contenido general de la evaluación, sin establecer por anticipado especificaciones de ese contenido ni de los procedimientos por seguir.

#### **c. PROPÓSITOS**

Comprender la dinámica interna del programa tratando de responder preguntas tales como: ¿qué factores son los que caracterizan este programa? ¿Cómo se selecciona a los usuarios y qué hacen una vez están en el programa? ¿Cómo perciben los usuarios el programa?

#### **d. CARACTERÍSTICAS**

Se focaliza la evaluación en los procesos, dando más importancia a cómo se logra un cierto resultado que al producto mismo. Se consideran los resultados, pero desde el supuesto de que estos serán diferentes para los distintos tipos de participantes, es decir, se aboga por la identificación de resultados individualizados.

#### **e. METODOLOGÍA**

Se comienza por identificar las personas que por su posición serán las que tomarán decisiones con respecto al programa y se contacta a los usuarios activos. Con ellos se focalizan los problemas y temas relevantes y luego se seleccionan las técnicas de recolección de información, especialmente de tipo cualitativo (observación, entrevistas, etc.), así como los procedimientos de análisis que se van a seguir. Se caracteriza a los diferentes participantes para el estudio individualizado de los resultados a través de estudios de caso.

#### **4. MODELO DE EVALUACIÓN DE D. FETTERMAN-EVALUACIÓN POR EMPODERAMIENTO**

La evaluación por empoderamiento es una innovadora propuesta de evaluación que ha venido siendo adoptada en la educación superior, el gobierno, distritos de educación pública, corporaciones sin ánimo de lucro y fundaciones, tanto en los Estados Unidos como en el extranjero. Un amplio rango de programas y sectores políticos emplean la evaluación por empoderamiento incluyendo entre estos los programas de prevención del abuso de sustancias, prevención del sida, prevención del crimen, reformas de bienestar social, protección a mujeres maltratadas, desarrollo agrícola y rural, libertad condicional de adultos, prevención del embarazo en la adolescencia, grupos o sociedades tribales para el abuso de sustancias, autodeterminación e individuos con discapacidades, programas doctorales y de escuelas aceleradas entre otros. Una descripción detallada de los programas que hacen uso de la evaluación por empoderamiento aparece en las obras *Empowerment Evaluation: Knowledge and tools for selfassessment and accountability* (Fetterman, Kaftarian and Wandersman, 1996) y *Empowerment Evaluation* (Fetterman, in press). Adicionalmente, esta propuesta ha sido institucionalizada por la American Evaluation Association y es consistente con el espíritu y los estándares desarrollados por el Joint Committee on Standards for Educational Evaluation.

El autor de la propuesta, David Fetterman, decano de la Facultad de Educación de la Universidad de Stanford, reconoce que los programas que han adoptado procesos de evaluación por empoderamiento están tomando responsabilidad de sus propias acciones y sosteniendo por ellos mismos la evaluación, convirtiendo dicho enfoque en una propuesta que ha ganado credibilidad de supervisores y agencias externas comisionadas de evaluación.

La evaluación por empoderamiento es definida como el uso de los conceptos de evaluación, técnicas y hallazgos para fomentar mejoramiento y autodeterminación: ésta es diseñada para ayudar a la gente a ayudarse a sí misma (Fetterman 1996). Éste es un proceso democrático en que las personas se empoderan asimismo, con la asistencia de un experto externo, el cual actúa como un conductor, un facilitador, un entrenador, un consejero, pero no como un evaluador convencional. Los participantes del programa deciden llevar a cabo su propia evaluación, ellos planean e implementan ésta, recogen y analizan sus datos, interpretan sus hallazgos, proponen sus recomendaciones y lecciones aprendidas e implementan sus recomendaciones.



Desde una perspectiva del empoderamiento, el ejercicio de la evaluación no es el punto final del programa, sino un proceso en marcha de mejoramiento en que los autoevaluadores aprenden a continuar midiendo sus progresos hacia la autodeterminación de sus metas y redireccionar sus planes y estrategias acordes con los hallazgos de los procesos de evaluación continua.

Se han reconocido diferentes niveles de acceso a la evaluación por empoderamiento, identificando tres niveles en que los cambios pueden ocurrir:

- a) El nivel individual, en que se da el empoderamiento psicológico (incluye nuevos conocimientos, habilidades, competencias percibidas y expectativas), apropiarse, apoderarse de su lugar.
- b) El nivel intraorganizacional, en que el empoderamiento de la organización hace posible el empoderamiento colectivo de sus miembros: en este nivel, la evaluación por empoderamiento ayuda a los miembros individuales de la organización a integrar sus necesidades, intereses y habilidades con los medios, metas y misión de la organización.
- c) El nivel extraorganizacional en que la relevancia social del sistema puede ser juzgada por el mayor o menor empoderamiento organizacional, esto es, exitoso en influenciar su desarrollo.

La propuesta de evaluación por empoderamiento está basada en los tres principios guías de la filosofía del empoderamiento (Rappaport, 1981):

- a) Todas las personas tienen fortalezas y capacidades, esa capacidad de llegar a ser más competentes.
- b) El fracaso de una persona para desarrollar competencias no es imputable a las definiciones de las personas sino al fracaso del sistema social de proveer o crear oportunidades para que dichas competencias sean desarrolladas o adquiridas.
- c) En situaciones en que existan capacidades que necesitan ser reforzadas o nuevas competencias que necesiten ser aprendidas, éstas son mejor aprendidas a través

de experiencias que conduzcan o guíen a la gente a realizar un autorreconocimiento de la influencia de sus capacidades en los eventos importantes de su vida.

## ETAPAS DE LA EVALUACIÓN POR EMPODERAMIENTO

- a) **ENTRENAMIENTO:** donde los evaluadores enseñan a las personas a conducir su propia evaluación, desmitificando e internalizando el proceso de evaluación. En una evaluación convencional, el proceso evaluativo concluye cuando el evaluador entrega los resultados al administrador. En un proceso de evaluación por empoderamiento, el proceso de evaluación es internalizado dentro de la organización y llega a ser un permanente y continuo ejercicio de autoevaluación para mejorar por su propia cuenta su desempeño.
- b) **FACILITACIÓN:** cuando el evaluador actúa como un entrenador, un facilitador, que ayuda a las personas a conducir su propia evaluación. El evaluador presenta las diferentes alternativas basado en metodologías y técnicas propuestas, explicando beneficios y preocupaciones de cada alternativa, pero los participantes controlan el proceso de toma de decisiones. Ellos deciden qué alternativa metodológica van a emplear, ayudados por el facilitador-evaluador.
- c) **ILUMINACIÓN Y LIBERACIÓN:** muchos participantes experimentan el ejercicio de la evaluación por empoderamiento como una iluminadora y reveladora experiencia que formula una nueva conceptualización de sí mismos. Muchas experiencias demuestran cómo, ayudando a las personas a encontrar el útil camino de la evaluación por ellos mismos, liberados de las tradicionales expectativas y roles, ellos se capacitan para encontrar nuevas oportunidades, redefinir sus roles e identificar y facilitar sus visiones de los recursos existentes desde una nueva luz.

Además de las ventajas antes descritas, la evaluación por empoderamiento facilita la integración de métodos cuantitativos y cualitativos, dado que la actividad desarrollada en el proceso de evaluación provee información cualitativa a los métodos cuantitativos, ayuda a desmitificar el proceso de evaluación repensando las formas de participación y la capacidad de aproximarse a la construcción, mantenimiento, reinención y refinamiento de métodos y técnicas y, por último, pero no menos importante, promueve la institucionalización e internalización de los procesos y métodos de evaluación.

## AUTOEVALUACIÓN N° 3

Responda el siguiente cuestionario:

1. ¿Son totalmente contrapuestos el paradigma cualitativo y el paradigma cuantitativo en la investigación evaluativa o, por el contrario, son compatibles y complementarios?
2. ¿Los paradigmas determinan lógicamente la elección del método de investigación evaluativa?
3. ¿Qué determina, en última instancia, la elección del enfoque investigativo?
4. ¿Cuál de las orientaciones de Investigación evaluativa propuestas en la unidad considerada ha sido la predominante en Colombia y a qué ha obedecido ello?
5. ¿Cuáles de los modelos propuestos considera usted tienen mayor posibilidad de aplicación en la evaluación de instituciones y programas en nuestro medio? Debe sustentarse y argumentarse la respuesta.

## ACTIVIDADES DE EVALUACIÓN

1. A partir de la propuesta por usted presentada para ingresar al programa de Especialización, identifique el MODELO que considere más pertinente para llevar a cabo dicha evaluación y justifique adecuadamente dicha selección.
2. Estructure la propuesta investigativa en mención, con la METODOLOGÍA definida por el modelo seleccionado, sin desarrollarla.

## LECTURA COMPLEMENTARIA

Richardt, Charles, S. y Cook, Thomas O. "Hacia una superación del enfrentamiento entre los métodos cualitativos y los cuantitativos". *Métodos cuantitativos de la investigación evaluativa*. Madrid: Ediciones Morata, S.A., 1986.

## REFERENCIAS BIBLIOGRÁFICAS

Alkin. Citado por Stufflebeam, 1969.

Briones, Guillermo. *Evaluación de programas sociales*. México: Trillas, 1991.

Briones, Guillermo. *Evaluación educacional*. Convenio Andrés Bello, volumen 4, Evaluación institucional. Bogotá: Editorial Guadalupe, 1998.

Conde, Fernando. "Las perspectivas metodológicas cualitativa y cuantitativa en el contexto de la historia de las ciencias". En: *Métodos y técnicas cualitativas de investigación en ciencias sociales*. Madrid: Editorial Síntesis, 1994.

Fetterman, D. M. (1996a). Ethnography in the Virtual Classroom. *Practicing Anthropology*, 18(2):

\_\_\_\_\_. (1995a). Empowerment Evaluation: An Introduction to Theory and Practice. In Fetterman, D. M., Kaftarian, S., and Wandersman, A. (1996). *Empowerment Evaluation: Knowledge and Tools for Self-assessment and Accountability*. Thousand Oaks, CA: Sage.

\_\_\_\_\_. (1995). Kaftarian, S. and Wandersman, A. *Empowerment Evaluation: Knowledge and Tools for Self-assessment and Accountability*. Thousand Oaks, CA: Sage.

\_\_\_\_\_. (1998). Innovation in Evaluation. Cyber Conference. September 14 to 19.

Guba, E. "The Failure of Educational Evaluation". *Educational Technology*, 1978.

Metfessel, N. S y Michael, W. B. A Paradigm Involving Multiple Criterion Measures for the Evaluacion of the Effectiveness of School Programs, *Educational and Psychological Measurement* 27, pp. 931-943, 1967.

Owens, T. R. "Educational Evaluation by Adversary Proceedings in House". E. R. (comp.) *Beyond the Numbers Game*. Londres: Mac Millan Education, 1977.

Patton, M. Q. *Utilization-Focused Evaluation*. Beverly Hills/Londres: Sage Publications.

Provus, M. *Discrepancy Evaluation*. Berkeley: Mc Cutchan, 1971.

Reichardt, Charles S. y Cook, Thomas D. "Hacia una superación del enfrentamiento entre los métodos cualitativos y los cuantitativos". En: *Metodos cuantitativos y cualitativos de la investigación evaluativa*. Madrid: Ediciones Morata, S. A., 1986.

Reinhard, D. L. *Methodology Development for Input Evaluating Using Advocate and Design Teams*. Tesis Doctoral Inedita, Universidad del Estado de Ohio, 1972.

Restrepo, B., Correa, S. Y Flórez, R. *Investigación evaluativa sobre situación actual y prospectiva de la formación abierta y a distancia en el Sena*. Medellín, 1989.

Restrepo, B. *Conferencias sobre evaluación*. Documento Mimeografiado, Facultad de Educación. Universidad de Antioquia, 1974.

Rist. 1997. Citado por Reichard y Cook.

Stake, R. E. *The countenance of educational evaluation, teachers college record*. 1967.

\_\_\_\_\_. *Evaluating the arts in education: A responsive approach*. Columbus, Ohio, Merrill, 1975a.

\_\_\_\_\_. *Program evaluation: Particularly responsive evaluation*. Kalamazoo: Centro de evaluación, Universidad de Western Michigan, Ocassional Paper n. 5, Noviembre, 1975b.

\_\_\_\_\_. *A Theoretical Statement of Responsive Evaluation, Studies in Educational Evaluation 2*, págs. 19-22, 1976.

\_\_\_\_\_. "The Case Study Method in Social Inquiry" *Educational Researcher 7*, pp. 5-8, 1978.

\_\_\_\_\_. "Should Educational Evaluation be more Objective or more Subjective?", *Educational Evaluation on Policy Analysis*, 1979.

Stufflebeam, D. L. *A Depth Study of the Evaluation Requirement, Theory Into Practice*, 5 n. 3, pp. 121-133, 1966.

\_\_\_\_\_. *The Citizen as Jurist: A New Mode of Educational Evaluation, Citizen Actionb in Education*, Illinois, 1974.

\_\_\_\_\_. "Evaluation as Enlightenment for Decision Making". En: Walcott A. (comp.). *Improving Educational Assessment and an Inventory of Measures of Affective Behavior*. Washington, D. C., ASCD, 1969.

\_\_\_\_\_. "The Relevance of the CIPP Evaluation Model for Educational Accountability". *Journal of Research and Development in Education* (otoño), 1971.

\_\_\_\_\_. *Metaevaluation. Occasional Paper Series*. Kalamazoo, Michigan, Centro de Evaluación, Universidad de Western Michigan, 3, 1975.

Stufflebeam, D. L. y Otros. *Educational Evaluation and Decision Making*. Itasca, l11., Peacock, 1971.

Taylor. 1974. Citado por Stufflebeam.

Tyler R. W. *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press, 1950.

Webster, W. J. *The Organization and Funtions of Research and Evaluation in Large Urban School Districts*. Texto presentado en el Congreso Anual de la American Research Association, Washington, D. C., 1975.

Wolf, R. L. *The Aplication of Select Legal Concepts to Educational Evaluation*. Tesis Doctoral Inédita, Universidad de Illinois. 1974.

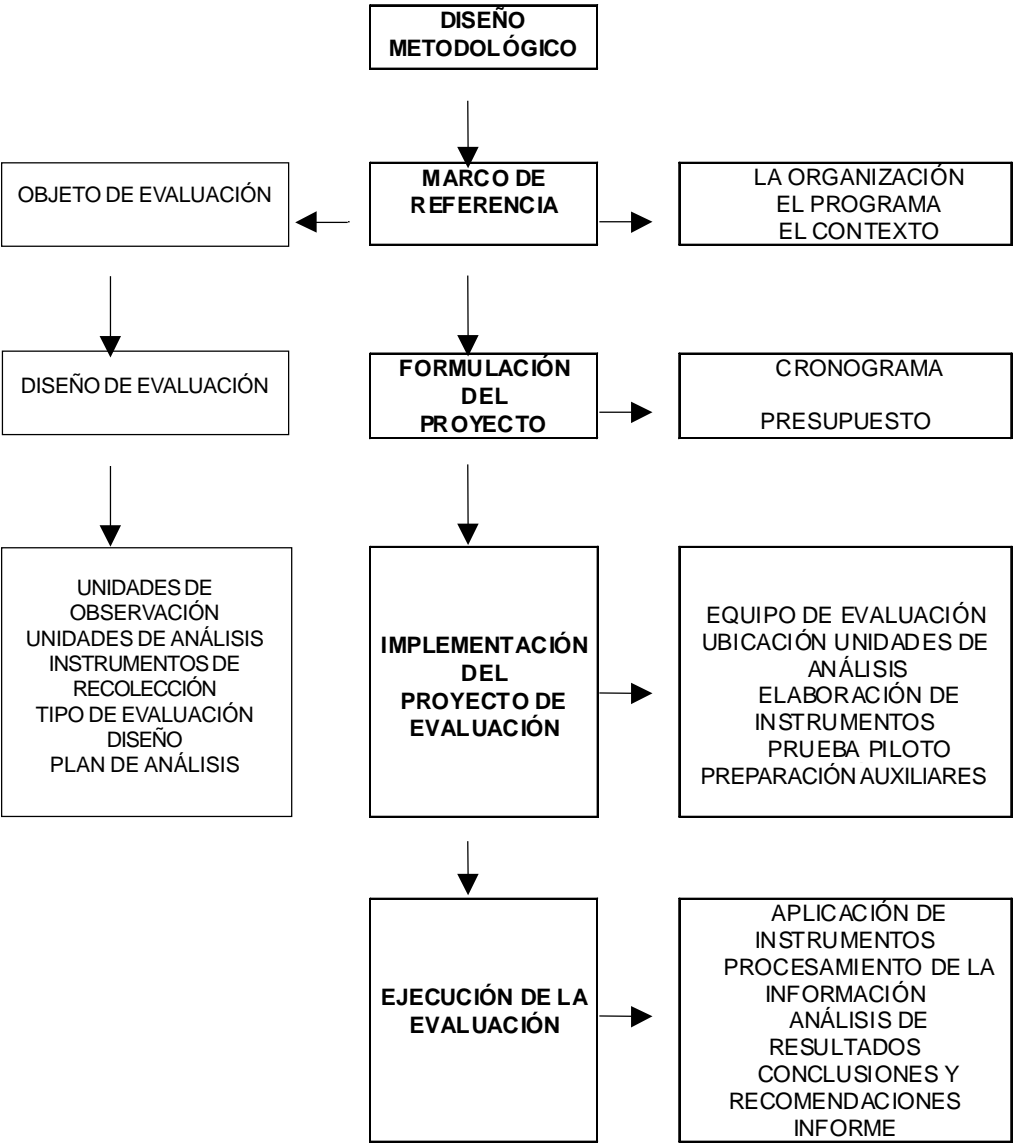


**Diseño metodológico,  
implementación, análisis  
e informes de la evaluación**

**Antonio Puerta Zapata**  
**Profesor Universidad de Antioquia**

**CUARTA UNIDAD**

Mapa conceptual



## INTRODUCCIÓN

En ésta unidad se desarrolla el proceso metodológico de la evaluación de programas. Se definen cuatro grandes momentos o etapas del proceso metodológico con sus correspondientes subrutinas. Estos momentos son: la formalización del marco de referencia, la formulación, la implementación y la ejecución del proyecto de evaluación.

La concepción aquí presentada se basa fundamentalmente en los diseños de investigación social, para reafirmar que la evaluación se concibe como una investigación evaluativa. Entendemos, que con Rossi y Freeman (1989:14), “la investigación evaluativa es la aplicación sistemática de los procedimientos de la investigación social, para valorizar la conceptualización y el diseño, la ejecución y la utilidad de los programas de intervención social”.

Varios de los elementos aquí referenciados fueron tratados en los módulos anteriores, por ello, el estudiante debe hacer una transferencia de esos elementos a la evaluación.

## OBJETIVOS DE LA UNIDAD

- Describir los momentos y componentes de la metodología de evaluación.
- Demostrar la relación lógica y secuencial de cada uno de esos componentes.
- Comprender el aporte metodológico que hace la investigación social a la evaluación de programas.

## Cuarta unidad

El diseño metodológico en una investigación evaluativa corresponde al momento en el cual se deben tomar las decisiones correspondientes a las estrategias o actividades que debe realizar el evaluador para el logro de los objetivos de evaluación o la respuesta a las preguntas de evaluación por él planteadas. En términos generales, el diseño metodológico responde a la pregunta de cómo se debe proceder en la evaluación.

Se denomina diseño, en el sentido en que avizora o planea lo que se debe hacer, y es una metodología puesto que establece un modelo, un camino y una manera de proceder.

Las decisiones que se deben tomar sobre el diseño no son de orden mecánico, es necesario tomarlas pensando fundamentalmente en la coherencia interna del proyecto. Esta certeza debe estar acompañada de la conciencia de que existen unos factores del contexto evaluativo que, en un momento dado, pueden favorecer u obstaculizar la toma de decisiones por parte del evaluador.

El diseño de evaluación en el contexto de la investigación evaluativa se asemeja, en buena medida, al diseño en una investigación social. El diseño metodológico de la evaluación “es similar al diseño correspondiente de la investigación social y consiste en la estrategia para reunir, procesar y analizar la información que se necesita para satisfacer los objetivos de evaluación” (Abello y Madarriaga, 1987:94).

Para una mejor comprensión de los diseños de evaluación se debe describir el proceso de la evaluación y ubicar en él los diseños o metodologías de evaluación. Si bien cada enfoque de evaluación propone una metodología específica, estamos abogando por un método ecléctico. En este sentido, nos sumamos a lo expresado por Stufflebeam y Shinkfield (1995:25).

...creemos que los evaluadores deberían conocer un amplio espectro de técnicas adecuadas y saber cómo aplicarlas en los distintos contextos

evaluativos. De este modo, podrían valorar, en cada situación, qué técnicas son potencialmente aplicables y cuáles pueden servir de un modo más eficaz a los propósitos particulares de la evaluación en cuestión.

De manera simplificada puede decirse que el proceso o recorrido de una evaluación comprende unos momentos o etapas.

## **ETAPAS DE LA EVALUACIÓN DE PROGRAMAS**

### **1. Especificación del marco de referencia**

El evaluador parte de la reconstrucción de un marco referencial que debe contener el detalle del programa que se quiere evaluar, la organización gestora del programa y el contexto o entorno social. A continuación se detallan cada uno de los componentes del marco referencial.

#### **1.1 *El programa de acción***

El evaluador puede encontrar dos posibles situaciones: una que el programa esté formalmente definido o en su defecto lo que existen son una serie de actividades que es necesario entrar a formalizar.

El programa objeto de la evaluación también se le define como programa de acción o simplemente proyecto de intervención. Por programa o proyecto social se entiende “el conjunto de actividades que desarrolla la empresa social, en un tiempo determinado, propiciando la resolución de un problema o necesidad de las personas, el grupo o la región” (Quintero, 1995:19).

Las teorías de la planeación y de la administración social hablan de planes, programas, proyectos y actividades como unidades de la planeación. Se entiende que el plan es lo genérico. En el caso de una ONG, el plan puede ser el mejoramiento de los ingresos de una población determinada. El programa es un componente del plan, es una unidad

menos compleja. Un programa para el ejemplo anterior podría ser el de creación de microempresas. El proyecto es un componente del programa, que está inscrito, a su vez, en un plan; es lo específico. Así, por ejemplo, puede ser un proyecto del programa enunciado con anterioridad, la capacitación técnica de los microempresarios. Finalmente, las actividades son aquellas acciones puntuales que garantizan la ejecución de los proyectos. Para continuar con el mismo ejemplo, pueden ser actividades del proyecto los cursos de capacitación en diferentes áreas.

Para el caso de la evaluación es más común hablar de programas o proyectos, pues son las unidades, que, por su relativa autonomía y limitada extensión, permiten la aplicación de estas metodologías de evaluación descritos en este módulo. Se asume que la evaluación de los planes se logra mediante la evaluación de los programas, proyectos y actividades que los conforman.

En este módulo se utiliza el concepto de programa de acción, para evitar la posible confusión con el concepto de proyecto de evaluación que es la planeación de la evaluación y para diferenciarlo de la evaluación de proyectos que tradicionalmente se ha entendido como la evaluación de factibilidad del proyecto.

En este contexto, el programa de acción es un diseño prospectivo que orienta la ejecución de unas acciones intencionadas para cambiar o transformar una realidad diagnosticada. Es un programa de acción porque contiene una serie de elementos o insumos que al ser puestos en marcha deben producir unos resultados relativamente predecibles.

El programa es un acto intencionado o deliberado que debe tener una coherencia interna o sea una lógica de relaciones entre sus componentes. El programa establece el deber ser y simula el orden de los acontecimientos.

Cuando se habla de coherencia interna del programa, se quiere decir que la naturaleza del programa armoniza con las necesidades o problema diagnosticado que se quiere resolver, que los objetivos sean viables, que las metas sean medibles, que los recursos sean suficientes y adecuados a las estrategias elegidas y que la metodología de intervención propuesta corresponda y garantice el cumplimiento de los objetivos de cambio formulados.

Para el caso de la evaluación, el programa se convierte en aquella variable que puede controlarse o manipularse, o sea que se induce deliberadamente en una realidad para impactarla; es la variable independiente o la variable conocida.

El programa debe dar cuenta de todos los ingredientes o insumos que entran en juego para alterar la situación diagnosticada como problema, de ahí la importancia de la elaboración técnica del programa, lo que facilita el proceso de evaluación.

Los elementos básicos de un programa responden a las preguntas:

- ¿Por qué existe el programa? ¿Quiere decir cuál es la justificación para su existencia? ¿Cuál es el diagnóstico o situación del problema que da origen al programa de acción? Un buen diagnóstico o una buena evaluación de necesidades es lo que justifica el diseño de una estrategia de acción denominada programa. El diagnóstico se convierte en un patrón de comparación, que permite contrastar los resultados del programa con la situación inicial de la cual se partió. De esta manera es posible decir cuáles son las características y la magnitud del cambio logrado.
- ¿El qué? ¿Cuáles son los objetivos o metas del programa? ¿Qué se pretende lograr con el programa? ¿Cuál es la situación que se quiere alcanzar? Es la utopía, el ideal, el deber ser. Ese deber ser se expresa mediante objetivos y la fijación de metas. A su vez, estas metas son las variables dependientes o de resultado que se buscan medir con la evaluación.

Los objetivos de un programa deben ser específicos, claros y medibles. Para French y Kaufman (1983:41), los objetivos deben tener en cuenta: - Los comportamientos deseados; ejemplo, la disminución del consumo de sustancias psicoactivas. - Unas circunstancias específicas; ejemplo, después de tres meses, los usuarios del programa deben haber disminuido el consumo de drogas. - El nivel de logro; ejemplo, sostener el bajo consumo o la abstinencia total después de un año de iniciado el tratamiento.

La mayoría de los programas están diseñados para producir unos efectos intermedios en los usuarios (objetivos intermedios), como mejorar la autoestima. Se asume, por parte del programa, que el logro de los objetivos intermedios es una condición necesaria y suficiente para el logro del objetivo final, por ejemplo, la reducción del consumo de alcohol (French y Kaufman, 1983:41).



- ¿El quién? Son los beneficiarios, participantes, usuarios o clientes externos del programa. Se refiere a la cobertura poblacional que tiene el programa. Esta cobertura puede corresponder a la cobertura espacial; ejemplo, un programa educativo que tiene como cobertura una región geográfica. La cobertura también puede ser una cobertura de usuarios; ejemplo, un programa de atención a discapacitados físicos tiene una cobertura, no en términos geográficos, sino en términos de las personas que tienen una determinada demanda que necesitan satisfacer o una característica que las hace susceptibles de ser agrupadas en una categoría.
- ¿El cómo? La metodología, procedimientos o actividades que deben cumplirse para el logro de los objetivos. Incluye, fundamentalmente, procesos de intervención.
- ¿El con qué? Cuáles son los recursos humanos, materiales, físicos, equipos y económicos que se requieren para el logro de los objetivos del programa. En otros términos, son los insumos requeridos para desatar los procesos que han de producir unos resultados o impactos.
- ¿El cuándo? El tiempo requerido para la implementación del programa. Corresponde al cronograma de las etapas del programa.
- ¿El dónde? Es la ubicación física o la sede del programa.
- ¿El con quiénes? Son las personas, profesionales, funcionarios que intervienen en la implementación del programa. Debe tenerse en cuenta el perfil, la cantidad y la calidad del talento humano.
- ¿Desde qué enfoque? Es la corriente del pensamiento que sustenta la línea de acción y garantiza, desde lo conceptual, la eficacia del programa. Debe explicar la relación entre los diferentes componentes del programa.

## 1.2 *La organización gestora*

El marco referencial incluye, además, información sobre la organización gestora del programa, o sea la agencia gubernamental, privada o no gubernamental responsable del proyecto.

En la mayoría de los casos, los procesos evaluativos están ligados a instituciones u organizaciones formales, en el sentido de que los programas de acción tienen asiento en ellas. Es, tal vez, imposible encontrar programas por fuera de las organizaciones.

En la evaluación, la organización debe ser considerada parte del objeto de evaluación, porque los programas tienen la impronta de las organizaciones que los patrocinan. El perfil del programa está, en buena medida, definido por la organización como un sistema complejo de componentes.

Los componentes de la organización que el evaluador debe tener en cuenta para comprender cómo ésta es parte del objeto de evaluación, son:

- La razón social de la organización. Su nombre.
- La naturaleza legal o jurídica de la organización (carácter público o privado).
- La misión, visión, principios de la organización.
- La cultura o clima organizacional.
- Los valores corporativos.
- La estructura de la organización o de gestión, entendida también como la jerarquía administrativa.
- La historia de la organización.
- Los programas de la organización.
- Los clientes internos. Perfil del talento humano vinculado, responsable de los programas.
- Los clientes externos. Perfil, necesidades y potencialidades de los usuarios de los programas.
- Las fuentes de financiación de la organización.

- La cobertura espacial o geográfica de la organización.
- La imagen corporativa hacia el exterior, o sea la opinión que el público tiene de las ejecutorias de la organización.
- Experiencias de evaluación.
- El control interno y sistemas de evaluación.

De otra parte existen una serie de características de las instituciones que pueden incidir positiva o negativamente en el desempeño de los programas y en el ejercicio de la evaluación. Estas características son:

- Las organizaciones tienen diferentes grados de complejidad que inciden también en el grado de complejidad de los procesos evaluativos. Hablamos de complejidad en términos del tamaño de la organización; por ejemplo, la evaluación de un programa en el contexto de una ONG, que administra un programa de protección de menores, es más elemental que la evaluación de un programa en otra ONG que administra simultáneamente varios programas.
- El carácter público o privado de la organización. La lógica de funcionamiento y la estructura de las organizaciones públicas es muy diferente a las organizaciones privadas. Las organizaciones públicas dependen, en buena medida, para su funcionamiento de los procesos políticos del entorno y que muchas veces producen inestabilidad de los programas y del personal, dificultades administrativas para la toma de decisiones y la ejecución de los presupuestos, la verticalidad de la autoridad y muchas veces el uso de la favorabilidad política en la prestación de los servicios.

Las organizaciones privadas, especialmente aquellas que han modernizado su gestión, por el contrario, tienen mayor estabilidad, la limitación de los recursos las obliga a realizar una planeación más rigurosa, la toma de decisiones es ágil, hay un estilo administrativo más participativo, generalmente gozan de mayor credibilidad en el medio y la prestación de los servicios es más profesional.

En términos generales, se puede afirmar que la evaluación encuentra un mejor campo de aplicación en el sector privado, que en el sector público, a pesar de que este último es el que con mayor urgencia requiere de la evaluación de sus programas, por utilizar dineros públicos, tener mayor cobertura y un mayor impacto.

- La misión de la organización determina la naturaleza de los programas que ofrece.
- El perfil de los funcionarios de la organización tiene que ver con las profesiones de sus funcionarios, la experiencia laboral, el tipo de compromiso con la organización, los grados de satisfacción personal y laboral, los estilos de gestión, las relaciones interpersonales y las experiencias en procesos de evaluación. Todas estas características del perfil de los funcionarios tienen que ver con las fortalezas y debilidades que demuestran estos mismos funcionarios en procesos concretos de evaluación. “Las personas que trabajan en una organización llegan a identificarse con sus metas, con su clientela o con aspectos específicos del programa. Los que se ocupan de la evaluación deben conocer este compromiso personal y emocional como una fuerza potente de la dinámica organizativa” (Hamilton-Smith, 1981:370).

En ocasiones, el perfil de las personas de la organización riñe con el perfil de los evaluadores, quienes están comprometidos con cierta objetividad, poseen una mirada crítica y global de los problemas, disponen de unas herramientas metodológicas propias de la evaluación, tienen otros hábitos de trabajo, manejan un enfoque teórico que no siempre es compatible con el del programa que se quiere evaluar y, por lo general gozan de una mayor libertad de pensamiento y de acción.

Las organizaciones tienen diferentes experiencias y tradición frente a la evaluación. Algunas nunca se interesaron por el tema; otras, en cambio, la incorporaron a las rutinas de gestión. Lo ideal es encontrar organizaciones que tengan un entrenamiento o al menos alguna sensibilidad por la evaluación.

La evaluación ideal será la de aquella organización que se evalúa a sí misma. Vigilaría sus propias actividades continuamente para saber si estaba cumpliendo sus metas o incluso si esas metas deberían continuar. Cuando el análisis sugiera cambiarlas o alterar los programas para alcanzarlos, estas proposiciones serían consideradas seriamente por los encargados superiores

de la toma de decisiones. Se introducirían los cambios necesarios y no habría derechos adquiridos para la continuación del programa. En cambio se iría siempre tras nuevas alternativas en un afán de servir mejor las metas propuestas (Wildavsky, citado en Hamilton-Smith, 1981:371).

- La evaluación es un evento que, de alguna manera, afecta las organizaciones en su interior; demanda de ellas procesos de concertación y de adecuación interna a los requerimientos propios de la evaluación. Existen organizaciones con climas más propicios a la evaluación; mientras que otras serán más hostiles y en las cuales es muy difícil que prospere la evaluación. “Las evaluaciones más útiles suelen darse cuando: hay acuerdo entre los interesados sobre la conveniencia de evaluar el programa. Se comprenda claramente el objetivo de la evaluación que se realizará. Hay acuerdo sobre las aplicaciones y posibles consecuencias de la evaluación” (Hamilton-Smith, 1981:368).

Para las Naciones Unidas (1978:45-46) existen unas dificultades, pero a la vez unas alternativas en las organizaciones para realizar procesos de evaluación. Éstas son:

- Falta de compromiso político administrativo. Cuando los encargados de adoptar decisiones creen que pueden comprobar la eficiencia sin necesidad de sistemas de evaluación o temen que los resultados indiquen que el rendimiento del programa no es tan satisfactorio como debería ser. Frente a esta dificultad se debe diseñar un sistema sólido de evaluación para entregar los resultados a los encargados del programa.
- Factores de costo. Es difícil comprobar que la evaluación reduce los costos del programa. Una medida paliativa para esta objeción es invertir de manera razonable en evaluación.
- Medidas institucionales. Para mejorar la economía de las operaciones de evaluación, se deben establecer éstas de manera regular e institucionalizada.

### 1.3 *Los factores del contexto*

Finalmente, como parte del marco de referencia, deben considerarse los factores del contexto; también llamados de coyuntura, que afectan directamente la vida de la organización y la ejecución del programa que se quiere evaluar. En la consideración de los factores del contexto deben tenerse en cuenta las características de los usuarios o beneficiarios directos e indirectos del programa.

Los factores de coyuntura tienen que ver con aquellas situaciones del entorno social, político, económico, cultural en las cuales se desenvuelven los programas y las cuales afectan de una manera positiva o negativa el desempeño de los mismos programas. Estas situaciones van a afectar los programas y sus efectos deben estar reflejados en los resultados de la evaluación.

Los factores de coyuntura no dependen del programa, no son siempre controlables por la organización, actúan como factores estructurales que, en la mayoría de los casos, aparecen en la marcha misma del programa y que también en la mayoría de los casos obligan a introducir cambios en la gestión del programa.

En cierto sentido, los factores de coyuntura pueden considerarse como variables intervinientes, o sea, que median entre el programa (variable independiente) y los resultados (variables dependientes).

Estos factores son de naturaleza diversa, por ejemplo: cambios en la legislación, situaciones de conflicto social, situaciones de alteración del orden público, desastres naturales, cambios en las relaciones internacionales, cambios de gobierno, decisiones de los organismos de cooperación internacional, etc.

Como parte del entorno social, están los usuarios directos de los programas. Estos usuarios tienen una serie de características que los hacen sensibles a la influencia del programa. Esto quiere decir que no todos los individuos, grupos o comunidades reaccionan igual al programa y, por lo tanto, los cambios o resultados no son los mismos.

El perfil de los usuarios, que debe estar determinado en el diagnóstico que da origen al programa, es un insumo que, como el programa, la organización y el mismo contexto,

interviene en los resultados de la intervención social. Por ello, las características de estos usuarios deben tenerse en cuenta como parte del objeto de evaluación.

En la consideración de estas características, debe darse especial prelación a aquellas directamente asociadas al tipo de programa desarrollado. Para la determinación de esas características, que podemos llamar variables, la teoría, la observación y la propia experiencia son fundamentales. Así, por ejemplo, los antecedentes ocupacionales de un grupo son importantes si se trata de un programa de capacitación profesional, mientras que para la evaluación de un programa de prevención de la farmacodependencia es más importante la historia psicosocial de los individuos. Una comunidad con experiencia en la conformación de organizaciones de base, seguramente, será más receptiva a un programa del sector solidario de la economía.

Son muchas las variables que pueden estar asociadas a los resultados de los programas: la historia de los individuos y los grupos, indicadores sociodemográficos, las experiencias anteriores de los grupos con relación a programas de intervención similares, las formas organizativas de las comunidades, etc.

## **2. La formulación del proyecto de evaluación**

La formulación del proyecto de evaluación comprende la determinación del objeto de evaluación, el diseño metodológico, el cronograma y presupuesto de la evaluación.

### **2.1 *El objeto de evaluación***

Si bien el programa de acción es el objeto de evaluación, es el evaluador quien debe definir qué asuntos, ejes, temas, variables o categorías le interesa evaluar. O sea que el evaluador es quien formula unas preguntas evaluativas que dan lugar a la definición de unos objetivos de la evaluación. Según Briones (1991:86), “los objetivos de evaluación son los conocimientos o informaciones relativas a los objetos, instancias o características del programa que se requiere recoger y analizar de acuerdo con las finalidades que sirven de justificación al estudio evaluativo”.

Los objetivos de evaluación pueden estar referidos al comportamiento de la organización frente a la ejecución del programa, a los productos, efectos, resultados o impactos del programa. Quiere decir, a medir los cambios en los usuarios del programa o de su entorno.

Además de la formulación de una preguntas u objetivos de evaluación, debe existir una delimitación temporal del objeto de evaluación. O sea, es el evaluador quien define el periodo de vida del programa que se quiere evaluar. Pueden evaluarse todas las etapas del programa o algunas de ellas.

Finalmente, se debe hacer una delimitación espacial, especialmente cuando se trata de un mismo programa ejecutado en diferentes espacios geográficos o institucionales o con diferentes tipos de usuarios o poblaciones.

## *2.2 El diseño de la evaluación*

El diseño de una evaluación comprende: la determinación de las unidades de observación, las unidades de análisis, los instrumentos de recolección de información, definición del tipo de evaluación, el diseño, el procesamiento e interpretación de la información.

El diseño metodológico comprende unas áreas técnicas que, según Stufflebeam y Shinkfiel (1995:25), deben conocer a fondo los evaluadores profesionales. Se encuentran las siguientes:

Las entrevistas, los informes preliminares, el análisis de contenidos, la observación, el análisis político, el análisis económico, el examen investigativo, los informes técnicos, el estudio de caso, la evaluación sin metas, la escucha de opiniones contrapuestas, las asociaciones de defensa, las listas de control, la elaboración de pruebas, el análisis estadístico, el diseño de la investigación, el análisis sistemático, la teorización y la administración del proyecto.

### *2.2.1 Las unidades de observación*

Las unidades de observación son las variables, categorías, asuntos o componentes que deben ser medidos.

Los objetivos o preguntas de evaluación contienen las variables que deben ser observadas o medidas. Para medir una variable es necesario traducirla a indicadores.

La construcción de indicadores es, tal vez, una de las tareas más complejas de la evaluación. Existe una literatura disponible sobre el sentido de los indicadores, pero su



diseño particular es un ejercicio dispendioso, pues se trata de traducir objetivos de evaluación a variables y éstas a indicadores para poderlas medir.

Los investigadores sociales se han preocupado por el tema de las definiciones operacionales o indicadores para medir las variables de sus investigaciones y han propuesto, en consecuencia, modelos para la medición de las variables. Los evaluadores pueden echar mano de esa lógica de construcción de indicadores y de los instrumentos para su medición, pero hay que entender que los indicadores de evaluación no pueden ser universales, sino que corresponden a la especificidad del programa que se quiere evaluar.

La evaluación de un programa demanda la construcción y validación de indicadores particulares que van a depender de los objetivos del programa que se quiere evaluar, de su enfoque teórico, de la población usuaria del programa, de los propósitos de la evaluación.

Los indicadores son medidas específicas (explícitas y objetivamente verificables de los cambios o resultados de una actividad). Los indicadores sirven de patrón para medir, evaluar o mostrar el progreso de una actividad respecto a las medidas establecidas, en cuanto a la entrega de sus insumos (indicadores de insumo), la obtención de los productos (indicadores de producto) y el logro de sus objetivos (indicadores de efecto e impacto) (Quintero y Velásquez, 1993).

Los indicadores pueden ser directos o indirectos. Los indicadores de naturaleza cuantitativa, por lo general, pueden medirse de una manera directa; ejemplo, el número de personas atendidas por un programa, el número de créditos otorgados para un proyecto asociativo. Los indicadores son indirectos cuando se trata de medir objetivos intangibles; ejemplo, medir liderazgo, participación, motivación, etc.

Los indicadores están determinados por la naturaleza de los objetivos, por los efectos y por el impacto que se busca con el programa de acción. La construcción de indicadores depende, en buena medida, de la claridad y especificidad de los objetivos. En muchos casos encontramos que los programas sociales no tienen objetivos y metas claramente definidas. Esta situación se debe, entre otras razones, a:

- En el pasado las acciones sociales eran de buena voluntad y estaban por fuera de la planificación del desarrollo. Hoy existe una mayor preocupación por la elaboración estratégica de los programas.

- La formulación de un objetivo en el campo del desarrollo social tiene mayores dificultades que en el campo de lo económico o de las obras físicas. Lo social, de alguna manera, es de más difícil control y medición.
- No siempre existe consenso por parte de los funcionarios en las organizaciones acerca de las metas de los programas debido a que priman las propias concepciones que ellos tienen sobre la organización y sus programas; igualmente, a la propia formación profesional de cada uno de ellos.

Cuando se presenta esta última situación de no consenso sobre las metas en una organización o de un programa, Weiss (1990:43) recomienda las siguientes alternativas, que puede adoptar el evaluador.

- Plantear la situación en el interior de la organización y esperar a que el personal del programa llegue a un acuerdo.
- El evaluador puede estudiar el programa, hablar con los ejecutores del programa y observar el desempeño del programa para luego formular por sí mismo las metas del programa.
- El evaluador, conjuntamente con el personal del programa, podrá formular las metas.
- Sugerir la realización de un estudio exploratorio antes de una evaluación.

Para la definición de los indicadores debe tenerse en cuenta la conceptualización que se tenga del programa y la experiencia de las personas que lo ejecutan. Además del empleo de este mecanismo, Weiss (1990:50-53) recomienda tres alternativas para determinar los indicadores de resultado.

- Buscar medidas existentes, empleadas por otros evaluadores en situaciones similares.
- Usar medidas múltiples. “Medidas separadas pueden combinarse para formar una medida general del éxito del programa”.

- Usar medidas aproximadas. Deben utilizarse las medidas de acuerdo con las metas, si están establecidas a corto, mediano o largo plazo.

Las evaluaciones producen diferentes tipos de información. Según las Naciones Unidas (1978:11), la evaluación proporciona tres tipos de información:

- Información sobre rendimiento del programa. Es la información que da cuenta de si la ejecución del programa se realizó de conformidad con el plan y el presupuesto asignado. Para este tipo de información se cuenta con indicadores establecidos en el cronograma de acción y en el presupuesto asignado a la ejecución del programa.
- Información sobre la eficacia del programa. Da cuenta si el programa logró los objetivos propuestos y en qué medida. Para este tipo de información, los indicadores son las metas del programa. Cuando no existen las metas, los objetivos deben ser traducidos a indicadores observables.
- Información sobre la eficiencia del programa. Determina si los resultados del programa se obtuvieron de la forma más económica, maximizando productos y minimizando insumos. Los indicadores adecuados a este tipo de información son aquellos que tienen como unidad de medida la monetaria.

Para Espinoza (1983:43), la evaluación debe partir de una traducción de los objetivos a metas, pues el autor considera que una meta es un objetivo cuantificado y calificado. Por lo tanto, a cada objetivo corresponde una o varias metas. O sea que para medir el grado de obtención de los objetivos es necesario medir el grado de obtención de las metas. Para el autor, la evaluación de las metas tiene varios aspectos.

- Evaluación de la cantidad. Consiste en medir el número de unidades logradas en comparación con las unidades programadas.
- Evaluación de calidad. La meta tiene unos atributos que conforman su calidad y se debe medir el grado de obtención de los atributos.
- Evaluación del tiempo. Mide el tiempo gastado en la consecución de la meta en comparación con el tiempo previsto.

- Evaluación de los objetivos. Mide el grado de obtención de los objetivos.
- Evaluación de la cobertura. Mide si las metas logradas cubrieron a toda la población usuaria del programa.
- Evaluación de costos. Mide la inversión realizada para el logro de las metas en comparación con los costos previstos.

Victor Manuel Quintero (1995:35) distingue dos tipos de indicadores: de gestión o eficiencia y de logro o eficacia.

Define los indicadores de gestión como aquellos que “tienen como misión valorar el rendimiento (productividad física) de insumos, recursos y esfuerzos dedicados a obtener ciertos objetivos con unos tiempos y costos registrados y analizados”.

Según el mismo autor, los indicadores de gestión permiten la valoración de los siguientes elementos: la utilización de los recursos humanos, físicos y financieros, el tiempo, el cumplimiento de actividades, tareas o metas intermedias, el rendimiento físico promedio (productividad), o sea la relación que existe entre los logros y los recursos utilizados, el costo promedio y el grado de eficiencia.

Los indicadores de logro o de eficacia, también denominados de éxito, son “hechos concretos, verificables, medibles, evaluables, que se establecen a partir de cada objetivo” (Quintero, 1995:36). Según el mismo autor, existen cuatro indicadores de logro.

- Indicadores de impacto. Logros a largo plazo, contribuyen al cumplimiento de la misión de la organización.
- Indicadores de efecto. Logros a mediano plazo, contribuyen al cumplimiento de los objetivos programáticos en una región específica.
- Indicadores de resultado. Logros a corto plazo, contribuyen a resolver las necesidades y problemas del grupo.
- Indicadores de producto. Logros del inmediato plazo, contribuyen al cumplimiento de los objetivos específicos del proyecto social.

Las Naciones Unidas (1978:7) utilizan algunos de los conceptos anteriores de manera diferente.

Entienden los productos como las consecuencias físicas producidas por el programa, como escuelas construidas, kilómetros de carreteras habilitados. También incluyen la medición de los servicios prestados, ejemplo: personas capacitadas, créditos concedidos.

Los efectos se entienden como las consecuencias directas o inmediatas de los productos, ejemplo: mayores tasas de asistencia escolar, mejores cosechas.

Las repercusiones, como aquellos cambios que ocurren en el nivel de vida del grupo, como consecuencia del programa.

### 2.2.2 Unidades de análisis.

Las unidades de análisis también son denominadas fuentes de información. Una vez definidos los indicadores, nos preguntamos quiénes (individuos, familias, grupos, comunidades, etc.) o qué objetos (documentos, monumentos, artefactos, situaciones, etc.) contienen esos indicadores. Hablamos, entonces, de las fuentes de información que pueden ser objetos o sujetos. La unidad de análisis “es un dato —observable y externo— que se convierte en la unidad de información básica. Es un concepto referido al objeto de estudio, por ejemplo, individuos, grupos, región, hechos, procesos” (Aigner, 1997:22).

La elección de las fuentes está condicionada por las unidades de observación (lo que se quiere medir), por la naturaleza del programa objeto de evaluación, por los objetivos de evaluación, por los recursos humanos, técnicos y financieros disponibles, por la posibilidad de acceder a las fuentes y por la disponibilidad de tiempo.

El modelo de evaluación es también determinante en la selección de las fuentes. Los enfoques cualitativos hacen énfasis en los muestreos estadísticos y algunos diseños, como el experimental, implican la selección de muestras aleatorias. Los diseños cualitativos trabajan con muestreos teóricos, muestreos intencionales, informantes claves, etc.

Para determinar las fuentes se deben tener claros los conceptos de universo, población, muestra, tipo de muestreo que fueron vistos en los módulos anteriores.

### 2.2.3 *Instrumentos de recolección de información*

Para capturar o recoger información sobre los indicadores que están presentes en unas fuentes debemos recurrir a los instrumentos de medición o recolección de información. La evaluación utiliza los instrumentos propios de la investigación social, bien sean de naturaleza cuantitativa, como los cuestionarios, o de naturaleza cualitativa, como las entrevistas en profundidad.

Decimos que son instrumentos de medición porque son sensibles a los indicadores y son de registro porque no basta con medir, sino que se debe guardar y ordenar la información de tal manera que pueda ser recuperada e interpretada.

Los instrumentos contienen a los indicadores y son aplicados a las fuentes de información.

La elección de los instrumentos depende de la naturaleza de lo que se quiere evaluar, el tipo de indicadores, las fuentes de información, los recursos humanos y económicos disponibles, el tiempo de ejecución de la evaluación, el diseño elegido y la perspectiva conceptual y metodológica del evaluador o del equipo de evaluación.

Los instrumentos de recolección de información fueron vistos en los módulos anteriores.

### 2.1.4 *Tipo de evaluación*

El tipo de evaluación puede ser ex-ante, formativa o sumativa, según el momento de ejecución del programa que se quiere evaluar.

### 2.1.5 *El diseño*

La evaluación toma prestados de la investigación social los diferentes diseños disponibles y que han sido probados en el estudio de la realidad social. Algunos de estos diseños, inclusive, no son exclusivos de las ciencias sociales, también son aplicados en las ciencias naturales, como es el caso de los diseños experimentales.

Se debe determinar si el diseño de la evaluación será no experimental o experimental. Los diseños de evaluación están asociados a los enfoques de evaluación: de corte cuantitativo o de corte cualitativo y a los modelos de evaluación existentes.

En los módulos de investigación cuantitativa y cualitativa se pueden encontrar referencias a estos diseños y a los enfoques.

#### **2.1.6 *El plan de análisis***

El evaluador debe prever, desde el proyecto, cómo va a realizar el procesamiento, organización e interpretación de la información.

El procesamiento y análisis de los resultados. Esta etapa implica la limpieza de la información (cuando se han administrado instrumentos de recolección de información estandarizados), la organización de la información, el procesamiento de datos, la representación de datos y el análisis estadístico y lógico de los resultados.

La evaluación, igualmente, comparte con la investigación social los mismos procedimientos para el procesamiento e interpretación de la información. Hoy están disponibles paquetes estadísticos de procesamiento de información para datos cuantitativos y cualitativos.

Se entiende que según el enfoque de evaluación elegido va a depender el procesamiento e interpretación de la información. Para ello, el lector debe remitirse a los módulos anteriores que desarrollan esta temática en detalle.

### **2.3 El cronograma**

El cronograma de la evaluación hace referencia a la determinación anticipada del tiempo que dura la planeación y ejecución de la evaluación. Normalmente los cronogramas se elaboran utilizando la semana como unidad de medida.

El cronograma se define por el tipo y número de actividades previstas para la formulación del proyecto, la implementación y la ejecución de la evaluación. Las etapas del proceso evaluativo sugeridas en esta unidad pueden servir de guía para la conformación del cronograma.

No sobra recordar que el cronograma del proceso de evaluación es diferente al cronograma de ejecución del programa de acción.

## **2.3. El presupuesto**

El presupuesto es la actividad de asignación del valor monetario a cada una de las actividades, recursos humanos, físicos, locativos, papelería, transporte, etc., que se requiera para la ejecución del proyecto de evaluación.

La elaboración de un cronograma y de un presupuesto para la evaluación permiten ser más realistas en el diseño de la evaluación y contar con instrumentos de control y monitoreo de la gestión del proceso de evaluación.

También hay que recordar que el presupuesto de la evaluación es diferente al presupuesto de ejecución del programa de acción.

## **3. Implementación del proyecto de evaluación**

La implementación del proyecto de evaluación implica: la conformación del equipo de evaluación, la elaboración de instrumentos de captura de información, la prueba piloto de los instrumentos si es requerida, la selección de las fuentes o audiencias y la preparación del personal auxiliar si es necesario.

### **3.1 El equipo de evaluación**

La evaluación puede ser planeada y ejecutada por un evaluador quien, obra como experto o en representación de una agencia evaluadora. Éste es el caso de las evaluaciones externas.

La evaluación puede, también, planearse por un equipo o grupo evaluador. En el caso de las evaluaciones internas, el grupo está conformado por personas vinculadas al programa o a la organización gestora del programa. Pueden existir grupos de evaluación que obran como agentes externos y también pueden darse grupos con una conformación mixta que incluye personal interno y personal externo.

El equipo de evaluación puede integrarse para formular el proyecto de evaluación solamente, para formularlo y ejecutarlo o sólo para ejecutarlo.

La tendencia moderna de la evaluación es la de ser una tarea colectiva, mas que individual y va más allá al plantear modelos más participativos de la evaluación. Se sugiere que los



diferentes actores del programa participen en su evaluación. Aparecen, entonces, nociones como la de empoderamiento, o sea, el poder de decisión que ganan los actores en los procesos autogestionados.

Otro concepto que está ligado a la evaluación es el de participación ciudadana, también denominada por algunos participación comunitaria, participación social o participación popular.

La participación ciudadana ocupa hoy un rol protagónico en los procesos de planeación, ejecución y evaluación de los programas sociales: los planes de gobierno, las agencias de cooperación internacional, la planeación regional y local incluyen la participación como un componente básico para el logro de los objetivos de mejoramiento de la calidad de vida que se proponen los programas sociales.

La participación ciudadana, según los modernos enfoques del desarrollo y la administración, debe darse en todas las fases, desde la evaluación de necesidades mediante los métodos de autodiagnóstico o diagnóstico participativo, la programación y la evaluación.

Para la evaluación, la participación ciudadana es un componente muy importante. Según Wagner y Dinkel (1981:9-10), la participación ciudadana en la evaluación se justifica por las siguientes razones:

- La participación hace más responsable a las instituciones con relación a la comunidad.
- La participación ciudadana refleja los valores de la comunidad.
- La participación ciudadana incrementa el uso de la información evaluativa.
- Es una oportunidad para que los ciudadanos adquieran habilidades y conocimientos.
- Incrementa el conocimiento de los ciudadanos acerca del programa.
- Proporciona un beneficio terapéutico para los usuarios del programa.

- Los ciudadanos proveen fuentes adicionales de información para la evaluación.
- Es un mecanismo para estimular la aceptación de los programas por la comunidad.

La participación ciudadana no reemplaza la función del evaluador, pues éste debe ser un miembro calificado del grupo y debe cumplir las funciones de coordinador. El profesional o líder del equipo, así la evaluación sea participativa, seguirá teniendo un papel determinante.

Desde el comienzo de la evaluación deben clarificarse las funciones del líder o coordinador y las de los demás miembros del equipo. Es indispensable organizar unas condiciones de trabajo: asignar recursos, fijar tiempos, establecer lugares para el trabajo, distribuir las funciones. Una adecuada atención a estos asuntos administrativos evita conflictos, racionaliza el trabajo y garantiza unos productos evaluativos más confiables y eficientes.

La evaluación comporta, además de los aspectos técnicos, otras consideraciones de orden político y administrativo que es necesario atender. Es claro que la actividad evaluativa exige habilidades políticas, además de las técnicas por parte del evaluador.

En este mismo sentido, Andrade y Otros (1987:19) advierten cómo el conocimiento técnico y la experiencia profesional no son suficientes para el evaluador. Éste debe identificar y trabajar con aquellas personas que toman las decisiones y todas las personas interesadas en los resultados de la evaluación como los clientes, los grupos organizados de la comunidad, el gobierno, las agencias financieras. En todos estos casos debe conformarse un grupo evaluador y el responsable debe ser el líder del equipo, el mediador y el facilitador.

El evaluador, responsable de orientar la evaluación según la naturaleza de los usuarios de la información evaluativa, debe identificar “quién necesita qué información, en que condiciones, para qué propósitos”. “El evaluador debería tratar de identificar a aquellas personas estratégicamente ubicadas, que sean dedicadas, competentes y que se interesen por el asunto, aunque no ocupen una alta posición en la institución” (Andrade y otros, 1987:22).

### **3.2 Ubicación de las unidades de análisis**

En el diseño metodológico se toma la decisión de determinar cuáles serán las fuentes de información, o sea cuáles son los sujetos u objetos portadores de los indicadores que se requieren medir. En el momento de la implementación del proyecto de evaluación se trata de ubicar las unidades de análisis; lo que quiere decir, la localización física de los sujetos u objetos portadores de la información. En el caso de los sujetos, se trata de concertar las entrevistas, hacer la motivación necesaria para lograr su cooperación. Si se está utilizando un sistema de muestreo, es preciso hacer la aplicación correspondiente, seleccionando los casos específicos representados en la muestra.

Si se están utilizando documentos, el equipo evaluador debe disponer de ellos mediante la localización de los originales, los textos, los recortes de prensa, informes estadísticos, publicaciones, fotocopias, etc.

### **3.3 Elaboración de instrumentos de recolección de información**

En este momento del proceso, deben elaborarse los instrumentos enunciados en el proyecto de evaluación: cuestionarios, guías de observación o entrevista, test, guías para historias de vida, guías para entrevistas en profundidad o para grupos focales. La manera de confeccionar estos instrumentos se puede consultar en los módulos anteriores.

### **3.4 Pruebas piloto**

Siempre que se utilizan instrumentos de recolección de información, es conveniente hacer pruebas piloto para probar la consistencia y confiabilidad de los instrumentos elaborados.

La aplicación de pruebas piloto sirve para revisar los contenidos y forma de los instrumentos. Así se puede mirar si están comprendidos todos los indicadores, si es el orden el más indicado, si son claras y precisas las preguntas, cómo es la reacción de los sujetos frente a cierto tipo de preguntas, si la forma de presentación en el caso de instrumentos autoadministrados es la más conveniente, etc.

### **3.5. Preparación de auxiliares**

Para el caso de aquellas evaluaciones en la cuales se requiere de personal adicional al del equipo evaluador, para llevar a cabo la evaluación, es preciso hacer una capacitación a cada una de las personas auxiliares según el tipo de tarea que tenga que desempeñar en el grupo. Generalmente, el personal auxiliar es empleado en tareas administrativas, de digitación, recolección de la información (entrevistadores, encuestadores, talleristas), para búsquedas documentales, establecimiento de contactos, etc.

La preparación de entrevistadores, de observadores y, en general, de personas que en sí mismas, según se afirma en los enfoques de investigación cualitativa, son instrumentos de recolección de información, estos esfuerzos de capacitación tienen que ser más especializados. Se deben tener criterios unificados, una postura ética y el desarrollo de una gran destreza en el ejercicio de estas responsabilidades.

Es menos exigente la capacitación cuando se utilizan instrumentos cuantitativos, los cuales son intermediarios entre el evaluador y la realidad que se quiere evaluar.

Además de la capacitación en las funciones específicas, debe instruirse a los auxiliares en todo lo relacionado con la evaluación: objetivos, propósitos, diseño, conceptualización, marco de referencia, indicadores, etc. Un conocimiento global del proyecto de evaluación permite desarrollar un sentido de identidad con las tareas que se requieren. Igualmente, estos auxiliares deben ser tratados con respeto y ofrecerles condiciones dignas para el desempeño de sus funciones, que garanticen la calidad del trabajo.

En el caso de utilizar instrumentos de mayor naturaleza cualitativa, deben hacerse simulaciones de estos instrumentos con informantes claves voluntarios a quienes se les explica muy bien el alcance de su participación.

## **4. Ejecución de la evaluación**

El momento de la ejecución de la evaluación significa la puesta en marcha de las etapas correspondientes a la formulación e implementación del proyecto de evaluación. Esta etapa corresponde al trabajo de campo que debe ser planeado. También se le denomina administración o gestión de la evaluación y comprende la recolección de la información, su procesamiento e interpretación, las conclusiones y recomendaciones y la escritura

del informe final. Aunque todas estas tareas deben estar previstas en el proyecto de evaluación, es muy probable que en el trabajo de campo sea necesario introducir cambios a la propuesta original. En el trabajo de campo, el propio evaluador confronta la coherencia del proyecto formulado.

#### **4.1 Aplicación de los instrumentos de medición y de registro**

Es el momento para aplicar los instrumentos de medición a las unidades de análisis o fuentes de información.

En la fase de construcción del marco referencial ya se ha recopilado una información muy rica que tiene que ver con los fundamentos conceptuales del programa, la organización y el contexto. Además, la prueba piloto de los instrumentos brinda a los evaluadores una orientación sobre el tipo y calidad de información que van a recoger.

En todos los casos es muy importante tener un profundo respeto por cada uno de los informantes, establecer con ellos una buena relación, ser claros en la explicación de los objetivos de la evaluación y demostrarles la importancia que ellos tienen para el resultado final de la evaluación.

El equipo de evaluación debe ejercer una supervisión especial de esta etapa, especialmente si se han contratado personas auxiliares para cumplir con las tareas de recolección de información. En algunos casos, la falta de una preparación suficiente y la ausencia de un acompañamiento puede viciar ostensiblemente los resultados de la evaluación.

#### **4.2. Procesamiento de la información**

Esta etapa implica: recibir la información, supervisar su calidad (limpieza de datos), ordenarla (establecer categorías y códigos para la información cualitativa, digitarla, procesarla y representarla).

Si bien el uso del computador facilita y garantiza rapidez y alta calidad en el procesamiento de la información, la intervención de los evaluadores es fundamental.

La información debe ser recibida, contada y supervisado su contenido. Antes de ser digitada debe ordenarse. Generalmente, los digitadores no tienen tiempo o por su labor

mecánica no detectan las inconsistencias en la información. Éstas deben detectarse por los evaluadores, quienes conocen la naturaleza de la información recogida.

Para el procesamiento de la información debe hacerse en un equipo adecuado y utilizando el programa que garantice el análisis de la información según los parámetros establecidos en el plan de análisis.

Desde la elaboración de los instrumentos, se debe tener muy claro cuál es el programa que se va a utilizar con el objeto de adecuar el sistema de codificación a los requerimientos del programa o paquete estadístico elegido, si es del caso.

#### **4.3. Análisis de los resultados**

Una vez procesada la información y teniendo los datos representados en cuadros o gráficos, se hace una lectura e interpretación de ellos. En la lectura se destacan aquellos datos más relevantes que se ubican en los extremos y aquellos que son típicos o promedios.

En esta etapa la formación teórica del evaluador, su capacidad interpretativa y el marco referencial son definidos para lograr que los datos adquieran su verdadero valor. No se quiere decir que éste es el único momento del análisis, pues en todo el proceso de evaluación, incluyendo la etapa de formulación del objeto de evaluación y de recolección de información, el evaluador debe tener una actitud de análisis e interpretación.

El análisis debe ser lógico y estadístico. El análisis estadístico va a depender de la naturaleza de los datos producidos por la evaluación (ver módulo de diseños cuantitativos). El análisis lógico o de contenido parte del marco referencial y conceptual del programa evaluado. Debe recordarse que los datos tienen sentido en la medida en que se hagan hablar desde una interpretación comprensiva (ver módulo de diseños cualitativos).

#### **4.4 Conclusiones y recomendaciones**

La evaluación es una investigación aplicada que debe producir conclusiones a manera de hallazgos conceptuales y metodológicos referidos al programa, pero además debe ofrecer alternativas de acción a manera de acciones mejoradoras para el programa. Los evaluadores deben tener, por lo tanto, habilidades administrativas, conocimientos sobre planificación y gerencia para proponer recomendaciones que mejoren el programa.

## 4.5 Informes de evaluación

El informe de evaluación debe dar cuenta no solo de los resultados finales de la evaluación, sino que debe recoger además el proceso que va desde la formulación del proyecto de evaluación, su implementación y ejecución.

No basta con la escritura de los informes de evaluación, también se requiere tener estrategias para la disseminación o entrega de los resultados a los diferentes públicos o usuarios de la información evaluativa.

Para la presentación de los informes pueden utilizarse formatos escritos o visuales, según los destinatarios de la información. Los formatos más comunes son: el informe técnico, el resumen ejecutivo, el reporte profesional, el artículo para publicación, una conferencia de prensa, un reporte para un medio de comunicación, una reunión pública, un taller, afiches, plegables, memos, audiovisuales, etc.

Los resultados correspondientes a una evaluación podrán presentarse simultáneamente de diversas maneras, de acuerdo con los criterios definidos a continuación:

- La naturaleza del programa que se está evaluando: el tipo de variables elegidas, los indicadores empleados, el diseño utilizado y las técnicas de análisis utilizadas.
- Los contratantes de la evaluación. Los que facilitan los recursos para financiar la evaluación podrán establecer en el convenio las condiciones para la presentación y definir los destinatarios de los resultados.
- Los usuarios de la evaluación, o sea los responsables de la utilización de los resultados. Los usuarios más comunes son: los administradores de los programas, los contratantes de la evaluación, los profesionales o mandos medios encargados de prestar los servicios, los usuarios o beneficiarios de los programas, las organizaciones políticas, las organizaciones civiles o sociales, los profesionales independientes del programa como investigadores universitarios u otros académicos.
- El nivel de comprensión de los destinatarios que tiene que ver con la experiencia en el manejo de estos estudios, los intereses intelectuales, la capacidad de entender cierto tipo de análisis.

- Las condiciones de tiempo y espacio en las cuales se presentará el estudio.
- Los recursos con que se cuenta en el presupuesto para invertir en el informe.

De otra parte, la diseminación de los resultados tiene que ver con la manera como son difundidos los resultados de una evaluación. La diseminación guarda una relación inseparable con los formatos de presentación de los resultados. Para Morris y Otros (1987:22), las normas que deben tenerse en cuenta para difundir los resultados son:

- La información debe ser entregada a los usuarios adecuados.
- Los informes deben tratar los asuntos que los usuarios consideren que son importantes.
- Los informes deben entregarse a tiempo para que sean útiles. También deben ser claros para los destinatarios.

Estos mismos autores, recomiendan unos criterios que deben ser tenidos en cuenta para entregar la información a los usuarios de ella.

- Diferentes usuarios desean información diferente, inclusive para responder a las mismas preguntas.
- Algunos usuarios no tienen claro lo que necesitan conocer.
- Algunos usuarios, esperan apoyo para sustentar su propio punto de vista.
- Algunos usuarios esperan que se produzcan cambios en la información durante el desarrollo de la evaluación.

En cuanto al contenido del informe de evaluación, éste debe tener en buena medida lo que fue previsto en el proyecto de evaluación, los pasos dados en la ejecución e implementación del proyecto. Debe recordarse que el informe debe ser lo más claro y completo posible pues va dirigido a personas que no participaron directamente de la experiencia evaluativa. Estos son los contenidos que debe tener un informe:



- El marco referencial, que incluye: una descripción completa del programa de acción, un perfil de la organización gestora y los factores del contexto que inciden en el desempeño del programa.
- Los antecedentes, justificación y propósitos de la evaluación.
- Los objetivos o preguntas de evaluación. El objeto de evaluación y su delimitación.
- El diseño metodológico. El enfoque utilizado y el modelo. El equipo evaluador. La agencia evaluadora (cuando se trata de una evaluación contratada).
- La ejecución de la evaluación.
- Los resultados de la evaluación.
- Recomendaciones o acciones mejoradoras.
- Evaluación de la propia evaluación. Tiempo empleado, costos, aciertos y dificultades del proceso evaluativo. Necesidad de otros estudios.
- Bibliografía.
- Anexos.

## AUTOEVALUACIÓN

Conteste el siguiente cuestionario:

- ¿Qué se entiende por diseño metodológico de la evaluación?
- ¿Qué toma la evaluación de la investigación social?
- ¿Cuales son los momentos del proceso evaluativo?
- ¿En qué consiste el marco de referencia de una evaluación?
- ¿Qué son los indicadores?
- ¿En qué consiste la administración de la evaluación?

## LECTURA COMPLEMENTARIA

Briones, Guillermo. Evaluación de programas sociales, México, Editorial Trillas, 1992.

## ACTIVIDADES DE EVALUACIÓN

El estudiante, al finalizar esta unidad, debe presentar un anteproyecto de evaluación que contenga los siguientes elementos:

- Formular por lo menos tres preguntas u objetivos de evaluación ex-post, correspondientes a un programa de intervención social.
- Enunciar las variables contenidas en esas preguntas u objetivos, que deben ser medidas a través de los indicadores.
- Definir cuáles serían las unidades de análisis o fuentes de información.
- Enumerar los instrumentos de recolección más adecuados para esos objetivos de evaluación, a fin de medir esas variables y para ser aplicados a esas fuentes de información.
- Diseñar y redactar un primer borrador de uno de los instrumentos enunciados en el punto anterior donde se tengan en cuenta los objetivos de la evaluación, las variables enunciadas y los indicadores de dichas variables.
- A partir del numeral anterior, elaborar un plan de análisis de la información recogida con el instrumento.

## BIBLIOGRAFÍA

Aignerren, A. Miguel. *Investigación cuantitativa en ciencias sociales. Procesamiento y análisis de datos*. Medellín: Universidad de Antioquia, 1997.

Abello, R. y Madarriaga, Camilo. *Estrategias de evaluación de programas sociales*. Barranquilla: Uninorte, 1987.

Andrade, Selly y Otros. *Métodos cualitativos para la evaluación de programas. Un manual para programas de salud, planificación familiar y servicios sociales*. Estados Unidos: The Pathfinder Fund, 1987.

Briones, Guillermo. *Evaluación de programas sociales*, México: Editorial Trillas, 1991.

Espinoza, V., Mario. *Evaluación de proyectos sociales*. Buenos Aires: Editorial Humanitas, 1983.

French, John F. and Kaufman, Nancy. *Hanbook for preventioon evaluation*, Usa, Department of Health and Human Services, 1983.

Hamilton, Smith, Elery. "La estrategia y la metodología de la evaluación de programas de desarrollo social", en *Planificación social en América Latina y el Caribe, Chile, UNICEF*, 1981.

Morris, Lynn. *How to communicate evaluation findings*. USA Sage Publications, 1987.

NACIONES UNIDAS. *Manual de vigilancia y evaluación sistemáticas de programas integrados de desarrollo*, Nueva York, 1978.

Quintero, Víctor Manuel y Velásquez, Mario. *Los indicadores sociales. Marco conceptual*, Cali: FES, 1993.

Quintero, Víctor Manuel. *Evaluación de proyectos sociales: construcción de indicadores*. Cali: Fundación FES, 1995.

Rossi, Peter H. y Freeman, Howard. *Evaluación. Un enfoque sistemático para programas sociales*, México, Editorial Trillas, 1989.

Stufflebeam, Daniel L. y Shinkfiel, Anthony J. *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Ediciones Paidós, 1995.

Wsgnrt, Dinober and Dinkel, Nancy R. *A trust of evaluation*. Tampa Florida: The Florida Consortium for Research and Evaluation, 1981.

Weiss, Carol. *Investigación Evaluativa*, México: Editorial Trillas, 1978.



# LECTURAS COMPLEMENTARIAS

1. **Introducción a la evaluación.** Daniel L. Stufflebeam y Anthony J. Shinkfield ..... 169
2. **La autoevaluación como proceso previo a la acreditación de programas de pregrado de la Universidad de Antioquia.**  
Bernardo Restrepo Gómez ..... 211
3. **Algunos elementos para una evaluación descentralizada de proyectos.** J. Forero ..... 230
4. **Hacia una superación del enfrentamiento entre los métodos cualitativos y los cuantitativos.**  
Charles S. Reichardt y Thomas D. Cook ..... 240
5. **Diseños de evaluación.** Guillermo Briones ..... 273





## **Lectura complementaria N° 1**

**Daniel L. Stufflebeam y Anthony J. Shinkfield**

### **INTRODUCCIÓN A LA EVALUACIÓN**

*Tomado de*

*Evaluación sistemática guía teórica y práctica*

La evaluación es una de las actividades fundamentales de los servicios profesionales más solventes. Los clientes de estos profesionales merecen una asistencia que satisfaga sus necesidades, que sea de una alta calidad, moderna y eficiente. Para que los profesionales se responsabilicen de satisfacer estos niveles, la sociedad debe someter regularmente a evaluación a los propios servicios profesionales. Una parte del trabajo evaluativo, la que consiste en la regulación y protección del interés público, debe, obviamente, ser dirigida por organismos independientes, como agencias gubernamentales y juntas acreditadas. Pero, fundamentalmente, las evaluaciones más importantes de los servicios profesionales están realizadas (o encargadas) por los propios profesionales.

Para mantener sus servicios al día y asegurarse de que, efectivamente, satisfacen las necesidades de sus clientes, los profesionales deben recabar continuamente información oportuna sobre el trabajo evaluativo. Este proceso incluye estudiar las necesidades de sus clientes, evaluar los métodos que proponen o utilizan otros, seguir muy de cerca el desarrollo del servicio, valorar los resultados a corto y largo plazo y buscar modos de que los servicios sean cada vez más eficientes y efectivos. La primera pregunta que todo grupo evaluativo exterior debe plantear es si el profesional realiza periódicamente un programa de autoevaluación que responde oportunamente, con las correspondientes pruebas y evidencias, a estas cuestiones.

Desafortunadamente, el alcance de las evaluaciones que realizan los servicios profesionales –por medio de los propios profesionales y sus agentes acreditados– ha sido, tradicionalmente, y en el mejor de los casos, desigual y, en general, pobre. Algunos distritos escolares grandes y agencias gubernamentales mantienen sistemas de evaluación competentes, realizados por un personal adecuado, y los evaluadores de agencia han sabido ayudar a las organizaciones centrales a ser responsables con todos los implicados y a obtener la capacidad de planificar y administrar sus servicios. Pero incluso el más competente de estos sistemas evaluativos no puede asumir las responsabilidades evaluativas de todos los profesionales de la agencia. No debe olvidarse el hecho de que la evaluación es una responsabilidad tanto personal como institucional. Las oficinas de evaluación pueden cubrir las principales responsabilidades de una agencia y proporcionar preparación para el servicio y ayuda técnica al personal. Pero, a la larga, de todas maneras, cada profesional debe ser responsable de evaluar su propio trabajo.

## ¿QUÉ ES LA EVALUACIÓN?

Se han dado muchas respuestas a la pregunta: ¿qué es la evaluación? Mientras algunas de ellas varían sólo en aspectos menores, otras mantienen diferencias sustanciales. Por ejemplo, algunas definiciones identifican esencialmente la evaluación con otros conceptos como la investigación o las encuestas, mientras otros distinguen claramente la evaluación de otras materias. Una definición importante, y que se viene dando desde hace mucho tiempo, afirma que la evaluación supone comparar objetivos y resultados, mientras que otras exigen una conceptualización más amplia, apelando a un estudio combinado del trabajo en sí y de los valores. Muchas de las distintas definiciones formuladas aparecen en los diferentes modelos de evaluación presentados en las unidades 3-10. Además, la unidad 2 proporciona un análisis comparativo de estos modelos y definiciones. Esencialmente, ese análisis examina todo lo que se ha argumentado contra la definición que damos en esta unidad.

Esta definición es consecuente con lo que dice comúnmente el diccionario y, de una manera más específica, es la que ha adoptado el Joint Committee on Standards for Educational Evaluation. La definición es la siguiente: *la evaluación es el enjuiciamiento sistemático de la valía o el mérito de un objeto.*

Esta definición se centra en el término *valor* e implica que la evaluación siempre supone juicio. Desde el momento en que la evaluación, en un sentido óptimo, emplea procedimientos objetivos para obtener una información segura e imparcial, ya está ligada a la valoración. En efecto, su meta esencial es determinar el valor de algo que está siendo enjuiciado. De acuerdo con esta definición, si un estudio no informa de cuán buena o mala es una cosa, no se trata de una evaluación.

Algunos autores se han mostrado contrarios a esta posición. Creen que esta orientación valorativa conduce a interpretaciones radicales y subjetivas que conceden un poder innmercedo a los evaluadores e impiden el progreso.

Afirman, por así decirlo, que “la belleza reside en los ojos del observador” que las valoraciones de un grupo no deben tener precedencia sobre las de otro a la hora de determinar si algo es bueno o malo. También afirman que la evaluación es, demasiado a menudo, una fuerza negativa y, en general, mala psicología. Pero sus argumentos sólo nos convencen de que la mayoría de las evaluaciones no son determinantes y de que es difícil hacerlas bien; quizás existen muchas evaluaciones distintas, y posiblemente contradictorias en apariencia, de un objeto dado, dependiendo de la posición valorativa que se adopta. Pero, inevitablemente, la valoración debe desempeñar un papel. Y si sirve al progreso y se utiliza para identificar los puntos débiles y los fuertes, y para tender hacia una mejora, puede ser una fuerza positiva. Como puede verse, la evaluación es una empresa compleja.

Creemos que este aspecto de la complejidad no debe ser interpretado en el sentido de que estas evaluaciones sólo deben ser consideradas como procesos mentales que se producen de un modo natural, conduciéndonos así hacia una gran variedad de interpretaciones sobre el valor de algo. Al contrario, significa que hay que poner mucho cuidado no sólo cuando se recopile la información, que debe ser de calidad, sino también cuando se clarifique y proporcione una base lógica que justifique las perspectivas de valoración utilizadas para interpretar los resultados.

Como se puede ver, en la afirmación anterior incluimos la posibilidad de emplear muchas perspectivas de valoración. Consideramos que esta posibilidad es fundamental para los evaluadores y profesionales en general. A menudo, existen distintos públicos, que exigen o necesitan cosas distintas de un servicio profesional. Por ejemplo, a las familias

acaudaladas quizá les importe más el hecho de que las escuelas públicas preparen a los estudiantes para entrar en colegios de elite y en universidades, mientras que los nuevos inmigrantes, que son pobres y hablan muy poco inglés, quizá rechacen las escuelas que no ofrecen una educación bilingüe, no atienden a los niños atrasados y no promueven la integración social. A la hora de evaluar un distrito escolar en concreto, un evaluador competente debe tener en cuenta, según nuestra definición, entre otras cosas, estas dos perspectivas de valoración. De hecho, podríamos afirmar que las necesidades que unen o separan a los clientes de un determinado servicio son las que deben considerarse como base para decidir qué tipo de información se debe recopilar y a qué criterios recurrir para determinar la *valía* o el *mérito* de un servicio.

Los términos *valía* y *mérito* también deben ser definidos para enriquecer nuestro concepto de evaluación. Scriven (1981) ya ha indicado las trascendentales diferencias que hay entre ellos y el importante papel que desempeñan a la hora de determinar el valor de algo. Por ejemplo, un programa para formar a profesores de historia, ¿podrá producir profesores que enseñen historia a otros de un modo efectivo y de manera que se pueda confiar en ellos? En general, ¿hará bien lo que se supone que debe hacer? Si lo hace, será meritorio, pero puede que no sea válido. Por ejemplo, si la sociedad ya no necesita más profesores de historia porque ya tiene más de los que puede emplear, entonces aunque el programa sea muy bueno –y quizá sea mejor que otros parecidos–, continuar con el mismo nivel de expansión del programa quizá no sea la mejor manera de gastar sus escasos recursos. Aunque un programa meritorio puede no ser válido, debemos indicar que un programa que no sea meritorio no puede ser válido. Puede que no sea bueno formar malos profesores de Historia aunque exista una gran escasez de ellos.

Hasta aquí, esta discusión sobre qué bases de valoración deben ser tenidas en cuenta a la hora de evaluar algo, ha proporcionado tres importantes fuentes de criterios:

1. Las distintas expectativas de los *clientes*.
2. El *mérito* o excelencias del servicio en cuestión.
3. Hasta qué punto el servicio es *necesario* (es decir, su validez potencial).

En general, recomendamos que se tengan en cuenta los tres aspectos cuando se realice una evaluación. Pero creemos necesario considerar dos aspectos más: la *viabilidad* y la *equidad*. Aunque un servicio pueda ser de mucha calidad, atienda los diferentes objetivos de distintos grupos de clientes y se centre en un área que de verdad lo necesite, quizá falle en lo que se refiere a los valores de la viabilidad y la equidad.

Por ejemplo, puede que consuma más recursos de los que necesita o puede causar infinidad de problemas políticos. En cualquiera de los dos casos, el programa, en principio, debe ser modificado en esas áreas para hacerlo más viable. Obviamente, una buena evaluación del servicio debería mencionar esto y, en su caso, proporcionar una solución para que el servicio sea más sencillo, utilice el tiempo y los recursos de un modo más apropiado y sea políticamente viable. Este argumento en favor de la viabilidad es, en principio, aplicable a todas las sociedades.

La última base valorativa que vamos a mencionar aquí, la *equidad*, está predominantemente relacionada con las sociedades democráticas. Se exigen las mismas oportunidades para todo el mundo y se reclama la libertad para todos. En los Estados Unidos, una evaluación puede resultar incompleta si no tiene en cuenta que un servicio público debe estar pensado y ser asequible para todos los miembros de la sociedad. Es importante indicar que nuestro concepto de equidad es complejo. No es suficiente decir que los servicios públicos educativos deben poder ser requeridos y usados por todo el mundo.

Como ha dicho Kellaghan (1982), para que exista verdaderamente la igualdad deben existir siete indicios de equidad:

1. Los servicios públicos educativos de una sociedad deben estar *pensados* para toda la población.
2. Todos los sectores de la sociedad deben tener el mismo tipo de *acceso* a los servicios.
3. Todos los grupos deben disfrutar de una *participación* equitativa en la utilización de los servicios.
4. Los niveles de *logro*, por ejemplo, los años en el sistema educativo, no serán significativamente distintos para los diferentes grupos.

5. Los *conocimientos* exigidos para alcanzar los objetivos del servicio no deben ser sustancialmente diferentes para los distintos grupos.
6. Las *aspiraciones* vitales que se persiguen deben ser similares en todos los grupos sociales.
7. El sistema del servicio deberá producir *resultados* similares en lo que se refiere a la mejora en el bienestar de todos los grupos sociales.

De la discusión anterior, se puede colegir que la definición de evaluación que hemos propuesto es engañosa en su aparente simplicidad. Cuando examinemos seriamente el término base *valor*, debemos inevitablemente tomar en consideración tanto las perspectivas de valoración como la información, la combinación de ambas con el fin de conseguir averiguar el valor de algo y, por último, la transformación de los resultados en conocimientos operativos que se utilizarán para suministrar o no suministrar (o para comprar o no comprar) ciertos servicios y para mejorar los que ya se están ofreciendo.

Antes de terminar con este tema de la definición de evaluación, debemos considerar otros tres aspectos de la evaluación. El primero se refiere a si la evaluación necesita ser comparativa; el segundo, a las principales utilidades de la evaluación; y el tercero se pregunta hasta qué punto los evaluadores necesitan cualificaciones especiales.

Algunos autores (Stake, Scriven, Cronbach) han indicado que la evaluación puede centrarse en un solo producto o servicio, o compararlo con otras alternativas válidas. Por otra parte, han debatido los méritos de estas dos posturas, para acabar quedándose con una de ellas. Nosotros creemos que, dependiendo de las circunstancias, una evaluación puede ser, legítimamente, tanto comparativa como no comparativa.

Algo que se debe tener muy en consideración es la naturaleza de la audiencia y qué es lo que necesita saber acerca del objeto de la evaluación. Si la audiencia está compuesta por consumidores que desean escoger un producto o servicio, la evaluación debe ser comparativa y ayudar a los consumidores a saber qué alternativas son convenientes y cómo compararlas críticamente. Si, por el contrario, la audiencia incluye promotores o consumidores que ya están comprometidos con el desarrollo o el uso de un programa determinado, es mejor que la evaluación se centre intensamente en la evolución del programa para intentar mejorarlo. A veces sucede, sin embargo, incluso si un grupo es

totalmente devoto de cierto servicio o producto, que, al considerar otras opciones, obtiene mejores resultados de su proveedor o encuentra una alternativa mejor.

En general, creemos que la evaluación debe ser comparativa antes de que empiece el proceso o antes de suscribirse a un servicio, no comparativa durante el proceso (para que permita concentrarse en lo que resulte mejor para el desarrollo del servicio), y periódicamente comparativa después del proceso o la utilización continuada, para que el camino quede abierto ante posibles mayores avances.

Hablando en general, que una evaluación deba ser comparativa o no depende de lo que se quiera conseguir de esa evaluación, en cuanto a su utilización se refiere. Nosotros consideramos tres utilidades principales: perfeccionamiento, recapitulación y ejemplificación. La primera utilización se refiere a la información que se proporciona para asegurar la calidad de un servicio o para mejorarla. Hay que prestar mucha atención a la naturaleza y las necesidades de los consumidores, y a la relación existente entre el proceso y los resultados. En este punto, los evaluadores deberían estar muy en contacto con el personal del programa y servir como guías para las tomas de decisiones. El plan de evaluación necesita ser flexible y receptivo. Cuando la meta principal es mejorar algo, la evaluación debería parecerse más al estudio de un caso que a un experimento comparativo.

La segunda función más importante de la evaluación es hacer recapitulaciones o informes sumarios. Se trata de análisis retrospectivos acerca de proyectos ya terminados, programas ya hechos o productos ya acabados. La información no va dirigida al personal encargado de su desarrollo, sino a los patrocinadores y consumidores. Debe incluir una valoración de la eficacia de lo invertido, siendo también importantes las comparaciones con los más importantes competidores. La información que se deriva de estos profundos estudios y de las encuestas comparativas es muy interesante para la audiencia en este tipo de situaciones.

La tercera utilización es la ejemplificación. Básicamente, la evaluación y la investigación son actividades distintas. La primera se dedica a considerar todos los criterios que se utilizan en la determinación de un valor, mientras que la segunda se limita más bien al estudio de ciertas variables que tienen un interés teórico. Las evaluaciones suelen utilizar métodos subjetivos y no están tan estrechamente controladas ni sujetas a manipulación

como las investigaciones. No obstante, los resultados obtenidos, en un cierto período de tiempo, evaluando servicios pueden facilitar información que luego será utilizada para crear y ensayar una teoría. Y, sin duda, los resultados de las evaluaciones a menudo deberían conducir y conducen, de hecho, a unas conclusiones más concretas e interesantes en el terreno de la investigación. Por lo tanto, creemos que es importante que los evaluadores, al planificar sus estudios, tengan en cuenta que sus conclusiones pueden ser utilizadas en el contexto de una investigación que se proponga una comprensión más amplia de los fenómenos examinados en la evaluación. Con una planificación más minuciosa y cuidadosa, las evaluaciones pueden servir no sólo como guía de programas operativos, o como resumen de sus contribuciones, sino también para investigaciones concretas o cuestiones teóricas.

Así pues, parece claro que la evaluación es un concepto altamente complicado. A la hora de planificar o dirigir una evaluación, deben tenerse en cuenta una serie de cuestiones:

- ¿Qué valores deben ser mencionados al juzgar los resultados?
- ¿Qué criterios deben ser adoptados y cuáles deben tener preferencia?
- ¿La evaluación debe ser comparativa?
- ¿La evaluación debe estar capacitada para guiar el desarrollo, responsabilizar a los profesionales, ordenar las opciones o promover nuevas intuiciones acerca de los fenómenos implicados?

Si se tienen en cuenta estas cuestiones con el fin de que la audiencia confíe en los resultados y los utilice, los evaluadores deben mantener ciertos criterios o normas en su trabajo. Como consecuencia, creemos que la distinción entre evaluaciones profesionales y evaluaciones informales es importante. Desde luego, todo el mundo recurre a las segundas cuando se trata de emplear, opinar, aceptar o comprar cosas. Pero ocurre demasiado a menudo que nuestras evaluaciones resultan poco o nada sistemáticas y en modo alguno convincentes, ni siquiera para nosotros mismos. En muchos casos, preferiríamos que nuestras conclusiones tuvieran más relevancia, rigor y objetividad. Estos requisitos, sin embargo, requerirían a menudo más tiempo, dinero y habilidad de los que nosotros tenemos.



Así, en muchas ocasiones, las buenas evaluaciones deben ser realizadas por un personal especializado. Aunque nuestra definición afirma que la evaluación es sistemática, muchas veces, sin duda, no lo es. Lo que intentamos proponer en este libro es una especie de evaluación que sea justificable. La clase de evaluación que tenemos en mente requiere un esfuerzo sistemático por parte de una o más personas que posean, en principio, un cierto nivel mínimo en lo que se refiere a las técnicas evaluativas que se exigen en estos casos. Pensamos que el concienzudo estudio que propone este libro puede ayudar al estudiante a alcanzar el nivel básico de habilidad que se necesita para llevar a buen término nuestro concepto de *evaluación sistemática* y para evaluar las evaluaciones realizadas por otros.

## Métodos

Un aspecto que diferencia la evaluación sistemática de la evaluación informal es, sin lugar a dudas, el área metodológica. Cuando pasamos de las evaluaciones realizadas con juicios rápidos e intuitivos a las que suponen una rigurosa recopilación de resultados que deben ser considerados de un modo efectivo, debemos necesariamente enfrentarnos a las complejas áreas de la epistemología, las reglas empíricas, las ciencias informativas, los proyectos de investigación, las valoraciones, las estadísticas, la comunicación, etc. Muchos principios, instrumentos y estrategias de estas áreas son importantes para la evaluación sistemática. El evaluador bien preparado debe dominar convenientemente todas las técnicas de estas áreas y mantenerse informado acerca de los avances tecnológicos que puedan resultarle útiles. Estos evaluadores, que pueden convertirse en pioneros y contribuir al desarrollo de su profesión, suelen aumentar las críticas de los métodos ya existentes y la aparición de otros nuevos.

A lo largo de los años, muchos evaluadores han escogido, e incluso defendido, la utilización exclusiva de unas pocas técnicas. Algunos han equiparado la evaluación a sus métodos favoritos, como el diseño experimental, las pruebas estandarizadas o las visitas personales. Otros han atacado duramente la estrechez de miras respecto de la adecuación de los métodos y, en algunos casos, han abogado por sustituir su técnica favorita, como por ejemplo el estudio de un caso. Nosotros creemos que ambas posiciones son ineficaces, inadecuadas y, a veces, parciales.

A cambio, recomendamos un método ecléctico. Así, creemos que los evaluadores deberían conocer un amplio espectro de técnicas adecuadas y saber cómo aplicarlas en los distintos

contextos evaluativos. De este modo, podrían valorar, en cada situación, qué técnicas son potencialmente aplicables y cuáles pueden servir de un modo más eficaz a los propósitos particulares de la evaluación en cuestión.

Entre las áreas técnicas que, según creemos, deben conocer a fondo los evaluadores profesionales se encuentran las siguientes: las entrevistas, los informes preliminares, el análisis de contenidos, la observación, el análisis político, el análisis económico, el examen investigativo, los informes técnicos, el estudio del caso, la evaluación sin metas, la escucha de opiniones contrapuestas, las asociaciones de defensa, las listas de control, la elaboración de prueba, el análisis estadístico, el diseño de la investigación, el análisis sistemático, la teorización y la administración del proyecto. Es recomendable consultar las fuentes de información general que, acerca de estas áreas técnicas, pueden encontrarse en Scriven (1974); Anderson, Ball y Murphy (1974); Brinkerhoff y otros (1983) y Smith (1981a y 1981b).

## **Normas**

Una de las características de las prácticas profesionales es que sus miembros deben cumplir exigentes normas referentes al trabajo y al servicio que realizan. Por lo general, estas normas ya vienen formuladas por los miembros de la profesión, y, en algunos casos, por el gobierno. Ejemplos típicos de ello son las normas existentes en las áreas de la ley, la medicina y la contabilidad, con el fin de asesorar y valorar la práctica en cuestión, y, en algunos casos, llegar a conclusiones acerca de ciertos procedimientos ilegales. Hasta muy recientemente, los evaluadores de programas no disponían de normas explícitas para la evaluación, pero esta situación está cambiando.

Recientemente, se han publicado dos conjuntos de normas para las evaluaciones. Uno consiste en un grupo de treinta normas formuladas por el Joint Committee on Standards for Educational Evaluation. El otro es una serie de cincuenta y cinco normas formuladas por la Evaluation Research Society (ERS).

Aquí vamos a centrarnos en el primero por unas razones determinadas. Cuidadosas comparaciones realizadas por investigadores independientes han demostrado que casi todo lo que aparece en las normas de la ERS es examinado con mayor profundidad en las Normas del Joint Committee. Es más, la primera serie ha sido elaborada por un

comité cuyos diecisiete miembros han sido nombrados por doce organizaciones profesionales que abarcan un total de alrededor de dos millones de miembros, mientras que las normas de la ERS han sido elaboradas por un grupo de alrededor de dos mil miembros. Otra diferencia clave es que las *Normas* del Joint Committee incluyen tanto al cliente como al que realiza los servicios evaluativos.

El texto *Standards for Evaluations of Educational Programs, Projects, and Material* (Joint Committee, 1981) fue elaborado en un período de cinco años. Este comité está formado por alrededor de doscientas personas relacionadas con la práctica profesional de la evaluación, que realizan la labor de crear, valorar y clarificar principios ampliamente compartidos con el fin de que sirvan como base para evaluaciones sobre programas educativos. El resultado son las treinta normas resumidas en la tabla 1-1. Es particularmente notable el continuo proceso que llevan a cabo reexaminando y revisando las normas. En esencia, el Joint Committee concluye afirmando que las treinta normas son necesarias y suficientes para asegurar que las evaluaciones sean útiles, factibles, precisas y válidas. Se aconseja al lector que estudie en profundidad ese libro de *Normas*.

En general, las *Normas* del Joint Committee aconsejan que los evaluadores y los clientes cooperen entre sí para que las evaluaciones puedan cumplir estas cuatro condiciones principales:

1. Una evaluación debe ser *útil*. Debe estar dirigida a aquellas personas y grupos que estén relacionados con la tarea de realizar aquello que se está evaluando o que sean sus más directos responsables. Debe ayudarles a identificar y examinar lo bueno y lo malo de este objeto. Debe poner muchísimo énfasis en plantearles las cuestiones de mayor importancia. Debe proporcionarles informes claros de una manera oportuna. Y, en general, debe facilitar no sólo informaciones acerca de virtudes y defectos, sino también soluciones para mejorarlo todo.
2. Debe ser *factible*. Debe emplear procedimientos evaluativos que puedan ser utilizados sin demasiados problemas. Debe tomar en consideración y aplicar controles razonables sobre aquellas fuerzas políticas que puedan, de alguna manera, inmiscuirse en la evolución. Y deben ser dirigidas de un modo eficiente.
3. Debe ser *ética*. Debe estar basada en compromisos explícitos que aseguren la necesaria cooperación, la protección de los derechos de las partes implicadas y la

honradez de los resultados. Además, debe proporcionar un informe equitativo que revele todas las virtudes y defectos del objeto.

4. Debe ser *exacta*. Debe describir con claridad el objeto en su evolución y en su contexto. Debe revelar las virtudes y defectos del plan de evaluación, de los procedimientos y de las conclusiones. Debe estar libre de influencias. Y debe proporcionar unas conclusiones válidas y fidedignas.

TABLA 1-1. *Resumen del texto “Standards for Evaluations of Educational Programs, Projects, and Materials”*

---

## **A    *Normas de utilidad***

Las normas de utilidad están concebidas para asegurar que una evaluación proporcione las informaciones prácticas que necesita una audiencia determinada. Estas normas son:

### **A1    *Descripción de la audiencia***

La audiencia relacionada con la evaluación o afectada por ella debe ser descrita con el fin de que todas sus necesidades puedan ser satisfechas.

### **A2    *Credibilidad del evaluador***

Las personas que dirigen la evaluación deben ser honradas y a la vez competentes para poder llevarla a cabo, con el fin de que sus conclusiones alcancen el máximo de credibilidad y aceptación.

### **A3    *Alcance y selección de la información***

La información recopilada debe tener el suficiente alcance y haber sido seleccionada de tal manera como para incluir en sí misma las cuestiones pertinentes acerca del objeto de la evaluación y responder a las necesidades e intereses de audiencias concretas.

**A4 Interpretación valorativa**

Las perspectivas, procedimientos y bases lógicas utilizadas para interpretar los resultados deben ser cuidadosamente descritas, con el fin de que queden claras las bases para el juicio valorativo.

**A5 Claridad del informe**

El informe de la evaluación debe describir el objeto que se está evaluando y su contexto, a la vez que los propósitos, procedimientos y resultados de la evaluación, con el fin de que las audiencias comprendan en seguida qué se ha hecho, cómo se ha hecho, qué tipo de información se ha obtenido, qué conclusiones se han sacado y qué recomendaciones se hacen.

**A6 Difusión del informe**

Los resultados de la evaluación deben ser dados a conocer a los clientes y a otras audiencias interesadas o adecuadas, para que puedan valorar y utilizar las conclusiones.

**A7 Oportunidad del informe**

La difusión de los informes debe ser oportuna, con el fin de que las audiencias puedan hacer el mejor uso posible de la información recopilada.

**A8 Impacto de la evaluación**

Las evaluaciones deben ser planificadas y dirigidas de manera que animen a la audiencia a continuar investigando por sí misma.

**B Normas de viabilidad**

Las normas de viabilidad están concebidas para asegurar que una evaluación sea realista, prudente, diplomática y moderada. Son éstas:

**B1** *Procedimientos prácticos*

Los procedimientos evaluativos deben ser prácticos, para que los problemas sean mínimos y pueda obtenerse la información necesaria.

**B2** *Viabilidad política*

La evaluación debe ser planificada y dirigida conociendo las distintas posturas de los diferentes grupos de interés, con el fin de obtener su cooperación, y para que cualquier posible intento, por parte de alguno de estos grupos, de abreviar las operaciones evaluativas o de influir en los resultados o utilizarlos, pueda ser prevenido o contrarrestado.

**B3** *Eficacia de los costes*

La evaluación debe producir información con el valor suficiente como para justificar lo que ha costado.

**C** **Normas de honradez**

Las normas de honradez están concebidas para asegurar que una evaluación sea realizada legal y éticamente, y, a la vez, con el debido respeto por el bienestar de todos los que intervienen en ella, así como de aquellos que están interesados en las conclusiones.

**C1** *Obligación formal*

Las obligaciones de las partes formales de una evaluación (qué hay que hacer, cómo hacerlo, quién debe hacerlo y cuándo) deben ser plasmadas en un compromiso escrito, con el fin de que estas partes estén obligadas a cumplir todas las condiciones del compromiso o a renegociarlo formalmente.

**C2** *El conflicto de intereses*

El conflicto de intereses, frecuentemente inevitable, debe ser afrontado abierta y honestamente, con el fin de que no interfiera en el proceso evaluativo y sus resultados.

---

**C3** *Informes completos y claros*

Los informes orales o escritos de una evaluación deben ser francos, directos y honestos en la revelación de las conclusiones pertinentes, incluyendo las limitaciones de la evaluación.

**C4** *El derecho del público a saber*

Las partes formales de una evaluación deben respetar y asegurar el derecho del público a saber, dentro de los límites de unos principios y estatutos determinados, como por ejemplo la seguridad del público y el derecho a la intimidad.

**C5** *Los derechos humanos*

Las evaluaciones deben ser preparadas y dirigidas de tal manera que sean respetados y protegidos los derechos y el bienestar de los seres humanos.

**C6** *Relaciones humanas*

Los evaluadores deben respetar la dignidad y la integridad humanas en su trato con otras personas relacionadas con la evaluación.

**C7** *Información equilibrada*

La evaluación debe ser completa y honrada en la presentación de las virtudes y defectos del objeto de la investigación, para que las virtudes puedan ser mejoradas y los problemas solucionados.

**C8** *Responsabilidad fiscal*

La cuota del evaluador y los gastos efectuados deben reflejar los procedimientos económicos del trabajo y, por otra parte, ser prudentes y éticamente responsables.

## **D Normas de precisión**

Las normas de precisión están concebidas para asegurar que una evaluación revele y transmita una información técnicamente exacta acerca de las características del objeto que está siendo estudiado, para poder determinar su valor o su mérito. Estas normas son:

### ***D1 Identificación del objeto***

El objeto de la evaluación (programa, proyecto o material) debe ser examinado en profundidad, de manera que su(s) estructura(s) pueda(n) ser claramente identificada(s).

### ***D2 Análisis del contexto***

El contexto en el que está incluido el programa, proyecto o material debe ser examinado en su más mínimo detalle, con el fin de identificar sus posibles influencias sobre el objeto.

### ***D3 Descripción de intenciones y procedimientos***

Las intenciones y los procedimientos de la evaluación deben ser concebidos y descritos con todo detalle, de manera que puedan ser identificados y valorados.

### ***D4 Fuentes de información fiables***

Las fuentes de información deben ser descritas con todo detalle, para que pueda valorarse si la información es suficiente.

### ***D5 Determinar la validez***

Los instrumentos y procedimientos para la recopilación de información deben ser escogidos o creados y luego utilizados de manera que aseguren que la interpretación final es válida para su uso.

### ***D6 Determinar la exactitud***

Los instrumentos y procedimientos para la recopilación de información deben ser escogidos o creados y luego utilizados de manera que aseguren que la información obtenida es lo suficientemente exacta para ser utilizada debidamente.



*D7 Control sistemático de los datos*

Los datos recopilados, sistematizados y finalmente presentados en una evaluación deben ser revisados y corregidos, con el fin de que los resultados de la evaluación no sean inexactos.

*D8 Análisis de la información cuantitativa*

La información cuantitativa de una evaluación debe ser adecuada y sistemáticamente analizada con el fin de asegurar que las interpretaciones sean fidedignas.

*D9 Análisis de la información cualitativa*

La información cualitativa de una evaluación debe ser adecuada y sistemáticamente analizada con el fin de asegurar que las interpretaciones sean fidedignas.

*D10 Conclusiones justificadas*

Las conclusiones a las que se llegue en una evaluación deben ser explícitamente justificadas, para que las audiencias puedan valorarlas.

*D11 Objetividad del informe*

Los procedimientos de las evaluaciones deben proporcionar las garantías suficientes para proteger los resultados y los informes de la evaluación contra la distorsión de los sentimientos e inclinaciones personales de cualquiera de las partes formales.

Las *Normas* del Joint Committee pueden aplicarse en cualquier etapa del proceso evaluativo: deben ser tenidas en cuenta, por ejemplo, a la hora de decidir si hay que realizar la evaluación o no, a la hora de planear la evaluación y llevarla a cabo, al presentar los resultados y al aplicar las conclusiones. Por lo tanto, la utilización de las *Normas* no sólo sirve para controlar la calidad de un informe ya terminado, sino también como asesoramiento durante la realización de la evaluación. Un segundo punto que hay que tener en cuenta es que las distintas normas son más o menos importantes en cada etapa de la evaluación. Así, quienes utilicen las normas deben estudiarlas y analizarlas comparativamente con el fin de poder aplicarlas juiciosamente en cada una de las etapas de la evaluación. El resumen de las treinta normas presentado en la tabla 1-1 es un punto de partida, pero recomendamos al lector que consiga y estudie el texto completo.

El trabajo del Joint Committee se ha basado en dos premisas básicas: que la evaluación es una actividad humana tan esencial como inevitable, y que la evaluación solvente debe proporcionar una comprensión más amplia y una mejora de la educación. Además, el comité se guio por la creencia de que una serie de normas concienzudamente elaboradas puede desempeñar un papel vital a la hora de mejorar la práctica de la evaluación educativa. En otras palabras: un papel metaevaluativo.

Otro punto trascendental con respecto a las *Normas* del Joint Committee es que se trata de un documento vivo. Un comité permanente, formado por doce organizaciones y respaldado por el Centro de Evaluación de la Universidad de Western Michigan, controla la utilización de las normas y carga con la responsabilidad de ponerlas al día y perfeccionarlas periódicamente. El comité anima a los usuarios para que les proporcionen información basada en críticas y, especialmente, en los usos más actuales de las normas.

## La historia<sup>1</sup>

Ninguna introducción a la evaluación entendida como una práctica profesional estará completa si no se presta la debida atención a la evolución histórica del tema. Cualquier profesión, para poder satisfacer las necesidades de sus clientes, debe evolucionar de acuerdo con los cambios de las necesidades sociales y teniendo en cuenta todos los

---

1. Este apartado sobre la historia de la evaluación está basado en un análisis de George Madaus y Daniel Stufflebeam, que aparece resumido en George MADAUS, Michael SCRIVEN y Daniel STUFFLEBEAM. *Evaluation Models*. Boston: Kluwer-Nijhoff, 1983

avances teóricos y técnicos. A menos que los miembros de una profesión elaboren y sostengan una perspectiva histórica sobre su trabajo, probablemente seguirán teniendo siempre una concepción limitada de su función social, sin poder estimular la innovación ni contribuir a ella. Se dice a menudo que los que no conocen su propia historia están condenados a repetirla. Nosotros facilitaremos únicamente un breve esbozo histórico con el fin de informar al lector sobre los acontecimientos más significativos ocurridos en el campo de la evaluación.

Nuestro análisis histórico se basa en un trabajo clave de Ralph W. Tyler (véase unidad 3), que a menudo es considerado el padre de la evaluación educativa. Utilizando sus contribuciones iniciales como principal punto de referencia, hemos establecido cinco períodos básicos: 1) el período pre-Tyler, que abarca hasta 1930; 2) la época tyleriana, que va desde 1930 hasta 1945; 3) la época de la “inocencia”, desde 1946 hasta 1957; 4) la época del realismo, que cubre el período 1958-1972; y 5) la época del profesionalismo, que incluye la evolución experimentada desde 1973 hasta el presente.

## **El período pretyleriano**

La evaluación sistemática no era desconocida antes de 1930, pero tampoco era un movimiento reconocible. El hecho de evaluar individuos y programas apareció, como mínimo, en el año 2000 a. C., cuando algunos oficiales chinos dirigieron unas investigaciones de los servicios civiles, y en el siglo V a. C., cuando Sócrates y otros maestros griegos utilizaron cuestionarios evaluativos como parte de su metodología didáctica. En el siglo XIX, en Inglaterra, se utilizaron comisiones reales para evaluar los servicios públicos. En 1845, en los Estados Unidos, Horace Mann dirigió una evaluación, basada en tests de rendimiento, para saber si las escuelas de Boston educaban bien a sus estudiantes. Entre 1887 y 1898, Joseph Rice estudió los conocimientos en ortografía de 33.000 estudiantes de un amplio sector escolar y sacó la conclusión de que la gran insistencia en la enseñanza de la ortografía, que entonces estaba en boga, no había producido avances sustanciales en el aprendizaje. Este estudio está generalmente reconocido como la primera evaluación formal sobre un programa educativo realizado en América. También en los últimos años del siglo pasado, empezó a crearse el movimiento para la acreditación de instituciones educativas y programas en los Estados Unidos; desde entonces, el proceso de acreditación se ha convertido en uno de los principales

medios para evaluar los servicios educativos. Una aproximación aun mayor a la evaluación se produjo, en los primeros años de este siglo, con la aparición de los tests estandarizados. Este acontecimiento formaba parte de un gran esfuerzo que pretendía que la educación fuera más eficiente para que un mayor número de estudiantes pudiera utilizar los recursos entonces disponibles en las escuelas.

El breve resumen anterior viene a confirmar que la evaluación sistemática tiene una larga historia. Mientras que sólo muy recientemente se ha llegado a considerar la evaluación como una práctica profesional, buena parte del trabajo evaluativo moderno continúa basándose en ideas y técnicas que ya se aplicaban hace mucho tiempo, como por ejemplo las encuestas, las comisiones la acreditación y las comparaciones experimentales de varios programas.

## La época tyleriana

En los primeros años de la década de los treinta, Ralph Tyler acuñó el término *evaluación educacional* y publicó una amplia y renovadora visión del currículo\* y la evaluación. En un período de alrededor de quince años, desarrolló sus intuiciones hasta que conformaron un método que suponía una clara alternativa a otras perspectivas.

La principal característica de su método era que se centraba en unos objetivos claramente fijados. De hecho, definía la evaluación como algo que determina si han sido alcanzados ciertos objetivos. Como consecuencia de esta definición, los evaluadores debían ayudar a quienes elaboraban currículos para clarificar el rendimiento de los estudiantes, que había sido precisamente determinado a través de la realización de un currículo. Los objetivos resultantes, referidos al rendimiento, proporcionaban entonces la base para elaborar un currículo y un test de desarrollo. La elaboración del currículo quedaba así influida tanto por los contenidos que debían aprenderse como por la evolución del rendimiento de los estudiantes. La tecnología de los tests de desarrollo iba a servir de base para la elaboración de tests relacionados con los objetivos, y también a los relacionados con las diferencias individuales y con las normas nacionales o del estado.

---

\* Dado el significado amplio que tiene la palabra "curriculum" en inglés, utilizada en general para designar todo aquello que gira en torno a la actividad docente y no sólo al plan de estudios, hemos preferido conservarla, una vez castellanizada, a lo largo de toda la obra. (T.)

Durante la década de los treinta, los Estados Unidos, como el resto del mundo, cayó en las profundidades de la Gran Depresión. Las escuelas, como otras instituciones públicas, se estancaron en una ausencia total de recursos y optimismo. Justo cuando Roosevelt intentaba con su New Deal sacar de este abismo a la economía americana, John Dewey y algunos otros intentaron contribuir a que la educación se convirtiera en un sistema dinámico e innovador, que a la vez se fuera renovando a sí mismo. Con el nombre de Educación Progresiva, este movimiento reflejó la filosofía del pragmatismo y empleó los instrumentos de la psicología conductista.

Tyler se vio metido de lleno en este movimiento cuando le encargaron dirigir la parte de investigación del hoy famoso Eight-Year Study (Smith y Tyler, 1942). Este estudio fue pensado para examinar la efectividad de ciertos currículos renovadores y estrategias didácticas que se estaban empleando en treinta escuelas a lo largo de toda América. El estudio resulta notable porque ayudó a Tyler a difundir, ensayar y demostrar su concepción de la evaluación educacional de una manera inmediata.

A través de este programa de ámbito nacional, Tyler pudo publicar lo que él veía como un método que tenía claras ventajas sobre los demás. Desde el momento en que la evaluación tyleriana incluía comparaciones internas entre los resultados y los objetivos, ya no necesitaba recurrir a costosas y problemáticas comparaciones entre grupos de control y grupos experimentales. El método se centra en determinar el grado de éxito, por lo que se opone a métodos indirectos que determinan factores como la calidad de la enseñanza, el número de libros que hay en la biblioteca, los materiales y la participación de la comunidad. Y además, las evaluaciones tylerianas no se ven demasiado afectadas por la fiabilidad de las diferencias entre las puntuaciones individuales de los alumnos ya que generalmente cubren una franja más amplia de variables exógenas que las cubiertas por los tests tipificados. Todos estos argumentos fueron bien recibidos por el sector de la educación americana, y a mediados de los años 40 Ralph Tyler ya había conseguido la categoría suficiente como para ejercer una gran influencia sobre el panorama educacional en los próximos veinticinco años.

## **La época de la “inocencia”**

Para la sociedad americana, los últimos años 40 y la década de los 50 fue un tiempo para olvidar la guerra, dejar atrás la Depresión, crear y desarrollar aptitudes, adquirir recursos, y lograr disfrutar de la “buena vida”. Podríamos haber calificado esta época como el

período de expansión, pero hay que tener en cuenta que también existía una amplia complacencia a la hora de enjuiciar los graves problemas de aquella sociedad. Como consecuencia, creemos que es mejor referirnos a esta época como al Período de la «inocencia» o incluso de la *irresponsabilidad social*.

Además de ser una época de plenitud, este período fue un tiempo de pobreza y desesperación en las ciudades del interior y en las áreas rurales, pero al parecer casi nadie, excepto los afectados, se dio cuenta. Fue un período de profundos prejuicios raciales y de segregación, pero gran parte de la población blanca parecía inconsciente de esta enfermedad. Fue un período de exorbitado consumo y de gran despilfarro de los recursos naturales, sin tener en cuenta que un día de estos recursos iban a agotarse. Fue un período de un abrumador desarrollo de la industria y la capacidad militar, pero se tomaron muy pocas precauciones contra los perjuicios que iban a sufrir el paisaje y las futuras generaciones.

Más que de la evaluación educacional, se trató de una expansión de las ofertas educacionales, del personal y de las facilidades. Se construyeron nuevos edificios. Aparecieron nuevos tipos de instituciones educativas, como los colegios comunitarios. Los pequeños distritos escolares se unieron a otros con el fin de proporcionar un más amplio abanico de servicios educativos, como ya lo hacían los sistemas escolares más grandes, incluyendo servicios para la salud mental y física, consejos escolares, comedores; enseñanzas musicales, amplios programas deportivos, educación comercial y técnica y educación social. Se incrementaron desmesuradamente las inscripciones en los programas educativos para maestros y, en general, las inscripciones en los colegios aumentaron espectacularmente.

Este panorama general de la sociedad y la educación se reflejó también en la evaluación educacional. Mientras se producía esta gran expansión de la educación, la sociedad no parecía tener un gran interés en formar profesores competentes, en localizar y solucionar las necesidades de los menos privilegiados o en identificar y solucionar los problemas del sistema educativo. Los educadores escribían sobre la evaluación y sobre la recopilación de los datos más importantes, pero, al parecer, no traducían estos esfuerzos en intentos de mejorar los servicios educativos. Esta falta de objetivos provocó también el retraso de los aspectos técnicos de la evaluación. Hubo un considerable desarrollo de los instrumentos y estrategias aplicables a los distintos métodos evaluativos, como los tests,

la “experimentación comparativa” y la “coincidencia entre resultados y objetivos”. Como consecuencia, los educadores dispusieron de nuevos servicios de tests y de nuevas maneras de calificarlos, de algoritmos para designar objetivos de comportamiento, de taxonomías de objetivos, de nuevos modelos experimentales y de nuevos procedimientos estadísticos para analizar los datos educativos. Pero estas contribuciones no se derivaban de un análisis de la información necesaria para valorar y perfeccionar la educación ni representaban un desarrollo de la experiencia escolar.

Durante este período, las evaluaciones educacionales dependían, como ya lo habían hecho antes, sobre todo, de los límites de los distritos escolares locales. Las escuelas, pues, dependían del interés y la habilidad locales para poder hacer o no una evaluación. Las agencias federales y estatales aún no estaban demasiado introducidas en el campo de la evaluación de programas. Los fondos destinados a las evaluaciones provenían de las arcas locales, de fundaciones o de organizaciones profesionales. Esta ausencia de ayudas exteriores y de apoyos para la realización de evaluaciones en todos los niveles de la educación iba a concluir con la llegada del próximo período de la historia de la evaluación.

## **La época del realismo**

La época de la inocencia en la evaluación tuvo un brusco final con el auge, a finales de los años 50 y principios de los 60, de las evaluaciones de proyectos de currículos a gran escala, financiadas por estamentos federales. Los educadores se encontraron durante este período, con que ya no podían hacer o no hacer las evaluaciones a su antojo, y que los futuros progresos de las metodologías evaluativas deberían estar relacionados con conceptos como utilidad y relevancia. Su violento despertar, en este período, marcó el final de una era de complacencia y ayudó a emprender profundos cambios guiados por el interés público y dependiendo del dinero de los contribuyentes para su financiación, lo cual condujo a que la evaluación se convirtiera en una industria y en una profesión.

Como consecuencia del lanzamiento, por parte de la Unión Soviética, del Sputnik 1 en 1957, el gobierno federal respondió promulgando la National Defense Education Act de 1958. Entre otras cosas, esta acta proporcionó nuevos programas educativos en matemáticas, ciencias e idiomas y creó servicios asesores y programas de encuestas en los distritos escolares. Aparecieron, así, un cierto número de nuevos proyectos

nacionales para el desarrollo de currículos, especialmente en las áreas de ciencias y matemáticas. Finalmente, se asignaron fondos para evaluar estos programas.

Los cuatro métodos evaluativos expuestos hasta aquí estaban ya representados en las evaluaciones que se realizaron en este período. Primero, el método Tyler fue usado para definir los objetivos de los nuevos currículos y para valorar el grado en que estos objetivos quedaban, más tarde, alcanzados. Segundo, los nuevos tests estandarizados de carácter nacional sirvieron para reflejar mejor los objetivos y los contenidos de los nuevos currículos. Tercero, el método del criterio profesional fue usado para valorar las propuestas y verificar periódicamente los resultados. Finalmente, muchos evaluadores se comprometieron a evaluar los resultados del desarrollo del currículo a través de pruebas concretas.

A principios de los años 60, algunos de los más importantes evaluadores educacionales se dieron cuenta de que su trabajo y sus conclusiones no ayudaban demasiado a los que elaboraban currículos, y no respondían a las preguntas sobre los programas que planteaban aquellos que necesitaban comprobar su grado de efectividad. La “flor y nata” de la comunidad evaluativa educacional participó en estos esfuerzos para evaluar los nuevos currículos; estaban adecuadamente financiados y aplicaban cuidadosamente la tecnología que había ido desarrollándose durante la pasada década o incluso antes. A pesar de todo esto, empezaron a darse cuenta de que sus esfuerzos no tenían éxito.

Esta valoración negativa quedó perfectamente reflejada en un histórico artículo de Cronbach (1963: véase unidad 5). Examinando los esfuerzos evaluativos del pasado reciente, criticó duramente los conceptos en los que se habían basado las evaluaciones por su falta de relevancia y utilidad, y aconsejó a los evaluadores que abandonaran sus inclinaciones de entonces y se dedicaran a hacer evaluaciones basadas en comparaciones de los resultados obtenidos en tests tipificados, realizados por grupos experimentales y de control. Cronbach recomendó a los evaluadores que reconceptualizaran la evaluación no como si se tratara de una carrera de caballos entre programas competidores, sino como un proceso consistente en recoger y formalizar información que pueda ayudar a quienes elaboran los currículos. Cronbach argüía que el análisis y la relación de los resultados de un test podrían ser, probablemente, de más utilidad para los profesores que informales sobre el promedio del total de las calificaciones. Cuando fue publicado por primera vez, el consejo de Cronbach y sus recomendaciones



pasaron inadvertidos durante mucho tiempo, exceptuando un pequeño círculo de especialistas en evaluaciones. De cualquier manera, el artículo era básico y contenía hipótesis sobre la conceptualización y la dirección de evaluaciones que iban a considerarse totalmente válidas al cabo de muy pocos años.

En 1965, conducida por la clarividencia del senador Huben Humphrey, el carismático liderazgo del presidente John Kennedy y la gran habilidad política del presidente Lyndon Johnson, se desencadenó la Guerra contra la Pobreza. Estos programas invirtieron billones de dólares en reformas que pretendían igualar y aumentar las oportunidades para todos los ciudadanos de los Estados Unidos a través de una amplia serie de servicios sanitarios, sociales y educativos. La expansión de la economía permitió que el gobierno federal financiara estos programas, y se consiguió un gran apoyo de toda la nación para desarrollar lo que el presidente Johnson llamó la Gran Sociedad.

Junto a este esfuerzo masivo para ayudar a los necesitados, se produjo, en ciertas regiones, una cierta preocupación por el posible despilfarro de las inversiones si no se exigían algunos requisitos en la contabilidad. En respuesta a estas preocupaciones, el senador Robert Kennedy y algunos de sus colegas del Congreso promulgaron el Acta de la Educación Elemental y Secundaria, de 1965, que incluía requisitos específicos para la evaluación. Como resultado, el primer artículo de esa acta (que aspiraba a proporcionar una educación compensatoria a los niños menos favorecidos) exigía específicamente que cada distrito escolar recibiera fondos, basándose en ese artículo, para que evaluara anualmente hasta qué punto los proyectos que había realizado al amparo del artículo I habían alcanzado sus objetivos, todo ello utilizando los datos de tests debidamente estandarizados. Este requisito, con su referencia específica a las encuestas y a la valoración de la congruencia entre resultados y objetivos, refleja cómo se encontraba la evaluación educacional en aquella época. Y, lo que es más importante, el requisito obligó a los educadores a que trasladaran sus preocupaciones, en lo que se refiere a la evaluación educacional, desde el campo de la teoría y las suposiciones al de la práctica y el perfeccionamiento.

Cuando los distritos escolares empezaron a responder a los requisitos evaluativos del artículo I, comprobaron rápidamente que los instrumentos y estrategias, que empleaban entonces los evaluadores, resultaban totalmente inapropiados para aquel trabajo. Los tests estandarizados que se habían elaborado con el fin de determinar el promedio del grado de

conocimientos de los estudiantes no sirvieron para casi nada cuando hubo que diagnosticar las necesidades y valorar los avances de los niños menos favorecidos, cuyo desarrollo educativo se encontraba muy por detrás del de sus compañeros de clase media.

Además, dichos tests eran relativamente inadecuados para discernir las diferencias entre escuelas y programas, principalmente a causa de sus propiedades psicométricas y el alcance de sus contenidos. En lugar de determinar los resultados directamente relacionados con una escuela o un programa en concreto, estos tests determinaban, en el mejor de los casos, el grado de aprendizaje, centrándose casi en los mismos rasgos que los tests de aptitudes (Kella Ghan, Madaus y Airasian, 1980).

La utilización de tests estandarizados planteó otro problema: un método de evaluación como éste se oponía a los preceptos del método tyleriano. Desde el momento en que Tyler reconocía y enfatizaba las diferencias entre los objetivos de un lugar u otro, a este modelo le resultaría difícil adaptar a todo el ámbito nacional sus programas de tests estandarizados. Para ser comercialmente viables, estos programas regularizados debían olvidar ciertos objetivos bastante extendidos e importantes para ciertos lugares, en favor de otros que eran importantes para la mayoría de los distritos. Es más, había una escasez de información acerca de las necesidades y niveles de éxito de los niños menos afortunados, lo cual limitaba a los profesores a la hora de fijar objetivos de comportamiento que fueran significativos para este grupo de alumnos.

Los intentos de aislar los efectos del artículo I a través de la utilización de planes establecidos por grupos de control o experimentales fallaron, debido, principalmente, a la incapacidad de cumplir los requisitos de estos planes.

Y las inspecciones *in situ* realizadas por expertos –aunque fueron muy utilizadas por responsables gubernamentales– no resultaron aceptables como primera estrategia evaluativa porque se consideró que este método estaba falto de la “objetividad” y el “rigor” estipulados en la legislación de la ESEA. Cuando se informaba de que no se había llegado a ninguna conclusión, lo cual ocurría frecuentemente, se daba muy poca información acerca de los procedimientos empleados y, a menudo, no se proporcionaban datos sobre cómo se habían llevado a cabo. Frecuentemente, el énfasis puesto en las puntuaciones de los tests desviaba la atención del examen de los procedimientos o de su ejecución.

Como resultado de la creciente inquietud producida por los esfuerzos realizados en las evaluaciones con resultados siempre negativos, el Phi Delta Kappa creó el National Study Committee on Evaluation (Stufflebeam y otros, 1971). Tras examinar el panorama, este comité concluyó que la evaluación educacional «era víctima de una grave enfermedad» y recomendó el desarrollo de nuevas teorías y métodos de evaluación, así como nuevos programas para preparar a los propios evaluadores. Al mismo tiempo, empezaron a aparecer nuevas conceptualizaciones sobre la evaluación. Provus (1971), Hammon (1967), Eisner (1967) y Metfessel y Michael (1967) propusieron una reforma del modelo Tyler. Glaser (1963), Tyler (1967) y Popham (1971) recomendaron los tests basados en criterios como una alternativa a los tests basados en normas. Cook (1966) pidió la utilización del método de análisis sistemático para la evaluación de programas. Scriven (1967), Stufflebeam (1967, 1971) y STAKE (1967) crearon nuevos modelos de evaluación que se alejaban radicalmente de los métodos anteriores. Estas conceptualizaciones reconocían la necesidad de evaluar las metas, examinar las invasiones, y analizar el perfeccionamiento y la prestación de los servicios, así como determinar los resultados que se desean (o no se desean) obtener del programa. También insistieron en la necesidad de enjuiciar el mérito o el valor del objeto de la evaluación. Los últimos años de la década de los 60 y los primeros de la de los 70 vibraron con las descripciones, discusiones y debates acerca de cómo debía ser concebida la evaluación. Las unidades que siguen tratan en profundidad los métodos alternativos que empezaron a tomar forma en este período.

## La época del profesionalismo

Aproximadamente en 1973, el campo de la evaluación empezó a cristalizar y a emerger como una profesión diferenciada de las demás, relacionada con (pero también distinta de) sus antecedentes de investigación y control. El campo de la evaluación ha avanzado considerablemente como profesión, aunque es instructivo considerar su desarrollo en el contexto del período anterior

En aquellos momentos, como ya hemos explicado, los evaluadores se enfrentaban a una crisis de identidad. No sabían muy bien cuál era su papel, si debían ser investigadores, adiestradores de tests, profesores, organizadores o filósofos. Cualquier calificación especial que se les pudiera dar no resultaría clara. No existían organizaciones profesionales dedicadas a la evaluación entendida como especialidad, ni tampoco publicaciones

especializadas a través de las cuales los evaluadores pudieran intercambiar información acerca de su trabajo. Esencialmente, no existía literatura acerca de la evaluación educacional, excepto trabajos inéditos que circulaban únicamente entre círculos de especialistas. Había muy pocas oportunidades de prepararse para ejercer un servicio evaluativo. Las normas claras para una buena práctica se veían confinadas a las encuestas educacionales y psicológicas. El campo de la evaluación era amorfo y fragmentado. Muchas evaluaciones estaban dirigidas por personal sin preparación, otras por investigadores metodológicos que intentaban sin éxito adaptar sus métodos a las evaluaciones educativas (Guba, 1966). Los estudios evaluativos estaban cargados de confusión, ansiedad y hostilidad. La evaluación educacional como especialidad tenía poca talla y ninguna dimensión política.

Con este telón de fondo, el progreso realizado por los evaluadores educacionales para profesionalizar su especialidad durante los años 70 es totalmente destacable. Algunas publicaciones, como *Educational Evaluation and Policy Analysis*, *Studies in Evaluation*, *CEDR Quarterly*, *Evaluation Review*, *New Directions for Program Evaluation*, *Evaluation and Program Planning* y *Evaluation News*, aparecieron en aquella época, y luego han demostrado ser excelentes vehículos para recoger y difundir información acerca de las diversas facetas de la evaluación educacional. A diferencia de lo que ocurría quince años atrás, numerosos libros y monografías tratan ahora exclusivamente de la evaluación. De hecho, el problema actual no es encontrar literatura sobre la evaluación, sino seguir el ritmo de lo que se publica. El grupo May 12, la Division H de AERA, la Evaluation Network y la Evaluation Research Society han proporcionado excelentes oportunidades para que pueda establecerse un intercambio profesional entre personas relacionadas con la evaluación y otros programas.

Muchas universidades han empezado a ofrecer por lo menos un curso de metodología evaluativa (distinguiéndola de la metodología investigativa). Unas pocas—como la Universidad de Illinois, la Universidad de Stanford, el Boston College, la UCLA, la Universidad de Minnesota y la Universidad de Western Michigan— han elaborado programas para graduarse en evaluación. Durante siete años, la Oficina de Educación de los Estados Unidos patrocinó un programa nacional de preparación para el servicio evaluativo destinado a educadores especiales (Brinkerhoff y otros, 1983), y otras organizaciones profesionales han ofrecido talleres e institutos que tratan distintos temas relacionados con la evaluación. Se han fundado centros para la investigación y el desarrollo de la evaluación, entre los que se

encuentran la unidad de evaluación del Northwes; Regional Educational Laboratory, el Center for the Study of Evaluation de la UCLA, el Stanford Evaluation Consortium, el Center for Instructional Research and Curriculum Evaluation de la Universidad de Illinois, el Evaluation Center de la Universidad de Western Michigan y el Center for the Study of Testing, Evaluation, and Educational Policy del Boston College. El estado de Louisiana ha establecido un sistema y un programa para la titulación de evaluadores, Massachusetts está trabajando actualmente en un programa evaluativo similar y Dick Johnson (1980) ha elaborado un primer borrador de una guía telefónica de evaluadores y agencias de evaluación.

De un modo creciente, la especialidad está considerando la metaevaluación (Scriven 1975; Stufflebeam, 1978) como un medio para asegurar y comprobar la calidad de las evaluaciones. Como ya se ha indicado, un comité (Joint Committee, 1981), formado por doce organizaciones profesionales, ha publicado una clarificadora serie de normas para juzgar las evaluaciones de los programas educativos, proyectos y materiales, y ha establecido un mecanismo (Joint Committee, 1981) para repasar y revisar las *Normas* y facilitar su uso. Además, se han publicado muchos otros grupos de normas importantes para la evaluación educacional (*Evaluation News*, mayo de 1981). Y han aparecido muchas técnicas nuevas, tal y como se describe a lo largo de este libro.

Este considerable desarrollo profesional de la evaluación educacional ha producido diversos resultados. Primero, al producirse un indudable aumento y mejora de la comunicación en la especialidad, ha aparecido también una enorme cantidad de “cháchara” (Cronbach, 1980). Segundo, aunque la preparación y titulación de evaluadores ha mejorado con el fin de que las instituciones puedan disponer de servicios realizados por personas cualificadas, algunos lamentan que este desarrollo pueda reducirse a una especie de club cerrado y exclusivista (Stake, 1981). Tercero, la cooperación entre organizaciones profesionales relacionadas con la evaluación educacional, promovida por el Joint Committee on Standards for Educational Evaluation, es un prometedor pero frágil compromiso para fomentar la ejecución y utilización de trabajos evaluativos de alta calidad. Cuarto, mientras la creación de nuevas organizaciones profesionales ha incrementado la comunicación y reducido la fragmentación de la especialidad evaluativa, aún persiste una profunda división entre la División H de AERA, la Evaluation Network y la Evaluation Research Society. Finalmente, aunque se ha incrementado la comunicación entre los partidarios de los métodos positivistas/cuantitativos y los que proponen métodos fenomenológicos/cualitativos, existe, actualmente, el peligro de que se produzca una polarización entre estos dos campos.

A pesar de la creciente búsqueda de métodos adecuados, del incremento de la comunicación y el entendimiento entre los principales teóricos, y del desarrollo de nuevas técnicas, la práctica actual de la evaluación ha cambiado muy poco en la mayoría de los casos. Está claro que es necesario renovar los esfuerzos para educar a los evaluadores en las nuevas técnicas, para poner a prueba e informar de los resultados conseguidos mediante estas nuevas técnicas y para desarrollar técnicas adicionales. En todos estos esfuerzos, debe ponerse el énfasis en conseguir que la metodología responda a las necesidades de los profesionales y de aquellos a quienes sirven.

Hemos presentado la evaluación educacional como una profesión dinámica y aún en desarrollo. Aunque se trata de una profesión aún inmadura, no hay duda de que poco a poco se va convirtiendo en una parte identificable del amplio aparato profesional y gubernamental de educación, salud y bienestar. La predicción que se solía hacer en los años 60, en el sentido de que la evaluación de programas formalizada era sólo una moda que pronto desaparecería, ha demostrado ser falsa, y hay fuertes indicios de que esta especialidad continuará creciendo en importancia, complejidad y valor. Los logros alcanzados desde el final de la década de los sesenta son importantes, pero aún existen muchas deficiencias obvias y son necesarias una investigación y evolución constantes.

## **Reparto de papeles en el trabajo evaluativo**

Hasta aquí hemos presentado y discutido una definición básica de la evaluación, descrito algunos de los métodos relacionados con el trabajo evaluativo, señalado las normas de esta profesión y proporcionado un panorama histórico. Hecho esto, concluiremos echando un vistazo a los papeles que se desempeñan en una evaluación.

Tanto si una evaluación es dirigida por una persona como si lo es por muchas, deben desempeñarse unos cuantos papeles, y cuando se considera el trabajo relacionado con apoyar y promover la evaluación como profesión, este abanico de papeles se convierte en uno mucho más amplio. Todos los que están relacionados con la evaluación, ya sea en general o como especialistas, deben tener una amplia visión de todos estos papeles con el fin de que su participación y su colaboración se integren en la labor de grupo. Quizás nada sea tan destructivo con respecto a la potencial contribución en una evaluación como el hecho de pensar que esa evaluación es una investigación independiente realizada

por un solo investigador. Por el contrario, la evaluación alcanza su mejor momento cuando se desarrolla como un trabajo de equipo.

Con la intención de proporcionar un amplio panorama del trabajo, relacionado con las evaluaciones individuales y con la profesión evaluativa en general, hemos descrito doce papeles básicos. Aunque otros autores describan un panorama distinto y le asignen funciones diferentes, nosotros creemos que nuestra clasificación de papeles abarca muchas, sino todas, de las responsabilidades laborales que los evaluadores pueden llevar a cabo. Es más, en nuestra experiencia, esos doce papeles se han asociado con la actual distribución del trabajo. Desde luego, no siempre existe una correspondencia unívoca entre papel y trabajo, pero creemos que esos doce papeles pueden resultar útiles para el lector, no sólo para que reflexione sobre la amplia serie de actividades evaluativas existentes, sino también para que tenga en cuenta los posibles estudios que se pueden realizar en torno a la evaluación. Lo que sigue, pues, son unas breves definiciones de los doce papeles básicos que pueden desempeñarse en una evaluación.

## **El cliente de la evaluación**

Cualquier evaluación de utilidad implica la plena participación de un cliente. Esto es cierto tanto si el estudio está dirigido por una persona para servir a sus propios propósitos como si lo está por muchas. El cliente es la persona o grupo que utilizará los resultados para algún propósito, como la selección o el perfeccionamiento de un programa. El grupo de los clientes incluye tanto a quien haya encargado la evaluación, en primer lugar, como a todos aquellos que esperan los resultados para hacer uso de ellos. De una forma u otra, el tiempo y los recursos requeridos para una evaluación son proporcionados por el cliente o por una o más personas pertenecientes a un grupo de clientes. Los clientes son, también, el origen crucial de las cuestiones planteadas y de los criterios empleados para interpretar los resultados; también tienen la responsabilidad primordial de aplicar las conclusiones; y, a menudo, una evaluación no puede realizarse si el cliente no lo dispone todo con mucho cuidado y trata de asegurar que existirá una plena cooperación entre todos los que intervienen en el trabajo.

Puede encontrarse una gran variedad de clientes en las organizaciones educativas, sociales y de la salud. Esto incluye, por ejemplo, inspectores de distritos escolares, administradores hospitalarios y directores de agencias; directores de colegios, presidentes de departamento

y administradores de becas; y maestros, psicólogos y médicos. Si la evaluación consiste en contribuir al máximo a la hora de proporcionar servicios meritorios y válidos a la gente, deberán intervenir y aprovecharse toda esta serie de clientes. Además, creemos que deben estar activamente comprometidos en el encargo y la planificación de las evaluaciones, con el fin de obtener la información que necesitan para hacer bien sus trabajos. Por lo tanto, para satisfacer estas expectativas, deben estar adecuadamente preparados, tanto como todas aquellas personas que van a realizar los otros once papeles. Queremos subrayar que, quizás, la misión más importante de los programas de preparación para una evaluación –tanto para antes del servicio como para su realización– es enseñar a los clientes de la evaluación a cumplir con sus responsabilidades. Si no tienen un concepto sólido de la evaluación, probablemente no intervendrán en ella, se mostrarán pasivos, no sabrán utilizarla para conseguir sus propósitos o considerarán que los resultados tienen poco interés. En cambio, los clientes entendidos pueden ayudar a asegurar que la evaluación sea un instrumento útil y poderoso para garantizar la calidad de los servicios profesionales.

## **El diseñador de la evaluación**

Uno de los aspectos más fascinantes del trabajo evaluativo es que se trata de una empresa creativa. Diseñar una evaluación para satisfacer la demanda de información de un cliente es algo más que un proceso mecánicamente rutinario. Entre otras responsabilidades, debe conceptualizar las cuestiones que serán abordadas, examinar el panorama político, trazar una recopilación de datos y un plan de análisis, idear los informes y el proceso informativo, encargarse del aspecto financiero y del personal y determinar cómo deben ser empleadas las normas profesionales de la evaluación. Para solucionar todo este complejo entramado de decisiones de planificación, no bastan modelos de otras evaluaciones o planes ya publicados. Por el contrario, el diseño típico de una evaluación consiste en un continuo proceso conceptual.

Para diseñar una buena evaluación, se debe ser un observador sensible de los ambientes sociales, saber entrevistar y escuchar, y estudiar con voluntad y capacidad lo esencial del programa que va a ser examinado. El evaluador debe ser también experto en un amplio campo de técnicas comunicativas y de encuestas, y capaz de reunir las distintas piezas de las premisas y la información en planes de trabajo que sean prácticos, coherentes y que sirvan como guías racionales.



Debemos considerar el papel del diseñador de la evaluación como el más importante del trabajo evaluativo, junto con el del cliente. Aquellos que se convierten en expertos en diseño evaluativo, a menudo, trabajan todo el tiempo como evaluadores y están muy solicitados como asesores para ayudar a los clientes de una evaluación a planificar y encargar los estudios evaluativos.

Las conceptualizaciones generales sobre la evaluación presentadas en las unidades 3 a 10 proporcionan un importantísimo material de estudio a todos aquellos que deseen adquirir mayor facilidad para conceptualizar planes de evaluación. Además de estudiar las contribuciones creativas de otros, subrayamos que el desarrollo de la capacidad para diseñar evaluaciones requiere práctica: se necesita haber participado en la planificación de muchas evaluaciones. Creemos que esta práctica debe incluir algún trabajo individual con un cliente. Además, se puede aprender mucho colaborando con grupos encargados del diseño de evaluaciones. Igualmente, el estudio de un gran número de diseños evaluativos preparados por otros proporciona muchos buenos ejemplos de intuiciones acerca de esta compleja área de trabajo. Creemos que estos programas de preparación para evaluaciones deben incluir ejercicios y otras formas de guías prácticas para la planificación de evaluaciones.

## **El coordinador de la evaluación**

Muchas evaluaciones requieren los esfuerzos de varios participantes, y la utilización de considerables recursos y servicios para múltiples clientes. Por lo tanto, se hace necesario coordinar de un modo eficaz todo este trabajo. Esta necesidad se amplía cuando una determinada agencia tiene en preparación varios proyectos de evaluación distintos. No es sorprendente, por lo tanto, que la profesión incluya muchos trabajos evaluativos con nombres como ayudante del supervisor de la evaluación, coordinador de proyectos evaluativos, director del departamento de evaluación o director de personal del proyecto de evaluación.

Este papel implica que las personas especializadas en evaluaciones deben también prepararse para funciones administrativas, e igualmente que aquellos a quienes se les asigne un trabajo administrativo relacionado con la evaluación deben poseer o adquirir conocimientos básicos sobre la planificación de evaluaciones. Finalmente, creemos que las más importantes instituciones educativas, sanitarias y sociales deben, a menudo, dirigir sus esfuerzos hacia la creación de grupos administrativos competentes.

## **Asistentes de la evaluación**

Frecuentemente, uno de los papeles más importantes —a menudo independientes— de la evaluación es el del asistente. Este papel supone contactos periódicos con los clientes, revisiones de los documentos del programa, inspecciones de las actividades del programa, hacer entrevistas y recoger otro tipo de informaciones acerca de las partes, redactar informes y asumir la principal responsabilidad a la hora de presentar los informes. A veces, el asistente es el encargado de realizar la planificación del estudio completo. En otros casos, él o ella tiene una responsabilidad menor. Pero siempre estará encargado de colaborar con el personal del programa y de cooperar en la realización de la planificación del estudio.

Para desempeñar el papel de asistente se requieren tanto habilidad técnica como capacidad para las relaciones humanas. Además, las personas asignadas a este puesto deben poseer o ser capaces de desarrollar una gran credibilidad ante los clientes y ante todos aquellos que participan en la realización del programa.

## **Respondedores de la evaluación**

Se trata de un papel al que, generalmente no se le concede mucha importancia. Son las personas que rellenan los formularios, responden los tests, contestan las entrevistas y permiten que los observen en su trabajo. En muchos casos, la calidad de la información obtenida depende casi por completo de su buena voluntad para cooperar y dar lo mejor de sí mismos. Es improbable que realicen cualesquiera de estas tareas si no están informados acerca de su papel y convencidos de la importancia del estudio. Como consecuencia, quienes están al frente de las evaluaciones deben hacer un esfuerzo especial para asegurar que el estudio servirá de algo a estos colaboradores, tal y como ocurre con los clientes. También deben cuidar que las actividades que promuevan los motiven y preparen para responder apropiadamente.

## **Especialistas en ayuda técnica**

La mayor parte de las evaluaciones supone una cierta cantidad de trabajo técnico que requiere la ayuda de expertos especializados. Por ejemplo, puede ser necesario supervisar

y anotar los resultados de unas entrevistas; crear, proporcionar y utilizar instrumentos para una recopilación de datos estructurada; analizar tanto la información cuantitativa como la cualitativa; y realizar presentaciones audiovisuales de los resultados.

En las evaluaciones a pequeña escala, un simple evaluador puede, a menudo, hacer todas estas tareas sin asistencia especializada. El agente evaluador está obligado a coordinar y realizar todos los trabajos que sean necesarios, así como a dominar todas las técnicas especializadas que se requieran. Este tipo de agente necesita una amplia preparación que le introduzca en las más importantes de estas especialidades. A menudo, esta persona debe ser capaz de coordinar la asistencia de los especialistas cualificados.

De todas maneras, es inevitable que los estudios a gran escala contraten especialistas para llevar a cabo estas tareas con el fin de que los asistentes puedan concentrar sus esfuerzos en su propio campo de trabajo: los clientes y los respondentes. Este grupo de expertos puede incluir especialistas en la elaboración de pruebas, especialistas en muestreos, especialistas en el procesamiento de datos, estadísticos, especialistas en el estudio de casos y redactores técnicos.

## **Especialistas en información**

Un papel técnico que hemos escogido para dedicarle una atención especial es el del especialista en información. Y lo hemos hecho porque ese papel supone una continua necesidad de información relacionada con otras evaluaciones y no con problemas técnicos determinados. El especialista en información es especialmente importante en las agencias que disponen de departamentos de evaluación. Las evaluaciones dirigidas desde estos departamentos deben tomar en consideración el panorama institucional y su contexto histórico, examinar los descubrimientos del pasado y de la actualidad para determinar las tendencias y evitar una innecesaria pluralidad de esfuerzos en lo que se refiere a la recopilación de datos. Es necesario, como mínimo, un buen sistema para ordenar y archivar los informes de manera que la información que contienen pueda ser fácilmente consultada en futuras evaluaciones. Además, muchas agencias se beneficiarían de una recopilación regular de información que las ayudaría a estar totalmente al corriente de entradas de datos, procesos y producciones, así como de la actitud de sus componentes hacia los servicios de la agencia.

El desempeño del papel de especialista en información requiere que el evaluador, o uno o más miembros del equipo de evaluación, conozca bien las ciencias informativas. Eso supone conocimientos y técnicas en áreas como la comunicación, los sistemas de información, el proceso de datos, la contabilidad financiera y la planificación de gran alcance. Las personas que piensan especializarse en evaluación deben adquirir un dominio general de las especialidades de la información, mientras que un cierto tipo de trabajo correspondiente a esta área resulta importante incluso para las evaluaciones más sencillas. Además, los departamentos de evaluación deben, a menudo, asignar una o más personas para el cumplimiento de esta importante función.

## **Especialistas en comunicación**

La evaluación bien hecha es una comunicación efectiva acerca del mérito y el valor de algún objeto. Esto quiere decir que los problemas de los clientes guían el trabajo de los evaluadores, y las conclusiones de estos llegan hasta los clientes para informarlos. Ninguno de estos objetivos se hace posible sin una comunicación efectiva entre el evaluador y el cliente.

Ocurre demasiado a menudo, de todas formas, que la comunicación efectiva se pierde en el trabajo evaluativo. Los evaluadores y los clientes asumen que los primeros conocen su oficio y, de algún modo, obtendrán y facilitarán las conclusiones adecuadas; en lugar de eso, los evaluadores, regularmente, se ocupan de las cuestiones equivocadas. Por otra parte, los evaluadores deben hacer un buen trabajo consiguiendo que los clientes ayuden a centrar la evaluación y a recopilar una información que resulte sólida y potencialmente útil, para luego «pasarles la pelota» y que sean ellos los que, sin ayuda, interpreten y apliquen la información. Desearíamos dejar totalmente claro que la evaluación efectiva supone no sólo recopilar y analizar la información apropiada, sino también asegurar, a través de una efectiva tecnología comunicativa, que la información sea presentada de una forma útil y utilizada en consecuencia. Sabemos que los evaluadores no deben ser considerados responsables de todas las falsas interpretaciones y empleos erróneos que se han hecho de sus informes, pero sí, entre otras cosas, de asegurarse de que sus informes sean relevantes para los intereses de sus audiencias y de que puedan entenderlos.

Todo esto requiere que los evaluadores tengan conocimientos sobre tecnologías de la comunicación. Básicamente, deben ser capaces de preparar informes escritos con claridad y que estén dirigidos a —y resulten comprensibles para— la audiencia o audiencias que

los han solicitado. También deben saber cómo comunicar conclusiones a través de medios distintos del material impreso: por ejemplo, deben saber cómo hablar en público, utilizar técnicas audiovisuales y servirse de los medios públicos. Finalmente, queremos señalar que deben poseer conocimientos sobre el análisis político y la solución de conflictos, ya que, a menudo, las evaluaciones son objeto de controversia.

El papel del especialista en comunicación tiene considerables implicaciones para la preparación de los evaluadores, su utilización de los especialistas y la elección de personal en los departamentos de evaluación. Sin duda, los evaluadores deben adquirir preparación en una amplia serie de técnicas comunicativas. En su práctica, deben, como mínimo, someter los borradores de sus informes al juicio de los editores. Y los departamentos de evaluación, a menudo, deben emplear a especialistas en comunicación para que asesoren la planificación de las evaluaciones y la presentación de las conclusiones. Estos especialistas también desempeñan un importante papel a la hora de evaluar y ayudar a mejorar el exhaustivo trabajo de un departamento de evaluación, comunicando los resultados a sus clientes.

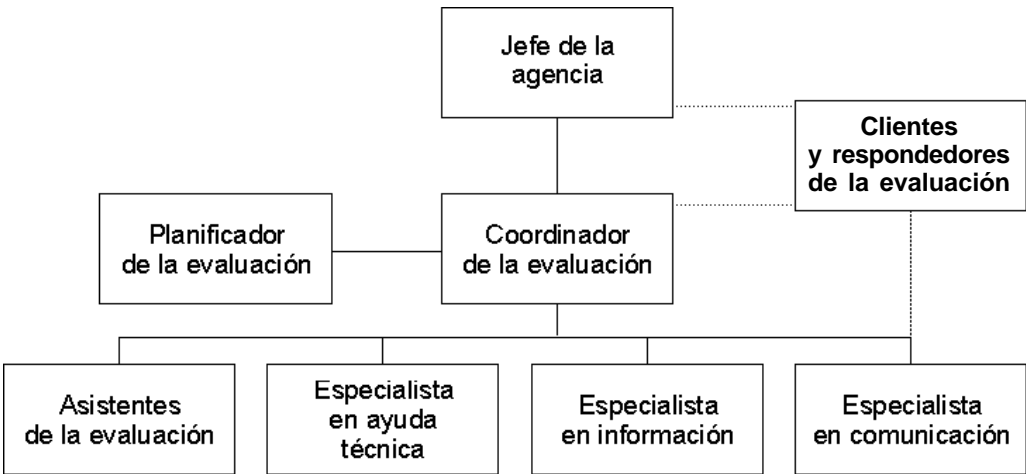
El análisis de los papeles realizados hasta este punto ayuda a clarificar la complejidad del trabajo evaluativo. Una gran agencia, que deba someter continuamente a evaluaciones a sus exhaustivos servicios, puede que necesite un personal completo de especialistas en evaluación. La figura 1-1 es un ejemplo de cómo debe estar organizado el departamento de evaluación de una agencia para que cumpla a la perfección las diversas funciones de una evaluación. Como puede verse en ese esquema organizativo, el coordinador de la evaluación tiene la responsabilidad de satisfacer las necesidades evaluativas de la agencia y de sus componentes. Para ayudarlo en el desempeño de este papel, el coordinador de la evaluación dispone de los esfuerzos de distintos especialistas. Que los servicios del departamento sean técnicamente sólidos e influyentes, depende de hasta qué punto su trabajo esté coordinado de un modo efectivo. Al mismo tiempo, queremos subrayar que la evaluación debe, a menudo, estar realizada por un solo evaluador. Para hacer bien su trabajo, debe llegar a desenvolverse bien en toda la amplia gama de papeles ya descritos. Probablemente, la mejor manera de encontrar y aprender el modo de conocer bien toda la gama de papeles de una evaluación es dirigiendo evaluaciones de una sola persona bajo la tutela de un experto en evaluaciones. Participar en equipos de evaluación también proporciona excelentes oportunidades de aprender. Los evaluadores deben ser inventivos a la hora de utilizar la ayuda de los especialistas, que pueden resultarles baratos e

incluso no costarles nada. Hasta el Llanero Solitario consideraba necesario y beneficioso colaborar con un socio leal y capaz.

Llegados a este punto en el análisis de los papeles en una evaluación, sólo hemos examinado aquellos que corresponden a una evaluación efectiva. Ahora vamos a ver otros cuatro que actúan como papeles secundarios en la profesionalización de la evaluación.

### Preparador de la evaluación

A lo largo del análisis precedente, nos hemos ido refiriendo a la necesidad de una buena preparación por parte de aquellos que participan en una evaluación. Estas necesidades incluyen tanto la orientación hacia unos planes de evaluación específicos como el desarrollo de unas técnicas y conocimientos más generalizados, y, más concretamente, los conocimientos que se adquieren examinando una práctica determinada. Además, la necesidad de una preparación corresponde tanto a los especialistas como a los clientes y a los respondedores.



**Figura 1-1 Esquema organizativo**

La evaluación, como cualquier otra profesión, debe asegurarse de que los que participan en ella disfruten de una amplia gama de sólidas oportunidades para prepararse bien. Esto incluye programas de preparación especializada que culminan en una titulación; cursos auxiliares que están concebidos para asesorar a todos aquellos que se estén preparando para una serie de funciones profesionales —como supervisores, asistentes sociales, profesores y médicos— sobre la evaluación de sus propios servicios profesionales; continuas oportunidades educativas para evaluadores y especialistas, y sesiones orientativas sobre estudios específicamente evaluativos destinados a clientes y respondedores. Este tipo de preparación debe presentar un contenido actual que resulte válido para los problemas concretos que deberán enfrentar los aprendices. Sin duda, la preparación es un papel crucial para mantener y hacer progresar la calidad de los servicios evaluativos.

### **Investigador evaluativo**

La misión de cualquier profesión es servir a sus clientes de la mejor manera posible, y la realización de este imperativo se ve siempre limitada por el estado actual de los conocimientos acerca de la especialidad. Por lo tanto, todas las profesiones deben tener un carácter investigativo. Los profesionales deben estudiar las necesidades relacionadas con la profesión y los problemas de los clientes, examinar sus prácticas a la luz de los conocimientos y principios de las disciplinas más relevantes, analizar su trabajo en su contexto histórico, teorizar sobre los servicios profesionales y hacer avanzar los conocimientos sobre su especialidad a través de un análisis estructurado de sus principales premisas e hipótesis.

Los programas de investigación más extensos se dan en las profesiones más veteranas, como la medicina y el derecho, o los negocios y la industria.

En el primer caso, la investigación, bajo la influencia de la ética profesional, se utiliza para rendir el mejor servicio posible. En el segundo, la principal motivación es la obtención de beneficios y la supervivencia en un mundo competitivo. Ambas áreas son muy conscientes de que un mejor servicio y la competitividad dependen casi totalmente de un programa de investigación dinámico y adecuado.

Aunque la evaluación es tan antigua como la civilización, sólo desde hace muy poco tiempo existe como práctica formal. Únicamente en las dos últimas décadas se han

realizado serios esfuerzos para recopilar la historia de la especialidad, teorizar sobre la práctica evaluativa y planificar estudios que resulten adecuadamente descriptivos y se basen en hipótesis apropiadas. Pero existe, y debe existir, una fuerte tendencia hacia el establecimiento y mantenimiento de una base investigativa para la especialidad. Esta tendencia es especialmente evidente en la gran cantidad de literatura aparecida sobre la especialidad evaluativa, aunque gran parte de este material no puede calificarse como investigativo. En cualquier caso, queremos subrayar que la investigación evaluativa es una función vital para asegurar que los evaluadores mejoren progresivamente sus servicios y sus resultados.

### **Promotor de la evaluación**

Se trata de otro de los papeles vitales de la evaluación. Tiene como base la función de ayudar colectivamente a los evaluadores a alcanzar y mantener el estatus de una profesión. El significado de este papel puede encontrarse en la definición de un diccionario que afirma que una profesión es “un oficio que requiere un conocimiento especializado y, a menudo, una preparación larga e intensiva, incluyendo el dominio de técnicas y métodos, así como también de las nociones científicas, históricas y eruditas en las que se basan estas técnicas y métodos, manteniendo, mediante la organización o la coordinación de conceptos, altos niveles de ejecución y éxito, y sometiendo a sus miembros a un estudio continuo y a una concepción del trabajo en la que lo que más importa es el ofrecimiento de un servicio público”.

La especialidad de la evaluación está empezando ya a presentar las características de una profesión, tal y como se promulgan en esta definición. Ya existen programas de preparación que terminan en doctorados y maestrías, así como también una extensa y boyante literatura sobre el tema. La literatura sobre evaluación recomienda contenidos y experiencias para ser incluidas en los programas de preparación, así como análisis sobre cómo deben elaborarse los programas a partir de disciplinas afines, como la filosofía, la psicología, la sociología y la economía. Muchas organizaciones profesionales y dos amplios conjuntos de normas profesionales ya están a disposición de los evaluadores. Y, desde luego, el gobierno, el público y las organizaciones de servicios continúan requiriendo evaluaciones de los servicios profesionales, así como fondos que los financien. Mantener estos avances hacia la profesionalización de la evaluación requiere iniciativas nacionales



y estatales, además del apoyo de los miembros de la profesión. Cada una de las personas que cobren por trabajar como evaluadores están obligadas a apoyar los distintos esfuerzos que se están haciendo para que la profesión siga adelante. Entre las funciones específicas que pueden cumplirse se encuentran las de editor de una publicación sobre la evaluación, elaborador de normas, educador de evaluadores y especialista y funcionario de una sociedad evaluadora. Compartiendo las responsabilidades de estas funciones, los evaluadores pueden contribuir a asegurar al público que los servicios profesionales que ellos ofrecen son modernos y de gran calidad.

## **El metaevaluador**

El último papel que queremos subrayar, el del metaevaluador, requiere mucha astucia y penetración. Está relacionado con las funciones de evaluar evaluaciones, pero también supone enjuiciar el valor y el mérito de todo lo que representa y hace la profesión: servicios de evaluación, utilización de las evaluaciones, preparación de la evaluación, investigación evaluativa y desarrollo organizativo. La metaevaluación utiliza las normas aceptadas de la profesión y trata de asegurar que son cumplidas. También se ocupa de examinar las propias normas con el fin de decidir en qué sentido deben ser revisadas.

De una manera u otra, la metaevaluación necesita incorporarse a cada una de las actividades de la profesión. A veces se trata de una actividad de autoasesoramiento, como cuando los evaluadores, los preparadores, los investigadores o los funcionarios de sociedades evaluativas examinan sus planes, la ejecución de su trabajo o los resultados, comparándolos con las normas de la profesión. En otros casos supone contratar a un agente independiente para que valore los servicios. El objetivo de la metaevaluación es asegurar la calidad de los servicios evaluativos, evitar o enfrentarse a las prácticas ilegales o a los servicios que no son de interés público, señalar el camino para el perfeccionamiento de la profesión y promover una mayor comprensión de la empresa evaluativa. Desde luego, todos los evaluadores profesionales deben mantener al día sus conocimientos de las normas de la profesión, incorporarlas a su trabajo, estar dispuestos a actuar como metaevaluadores de los trabajos de otros profesionales y actuar profesionalmente a la hora de evaluar e intentar mejorar las contribuciones que la evaluación hace a la sociedad.

Con este breve comentario sobre la metaevaluación, terminamos nuestro repaso de los papeles que se desempeñan en una evaluación. Hemos analizado minuciosamente, hasta

cierto punto, la función del evaluador profesional y hemos echado un vistazo a las divisiones que presenta. Creemos que, sin duda, todo ello ha contribuido a apoyar la idea de que la evaluación es a menudo un trabajo de equipo. Pero sacar la conclusión de que las evaluaciones necesitan de los esfuerzos de un equipo sería una falsa interpretación: por el contrario, la evaluación debe ser una parte necesariamente integrante de las funciones de todo profesional. Nuestra conclusión es que estas funciones (o papel) son tan difíciles como importantes. Y la conclusión más obvia y fundamental es que todos los profesionales deben desarrollar sus responsabilidades evaluativas, puedan o no conseguir la ayuda de un equipo y aunque tengan que desempeñar ellos solos todas las funciones de la evaluación.

Ya daríamos un gran paso en la dirección adecuada si consiguiéramos una mayor facilidad para conceptualizar los problemas evaluativos y los métodos adecuados para solucionarlos. El resto de este libro ha sido pensado para llevar a cabo este proceso. Presenta y examina en profundidad varias conceptualizaciones alternativas de la evaluación.

## **Lectura complementaria N° 2**

### **La autoevaluación como proceso previo a la acreditación de programas de pregrado de la Universidad de Antioquia**

#### **Proyecto general**

**Bernardo Restrepo Gómez**

**1998**

#### **1. Antecedentes**

La Ley 30 de 1992, Ley de la Educación Superior, creó en Colombia el Sistema de Acreditación de programas universitarios. El Comité Nacional de Acreditación, CNA, organismo adjunto al CESU e integrado al reglamentar la Ley, produjo un conjunto de normas o lineamientos para orientar la acreditación. Entre estas normas estableció procedimientos específicos en relación con las etapas que se deben seguir en el proceso. Estas etapas incluyen la revisión de condiciones iniciales del programa por acreditar, que es, en realidad, un paso previo al proceso de acreditación en sí, y que se lleva a cabo a través de una visita del CNA para verificar la existencia de tales condiciones y dar luz verde tendiente a la acreditación. La realización de la autoevaluación, cuyo informe final debe ser sometido al CNA; la visita de pares como expertos externos verificadores de la autenticidad del informe de autoevaluación; la evaluación final realizada por el CNA teniendo en cuenta la autoevaluación y la evaluación de pares y, finalmente, la decisión de acreditación por parte del Ministerio de Educación.

La Universidad de Antioquia, teniendo en cuenta este acto legal, y considerando tanto la necesidad de mejorar la calidad de los programas académicos y profesionales, como la

obligación de responder ante la comunidad por el servicio educativo que presta, comenzó a motivar las facultades, escuelas e institutos para que sopesasen la posibilidad de asumir el proceso de acreditación. El propósito fundamental, eso sí, es el mejoramiento del currículo, la enseñanza, la investigación y otros aspectos institucionales y de los programas académicos. En 1997, año señalado como el «año del cambio curricular» en la Universidad, se dio inicio a la acreditación de los primeros programas.

Dieciséis programas de pregrado respondieron positivamente a esta solicitud y otros manifestaron su interés de participar en el proceso como observadores para aprender de él y emprender luego el proceso de acreditación. El proceso de autoevaluación tomó cerca de año y medio. En septiembre de 1998 fueron enviados al CNA los informes de 14 programas.

## **2. La autoevaluación**

### **2.1 Marco referencial**

#### **2.1.1 La decisión sobre la evaluación y el objeto por evaluar**

La decisión de evaluar y acreditar es tomada por las autoridades universitarias. Los consejos de facultad deciden qué programas acreditar y autoevaluar. Los primeros programas en entrar al Sistema Nacional de Acreditación fueron los de Enfermería, Nutrición y Dietética, Educación Primaria, Ingeniería Sanitaria, Medicina Veterinaria, Trabajo Social, Diplomatura en Filosofía, Bibliotecología, Contaduría, Biología, Química Farmacéutica, Bacteriología, Educación Física, Odontología y Medicina.

Los programas académicos, no las facultades ni la Universidad, son los objetos de autoevaluación. Ahora bien, muchas cosas pueden mirarse y evaluarse. Un programa académico comprende un ambiente, una filosofía, metas y objetivos, políticas, un currículo, prácticas y actividades, recursos y actores puestos al servicio de la preparación de profesionales que sean capaces de desempeñarse apropiadamente en sus respectivas áreas en la sociedad. Volviendo a los objetivos fundamentales de la autoevaluación, calidad y responsabilidad ante la sociedad, un primer impulso es identificar resultados del programa acordes con dichos objetivos. Esto, en gran medida, ha sido hecho ya por el CNA, luego

de consulta con universidades y profesores de todos los programas en talleres que se llevaron a cabo en diferentes regiones del país. Tal precedente enmarca la búsqueda de resultados esperados de los programas.

### **¿Qué es importante para el profesorado universitario y cómo sus opiniones fueron tenidas en cuenta?**

El desarrollo de la Ley 30 de 1992 autorizó al Consejo Nacional de Acreditación, CNA, para construir unos lineamientos orientadores del proceso de autoevaluación. El consejo solicitó ideas a las universidades. Variadas fueron las reacciones de éstas, incluyendo el rechazo por parte de algunas universidades privadas renuentes a la acreditación. La Universidad de Antioquia conformó, junto con otras nueve universidades, entre ellas las más importantes del país, un grupo de estudio para establecer grandes logros e indicadores de estos y redactar con ellos una propuesta en relación con los lineamientos generales para la autoevaluación. La propuesta salió y fue entregada al CNA. Éste, a su vez, convocó, como ya se dijo, a profesores de todas las universidades del país, por programas, para debatir una propuesta más refinada de lineamientos, solicitando complementación, críticas, sugerencias sobre indicadores y criterios. Después de estos talleres, los lineamientos definitivos fueron finalmente publicados y distribuidos a las universidades para su uso. No obstante ser obligatoria su aplicación, se ha dado cierto margen para que las universidades incluyan nuevos factores de acuerdo con su filosofía y proyecto educativo.

Los lineamientos se basan en los logros esperados de la educación superior, así como en los insumos y procesos necesarios para alcanzar los primeros ¿Cuáles son unos y otros? Un modelo lógico, construido a partir de los propósitos y funciones de la educación superior deja ver la estrategia que tanto el «Grupo de las Diez» como el CNA deben haber seguido para elaborar su propuesta.

Ante todo, se pone atención a la misión, la filosofía y el proyecto educativo del programa. Estos tres componentes están enraizados en los ambientes cultural, social, económico y político, en los problemas, las necesidades y las condiciones que circundan el programa. Ello quiere decir que la evaluación tiene que considerar el **contexto** donde los profesionales trabajarán. El éxito o fracaso de un programa puede explicarse, en buena medida, debido a estas condiciones prevalecientes. Para comenzar un programa nuevo o rediseñar uno ya existente es necesario, entonces, diagnosticar estas necesidades y condiciones.

Un segundo componente en el modelo lógico proviene de los **insumos** que el programa recibe de la misión, la filosofía y la historia de la institución. A la par con estos insumos institucionales, la implementación y calidad del programa dependen también de los insumos de personal y de los insumos tecnológicos y materiales, como estudiantes, profesorado, currículo, equipo y facilidades físicas, entre otros, y de las estrategias actual y alternativas para poner en funcionamiento el programa con alta probabilidad de efectividad y eficiencia.

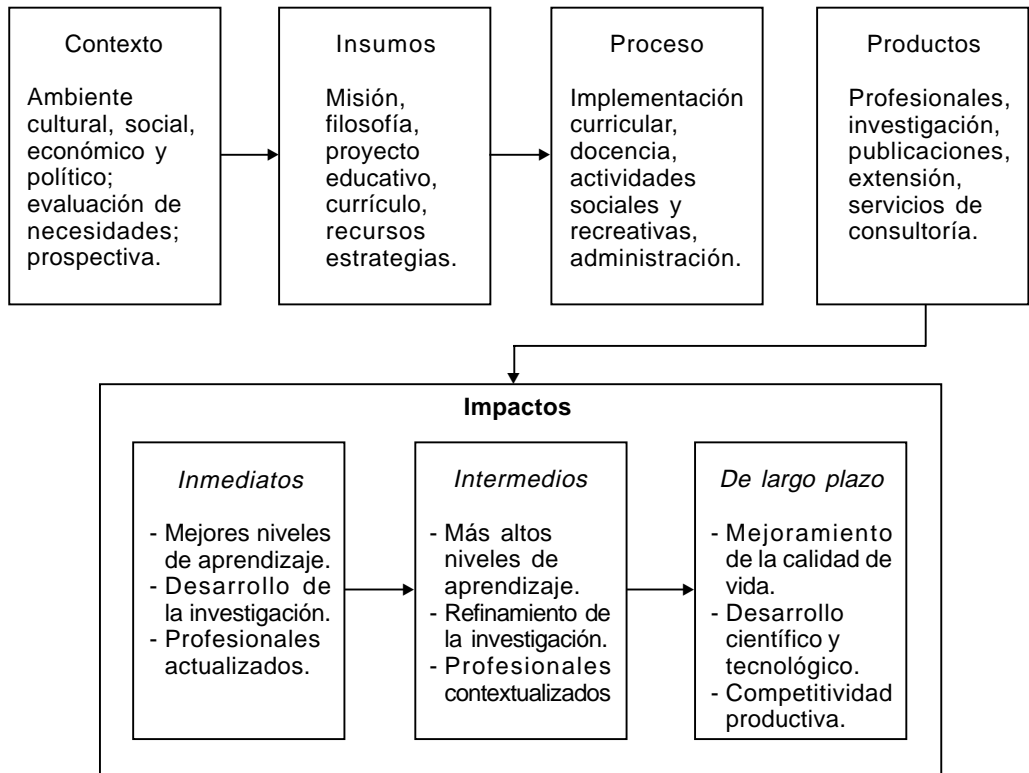
Un tercer componente, necesario para alcanzar los resultados, es el relacionado con la implementación del programa, esto es, con el **proceso o procesos**. Para poner en marcha un programa universitario es menester desarrollar un currículo; impartir docencia; hacer investigación; llevar a cabo actividades sociales, culturales y recreativas y contar con un funcionamiento adecuado y oportuno de administración y toma de decisiones.

Un cuarto componente por evaluar es el de los **resultados o productos** esperados de un programa universitario. Tales productos comprenden, en primer lugar, la entrega a la sociedad de profesionales idóneos, de resultados de investigación, innovaciones, publicaciones, servicios de consultoría y asesorías, y programas de extensión a la comunidad.

Finalmente, los lineamientos del CNA tienen en cuenta algunos impactos o repercusiones de los programas, a través de la información concerniente al desempeño de los egresados en sus puestos de trabajo y a través también de la repercusión de la investigación y las publicaciones.

El siguiente esquema trata de sintetizar el proceso lógico de derivación de componentes fundamentales en un programa y, por ende, de los componentes esenciales en la evaluación del mismo.

## Modelo lógico de los programas de pregrado



De este esquema, supuestamente, o modelo ideal de componentes de un programa, el CNA debe haber seleccionado componentes específicos para la autoevaluación, tales como la misión y el proyecto educativo de 105 programas, los procesos académicos (implementación curricular, procesos de enseñanza, investigación, publicaciones), gestión y toma de decisiones, bienestar institucional, servicios de extensión y asesoría, facilidades físicas y retroalimentación de egresados. Éstas serían las categorías obligatorias. El proceso de autoevaluación, sin embargo, puede ir más allá y, en el caso de la Universidad de Antioquia, ésta lo hace para ser coherente con su concepción de calidad y de proyección de los programas.

## **2.1.2 Preguntas relevantes, marco referencial, criterios e indicadores**

### **2.1.2.1 Preguntas relevantes y marco referencial**

Varias preguntas críticas presiden la evaluación interna, a saber:

¿Está el programa logrando sus objetivos?

¿Qué tan pertinentes son los programas en relación con la misión de la universidad, con las necesidades del país y con el progreso científico?

¿Cuál es el impacto real sobre la sociedad?

¿Qué factores pueden estar facilitando u obstaculizando la capacidad del programa para alcanzar sus metas?

¿Cómo mejorar el programa para alcanzar más satisfactoriamente sus metas e incidir más positivamente en la sociedad?

El modelo lógico debe estar apoyado en una teoría de base, en un marco referencial que explique las modificaciones, que nuevas necesidades y cambios en el contexto y la tecnología producen en las profesiones. Tal marco debe basarse en la teoría sobre educación superior, en la teoría social aplicada al análisis del contexto, en teorías particulares de los respectivos campos profesionales y en opciones teóricas hechas por los programas entre las diferentes escuelas de pensamiento con respecto a las ciencias básicas y aplicadas, que aportan su concurso a los programas. En esta forma, el marco referencial permite seguir y al mismo tiempo tomar distancia de los lineamientos del CNA, es decir, cumplir la ley al tiempo que salvaguardar la autonomía, y confiere a la evaluación un mayor sentido de utilidad, identificación de valores y viabilidad política, sobre todo ahora que los estudiantes cuestionan la acreditación relacionándola con políticas neoliberales, con la apertura económica y con la reducción del estado, particularmente con la política de privatización. Un buen marco referencial permite a profesores y estudiantes debatir la razón de la evaluación y de una evaluación en términos de la universidad, no en términos del gobierno y de influencias externas.

Algunas pautas básicas, no contenidos, fueron desarrolladas por los asesores, por solicitud expresa de los comités coordinadores, como un punto de partida para que cada programa



las enriqueciera y modificara de acuerdo con su situación específica y el énfasis dado al mismo.

Se da como un hecho que cada programa tiene su racionalidad, esto es, su razón de ser, de haber sido instaurado en un momento dado y de mantener su vigencia. En otras palabras, cada programa universitario tiene una historia y un trasfondo filosófico, social y científico.

Un resumen de estas pautas básicas para construir el marco referencial es el siguiente:

– **Antecedentes del programa**

Los antecedentes deben incluir la historia del programa y los cambios curriculares por éste experimentados, así como las explicaciones de dichos cambios, relacionadas tanto con las necesidades sociales como con el desarrollo científico y las concepciones profesionales.

– **Filosofía institucional**

Esta pauta tiene que ver con la filosofía de la universidad y con la específica del programa: su visión del mundo, sus estándares profesionales, sus valores, su compromiso con la sociedad.

– **La estructura curricular**

Es indispensable que quede claro cómo el programa establece la múltiple relación ciencia-cultura-tecnología-sociedad, básica para emprender un diseño y un rediseño curricular. En la base de esta múltiple relación debe producirse una reflexión acerca de la opción teórica con respecto a las ciencias y disciplinas básicas del programa y acerca del tejido de interdisciplinariedad entre las ciencias básicas y las aplicadas que lo conforman.

– **Pertinencia de programa**

Qué tan pertinente es o debe ser el programa, tanto desde el punto de vista social como del científico. A quién sirve y qué problemas intenta resolver y desde qué bases de

conocimiento. Es fundamental lograr un equilibrio entre las pertinencias social y científica; la primera se logra mejor si se afirma sobre la segunda.

### – Prospectiva programa

Esta pauta se relaciona con las nuevas tendencias y rumbos que las profesiones están tomando y con los factores determinantes de dichos cambios. Es cada vez más importante no perder de vista este punto debido al mundo globalizado en que vivimos y a la «aldea planetaria» en que, según muchos, se está convirtiendo el mundo.

#### 2.1.2.2 Criterios

Teniendo presentes los propósitos, categorías principales definidas por el CNA y el marco referencial en que se basa la evaluación interna, el siguiente paso consistió en establecer criterios coherentes para juzgar la situación del programa. El CNA sugirió algunos criterios, que fueron complementados así:

- |                                   |                                      |
|-----------------------------------|--------------------------------------|
| - Efectividad                     | - Colaboración internacional         |
| - Calidad                         | - Orientación hacia la investigación |
| - Pertinencia social y científica | - Prospectiva                        |
| - Eficiencia                      | - Modernización de la enseñanza      |

#### 2.1.2.3 Indicadores

El Comité Nacional de Acreditación publicó a la par con los *Lineamientos generales* un conjunto de indicadores de las categorías definidas como obligatorias y sugirió, incluso, fuentes para extraer los datos requeridos. La Universidad, sin embargo, consultó indicadores utilizados por otros sistemas de acreditación, como los de las asociaciones norteamericanas de acreditación y los sistemas de algunas universidades españolas, mexicanas y chilenas, antes de construir su propia lista de indicadores.

Para cada componente delineado en el modelo lógico se identificaron series de indicadores. Para el componente de investigación, por ejemplo, se desarrollaron tentativamente los siguientes indicadores, que cada programa entraría a revisar y redefinir según sus necesidades:

- Porcentaje de profesores que conocen las políticas de investigación.
- Porcentaje de profesores que conocen los recursos disponibles para investigación.
- Porcentaje de profesores del programa que participan en investigación.
- Premios y reconocimientos a la investigación recibidos por el profesorado en los ámbitos nacional e internacional.
- Mecanismos para promover las actividades de investigación.
- Lista de proyectos de investigación en curso.
- Lista de publicaciones en los tres últimos años.
- Criterios para juzgar la calidad de la investigación.
- Relación entre los proyectos de investigación y los objetivos del programa.
- Apoyo a la producción en las áreas de las humanidades, la estética y la plástica.
- Existencia de programas conjuntos de investigación en los ámbitos nacional e internacional.
- Porcentaje de profesores vinculados a asociaciones y redes nacionales e internacionales de investigación.
- Lista de revistas y publicaciones periódicas bajo suscripción.
- Evidencia acerca de la consulta y lectura de las revistas de investigación.

Estos son algunos de los indicadores del componente investigativo. De manera similar, todos los elementos de los componentes de contexto, insumos, procesos y productos se someten a identificación de indicadores adecuados, no solo cuantitativos, sino también cualitativos, con miras a facilitar luego la construcción de los instrumentos.

### **2.1.3 Quién evalúa el sistema de manejo de la evaluación**

Dado que la autoevaluación es una evaluación interna, los evaluadores son actores del programa evaluado. Los Consejos de Facultad, luego de ambientar la acreditación en sus respectivos espacios, toman la decisión de acreditar determinados programas. El Consejo Superior ha integrado un Comité Central Coordinador del Programa para toda la Universidad. Este comité es coordinado por uno de los asistentes del vicerrector de docencia.

Buscando ofrecer a los programas asesoría experta en evaluación, el vicerrector integró, también, desde el comienzo del proceso, un equipo de tres personas con experiencia en evaluación de programas y les solicitó acompañar el proceso, comenzando por la preparación de los evaluadores internos. Cada facultad, escuela o instituto constituye comités coordinadores (tres profesores) para conducir el proceso a nivel de programa. Estos docentes son preparados en teoría evaluativa, enfoques, modelos, criterios de evaluación y diseño metodológico antes de iniciar el proceso y durante éste, de modo que ellos puedan a su vez coordinar el trabajo de todos los actores del programa y lograr así una amplia participación en la evaluación. El propósito, dentro del enfoque de «empoderamiento», es, pues, capacitar a los comités para que estos a su vez capaciten a administradores, docentes, estudiantes y empleados, propiciando así una evaluación participativa que de a los programas el poder del aprendizaje organizacional y con él el mejoramiento continuo de la calidad y la efectividad. El proceso estuvo abierto al debate. Dada la diversidad de ideologías, puntos de vista políticos, se mantiene abierto a criterios y opiniones sobre quién y cómo debería adelantarse la evaluación. El marco referencial se constituye en valioso mecanismo para manejar la diversidad y permitir ampliar las contribuciones teóricas y metodológicas.

Un cronograma debe ser preparado, discutido y acordado por los comités de programa con miras a tener terminado el trabajo en 10 meses y máximo en un año. Los hitos principales de este cronograma son:

- Preparación del marco referencial.
- Entrega de primer borrador de instrumentos.
- Entrega de instrumentos finales, incluyendo el 20% de ítems específicos de cada programa.
- Sesión especial para discutir los problemas surgidos en la recolección de datos.
- Sesión especial para discutir las categorías formuladas para el análisis e interpretación de datos.
- Entrega de primer borrador de análisis de datos.
- Sesión especial para discutir el análisis e interpretación de datos.
- Entrega del primer borrador de informe.
- Sesión especial para discutir informes.
- Entrega de informes finales al terminar el noveno mes.

El seguimiento y monitorización de este calendario recae en manos del asistente del vicerrector de Docencia. A través de memorandos recordatorios y de llamadas telefónicas se asegura el cumplimiento de las fechas y la calidad de las tareas y sus productos.

Terminada la autoevaluación, es necesario adelantar un proceso de crítica a su implementación. El tiempo real del proceso de autoevaluación de los primeros 14 programas se dobló virtualmente. Se había calculado un periodo de 9 meses y los programas trabajaron 17 meses hasta el momento de enviar los informes al CNA. De la reflexión sobre el proceso recorrido quedan lecciones importantes para acometer autoevaluación de otros programas: los instrumentos son demasiado largos y cargados de mucha información, no siempre pertinente. El proceso de sistematización de datos debe hacerse en cada programa para evitar errores y repetición de procesos. La jerarquización de componentes y características, antes de emprender los procesos técnico de indagación, es una actividad altamente significativa y sobredeterminante de

todo el proceso, al punto de centrarlo y hacerlo más eficiente. La etapa de evaluadores internos da un fruto valioso en resultados.

### **2.1.2 El enfoque de la evaluación**

El principal objetivo de la universidad, como se dijo antes, es mejorar sus programas y rendir cuentas a la comunidad. Por lo tanto, la evaluación ha sido tomada como un proceso de aprendizaje organizacional. La evaluación sistematiza este aprendizaje mediante la vinculación de todos los interesados en los programas a los cuales se les da la oportunidad de observar cómo marchan estos, cómo se llevan a cabo y cómo responden a las necesidades de la comunidad. Muy ligado a lo anterior es el hecho de que en varios programas, a medida que se evalúa, se señalan necesidades y no solo se prospectan acciones mejoradoras, sino que se diseña una actividad de gestión de las mismas, de modo que al terminar la evaluación ya varias acciones se encuentran en curso para llenar vacíos o corregir deficiencias o modernizar procesos.

Teniendo presente lo anterior, no es apropiado hablar de un modelo o enfoque para orientar la evaluación. Primero que todo porque hay dos momentos distintos en ésta: la autoevaluación y la evaluación de pares. En efecto, cuando el informe final de la autoevaluación llegue al CNA, este consejo envía un grupo de pares a contrastar el informe con la realidad de los programas y a contribuir con algunos juicios y sugerencias para la evaluación final que hace el CNA y para la recomendación que éste hace al Ministro de Educación. Quiere ello decir, entonces, que tenemos dos tipos de evaluación, una interna y otra externa. La primera, de acuerdo con el esquema lógico planteado y los comentarios del párrafo anterior, seguirá una adaptación del modelo CIP (Stufflebeam, 1971: 1998) combinado con los recientes enfoques de evaluación-acción (Rothman, 1998) y evaluación de «empoderamiento» (Fetterman, 1996: 1998), mientras que la segunda es más bien del tipo de evaluación de expertos.

La combinación de los enfoques CIPP, evaluación-acción y «empoderamiento» proviene del hecho, ya señalado, según el cual se quiere que todas las audiencias y dentro de ellas el mayor número de personas participen en la autoevaluación para ganar el mayor aprendizaje organizacional posible, lo cual plantea acciones relacionadas con evaluación participativa y de empoderamiento, esto es, el uso de conceptos de evaluación, técnicas y hallazgos para fomentar la autodeterminación, la autorregulación y el mejoramiento de

los programas (Fetterman, 1996: 1998), y ello no sólo como resultado de todo el proceso, sino a medida que éste avance (Rothman, 1998). Estudiantes y profesores son parte especialmente interesada en el programa y sus contribuciones son muy importantes, tanto para generar nuevos factores y categorías de análisis, como para identificar problemas y necesidades en el camino y contribuir a diseñar su solución.

Podemos decir, entonces, que en lo que a enfoque se refiere, el programa de autoevaluación de los programas académicos de la Universidad de Antioquia asume una posición ecléctica, una combinación del modelo CIPP, el enfoque de expertos y la evaluación participativa, concretamente los enfoques de la evaluación-acción y la evaluación de empoderamiento.

## **2.2 Diseño de la evaluación**

### **2.1 Tipo de diseño**

El diseño escogido, dados los lineamientos del CNA y el tiempo definido para terminar la autoevaluación, es el diseño descriptivo, aunque con algunos elementos predictivos, al menos en el ámbito lógico, no en el estadístico. Es descriptivo porque su intención es describir la estructura y funcionamiento actuales de los programas, su calidad, efectividad, elementos fuertes, carencias y perspectivas. En relación con este último aspecto, el marco de referencia insiste en la prospectiva de las profesiones, lo que da a la evaluación un toque de predicción.

El diseño en un proyecto investigativo o evaluativo contempla principalmente dos tipos de planes: el plan para recoger los datos necesarios alrededor de las preguntas importantes del estudio y el plan para analizar e interpretar los datos.

#### **2.2.1.1 Recolección datos**

El primer plan incluye una revisión de las categorías o los componentes del programa evaluado, las fuentes de datos, esto es las audiencias y posibles muestras, así como las técnicas e instrumentos para obtener los datos en las fuentes identificadas.

- Categorías u objetos de evaluación.

- Misión y filosofía.
- Estructura e implementación curricular.
- Métodos y materiales de enseñanza.
- Información sobre características del profesorado.
- Información sobre características de los estudiantes.
- Desempeño y calidad de la investigación.
- Estructura administrativa y toma de decisiones.
- Bienestar institucional.
- Políticas y programas de extensión.
- Facilidades físicas y equipo.
- Desempeño de egresados.

El CNA agrupa estos objetos en siete factores.

- Audiencias, muestras y técnicas de recolección de datos.

Para obtener la información debida acerca de las categorías acordadas es necesario identificar audiencias apropiadas o poblaciones poseedoras de los datos buscados. Es conveniente, asimismo, tener en cuenta los actores afectados por los programas o interesados en él. Las audiencias determinadas en relación con los programas universitarios son:

- Los administradores de los programas, de las facultades y de la universidad. Todos son entrevistados a nivel de facultad, así como una muestra intencional en el nivel de la administración central de la universidad.



- El profesorado. En programas pequeños, todos los profesores son contactados para diligenciar un cuestionario y luego se convocan grupos focales de análisis. En programas más grandes se diseñan muestras y se llevan a cabo también grupos focales para profundizar sobre temas especiales.
- Los estudiantes. Una muestra estratificada al azar es diseñado para cada programa, teniendo en cuenta el sexo y el semestre cursado. Se usa cuestionario para la recolección de datos y se realizan también grupos focales.
- Egresados. Se aplica un cuestionario a una muestra intencional.
- Empleadores. La muestra es intencional y a ella se aplica un cuestionario.
- Empleados. Se considera realizar un censo de empleados, dado su escaso número. Se utiliza un cuestionario y se convoca grupo focal.
- Ex decanos y fundadores son entrevistados con el fin de reconstruir la historia de los programas y los cambios principales experimentados por estos. Expertos profesionales en los diversos campos son también entrevistados.
- El sistema de información de la universidad entrega oportunamente un paquete de datos atinentes a características demográficas, profesionales y laborales de los docentes, así como a características demográficas y académicas de los estudiantes.

Las decisiones sobre identificación de audiencias, diseño de muestras, envío de cuestionarios, definición de entrevistas, recolección de datos y convocatoria de grupos focales son tomadas por los comités de programas en estrecha comunicación con el profesorado y otros interesados en los programas.

#### **– Instrumentos**

- Los comités coordinadores de programa preparan cuestionarios para docentes, estudiantes, egresados, empleadores y empleados. El 80% de dichos cuestionarios deben reflejar los indicadores sugeridos por el CNA, mientras que el 20% restante debe reflejar la particularidad de los programas.

- Pautas de entrevistas han sido desarrolladas para obtener información de los administradores, ex decanos, fundadores y profesionales expertos.
- Protocolos para llevar a cabo sesiones con grupos focales.
- Protocolos para análisis de contenido son desarrollados con sus respectivas categorías para analizar documentos de apoyo.

Los comités coordinadores, conjuntamente con los asesores y con el asistente de la Vicerrectoría de docencia, se reúnen regularmente para discutir los borradores de los instrumentos y convalidar su contenido. Los asesores, luego, revisan los instrumentos, etapa por etapa e ítem por ítem, con miras a refinar su contenido, refinamiento que se considera suficiente sin tener que recurrir a pruebas piloto de los instrumentos.

La oficina de estadística de la universidad ha sido requerida para apoyar la codificación de los instrumentos y el posterior análisis estadístico de los datos.

Refinados los instrumentos, se aplican con el acompañamiento de los asesores de la autoevaluación.

### **2.1.1.2 Plan de análisis**

Dado que el diseño es descriptivo, los datos cuantitativos se reducen a porcentajes y sus resultados se presentan en barras simples y agrupadas en gráficas de línea múltiple y en gráficas de pastel, a partir de las categorías generales acordadas por los comités. Tales categorías actúan como hitos básicos del análisis cuantitativo.

Los datos cualitativos, recogidos en las entrevistas no estructuradas, en los análisis de contenido y en los grupos focales, son reducidos a categorías y analizados en relación con las preguntas relevantes y los criterios definidos al plantear la evaluación. Estos referentes, sin embargo, no son obstáculo para el desarrollo de nuevas categorías que pueden emerger en la primera lectura de los datos y en la observación realizada a lo largo del trabajo de campo. Lo que importa es obtener buenas respuestas a las preguntas centrales sobre la calidad y pertinencia social y científica de los programas, confirmar en qué grado estos están logrando sus metas y qué se ha logrado aprender durante la

evaluación para mejorar la calidad y efectividad. Entra aquí el análisis lógico a complementar el análisis descriptivo. Logrado éste, los evaluadores deben escrutar holísticamente el panorama para hacer contrastaciones con el marco de referencia, plantear discusiones, extraer inferencias y sugerir cambios.

Éste es el momento para usar los criterios en el juicio sobre la situación real de los programas y para sostener reuniones con los interesados para identificar argumentos facilitadores de la interpretación. Miembros destacados del profesorado, personas que son críticas del desarrollo del programa, son fuentes para ampliar la visión del análisis. Por supuesto, mucho trabajo de equipo deberá llevarse a cabo para interpretar lo que los datos estén diciendo y mucha imaginación y perspicacia son necesarias para sacarle lo máximo a los datos brutos, cuantitativos y cualitativos. Los comités coordinadores deben señalar las fortalezas y debilidades de los programas: aquello que es sorprendente en los hallazgos, las necesidades identificadas, preocupaciones de los estamentos, cambios indispensables y cómo encauzar dichos cambios, y quizás algunos elementos para un plan de acción.

Es importante resaltar cómo la triangulación de enfoques evaluativos, la triangulación de métodos y técnicas de recolección y análisis de datos y la triangulación de pares académicos internos y externos enriquecen el proceso de evaluación en sus distintas etapas.

### **2.3 Informes, metaevaluación y diseminación**

Terminado el análisis e interpretación de resultados, se pasa a la etapa de elaborar informes y diseminar los hallazgos a través de diferentes formas y medios.

Se ha considerado la posibilidad de invitar pares colombianos, conocedores profundos de los programas y las profesiones, a que se desempeñen como metaevaluadores y juzguen la utilidad, propiedad, factibilidad y exactitud de la evaluación. Se piensa que estos metaevaluadores podrían actuar en dos momentos: durante la recolección de los datos y en el momento de elaboración de los informes finales.

Los informes finales van al CNA para efecto de envío de pares verificadores y a los consejos académico y de facultades, escuelas e institutos para conocimiento, difusión a los

interesados y afectados y toma de decisiones. Deberán escribirse también resúmenes ejecutivos para facilitar la diseminación a ciertas audiencias. Y presentaciones orales, resaltando los hallazgos principales y los cambios necesarios, se harán ante el Consejo Superior de la universidad. El vicerrector de docencia se encarga de presentar los informes en los niveles superiores de la administración y de ver que estos tomen conciencia de los cambios que se requieran. A nivel de programa, la responsabilidad de diseminar los resultados recae en los comités coordinadores, en concertación con los respectivos consejos de facultad, escuela e instituto. El periódico de la universidad, así como su emisora cultural, será utilizado para informar a la comunidad universitaria y a la sociedad acerca de los resultados y cambios prospectados.

Los informes finales siguen la estructura de las etapas descritas en este documento: antecedentes, conceptualización de la evaluación, diseño de recolección y análisis de datos, situación del programa en el momento de emprender la acreditación y, finalmente, los resultados y su interpretación. Debe ponerse especial atención en las recomendaciones de acción, ya que el propósito primordial de la autoevaluación es el mejoramiento de los programas.

El paso siguiente consiste en la concertación de acciones mejoradoras entre el programa y la dirección central, comprometiendo tareas, tiempos y recursos adecuados a los resultados específicos previstos. Esto, sin restar fuerza a los cambios que durante el proceso mismo de la evaluación se van poniendo en marcha a la luz del enfoque de evaluación - acción ya mencionado.

## Referencias

Feiferman, D., Kaftarian, S., and Wandersman, A. *Empowerment Evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks: Sage, 1996.

Feflerman, D. *Empowerment Evaluation: Collaboration, Action Research, and a Case Example*. *Innovations in Evaluation*. Cyber Conference. Sep. 14-19, 1998.

Rothman, J. *Action-Evaluation: Integrating Evaluation into the Intervention Process*. *Innovations in Evaluation*. Cyber Conference. Sept 14-19, 1998.

Stufnebeam, D. and others. *Educational Evaluation and Decision Making*. Itasca, 111.: F. E. Peacock Publishers, inc. 1971. El modelo ha sido actualizado por el autor en 1997.

## **Lectura complementaria N° 3**

*Tomada de Forero, J. Evaluación general del programa de desarrollo rural integrado, DRI, en Colombia. Bogotá: Universidad Javeriana y Ministerio de Agricultura, 1990.*

### **Algunos elementos para una evaluación descentralizada de proyectos**

La idea central de esta propuesta es que la evaluación de un proyecto dentro de un programa que, como el Desarrollo Rural Integrado, se propone contar con la participación de los campesinos y de la municipalidad en esta tarea, rebasa la metodología de análisis costo beneficio. El análisis de la rentabilidad que afronta esta metodología es apenas una herramienta de apoyo. Lo determinante como parámetro de selección de un proyecto y de su evaluación es el proceso de concertación entre comunidad e instituciones: el grado de compromiso de quienes van a ser sus actores y las condiciones reales para sacarlo adelante. Condiciones dentro de las cuales, nuevamente, la voluntad y la comunicación humana son los elementos centrales.

Otra consideración esencial para los proyectos es su pertenencia a programas más amplios que fijen metas generales de carácter regional y nacional. Es preciso, por lo tanto, que el Desarrollo Rural Integrado recupere su capacidad de liderar la política de desarrollo rural, entre otras cosas, en cumplimiento del mandato legislativo. Dentro del espíritu de nuestra propuesta, este liderazgo implica contar con una participación orgánica de los funcionarios en todos los niveles y de las organizaciones campesinas y las comunidades rurales.

Una planeación dirigida a cumplir objetivos precisos en un plan de cofinanciación de proyectos no puede soslayar la planificación normativa a mediano y largo plazo en función de tal proyecto nacional de fortalecimiento, en nuestro caso, de la sociedad rural y la economía campesina. Estructurar los elementos institucionales y los procesos que implica el cumplimiento de esta tarea no es un objetivo de esta propuesta. Nos circunscribimos a proponer algunos elementos dirigidos a hacer posible la participación dentro de la planificación y la evaluación participativa.

## 1. La elegibilidad de proyectos

A nuestro juicio, cómo elegir un proyecto no es el problema central de la planificación del desarrollo rural; la cuestión crucial está en cómo construir un proyecto.

La sutileza y la especificidad de la organización social y económica del campesinado no encuadran dentro de los estudios de factibilidad de los proyectos ni dentro de la evaluación de los mismos.

El reto central está en cómo conseguir que las comunidades se apropien de los proyectos, lo cual remite a preguntarse si un proyecto surge realmente de la comunidad y cómo surge y cómo lo puede apoyar una entidad como el Desarrollo Rural Integrado<sup>1</sup>.

La conclusión de esta discusión no es que haya que eliminar el proceso de calificación técnica de proyectos (léase evaluación o estudios de factibilidad) para dar pautas para su elegibilidad. No, nos proponemos descartar este instrumento de planificación.

Queremos advertir, eso sí, que elegir proyectos centralmente, en Bogotá o en las capitales departamentales, genera un obstáculo técnico y político para un verdadero desarrollo rural y una franca contribución a la derrota precoz de cualquier intento de crear soluciones para los campesinos. El papel de quienes asuman la función de evaluar técnicamente los proyectos debe consistir en aportar un insumo técnico al proceso de concertación entre campesinos y entidades gubernamentales para que las decisiones se tomen en los municipios, con una representación real de las comunidades rurales que conforman el espacio municipal.

Aquí se nos plantea con toda claridad que un proceso técnico de evaluación no tiene sentido en sí mismo si se plantea exclusivamente como un sistema de flujo de información para la toma de decisiones. Se tiene que involucrar necesariamente una relación vivencial con el medio, con el medio rural en nuestro caso. Quizás las cosas sean distintas para otra clase de proyectos de naturaleza estrictamente privada, pero en materia de desarrollo social y económico hay que contar con instrumentos pragmáticos de concertación, para no repetir los modelos de imposición vertical del desarrollo rural.

---

<sup>1</sup> Conformarse con ir a preguntarle a unos cuantos líderes si les gustaría que el Estado trajese un dinero para hacer una obra significa reproducir el rito inútil de la participación formal. Inútil desde el punto de vista de la necesidad planteada atrás de construir con los campesinos los proyectos de desarrollo rural.

Decíamos que la evaluación técnica de proyectos no sólo no se descarta sino que es un instrumento útil y quizás imprescindible tanto para las entidades financiadoras (el Desarrollo Rural Integrado y otras que actúen como cofinanciadoras) como para quienes los llevan a cabo. El problema está en superar la evaluación convencional que, como dijimos, no se ajusta ni a los objetivos ni a la dinámica del desarrollo rural y en utilizar la evaluación como un elemento secundario, aunque imprescindible, para la toma de decisiones.

Algunos elementos que contribuirían a superar la evaluación convencional y manejar herramientas de evaluación que involucren asuntos (variables) centrales de los proyectos de desarrollo rural son, a nuestro juicio, los siguientes:

- a. Evaluación financiera y económica-social: no siempre aunque las metodologías lo contemplan y han desarrollado un instrumental para ello se introduce la dimensión social en los proyectos de desarrollo rural la experiencia muestra que hay aspectos de muy sencillo manejo que se pueden valorar cuantitativamente algunos ejemplos:
  - Impacto en el nivel de vida del autoconsumo o de mejoras en la calidad del agua y de la vivienda.
  - Un proyecto de producción debe considerar, entre otros aspectos, calidad de los alimentos e impacto en la dieta del productor y de la localidad.
  - Uno de mercadeo debe involucrar el impacto global sobre el mercado local o regional (y aun nacional si es el caso) en términos de los precios (y del ingreso) en todo el espacio afectado por las variaciones de estos sin restringirse al impacto en los productores directamente participantes, no bastaría tampoco con medir el efecto de un cambio de precios dentro del período en que se alteran las condiciones del mercado, sino que habría que tener en cuenta el impacto que hipotéticamente significaría su retiro en una situación de nuevos precios consolidados. Para tratar de aclarar más este punto, vale la pena mencionar que en el caso de proyectos de comercialización hay que valorar el surgimiento de liderazgos, la capacitación de cuadros directivos y socios de organizaciones de mercadeo en gestión empresarial, la consolidación de organizaciones, etc.



- b. Análisis (evaluación) de los factores cruciales hacen viable un proyecto en la realidad particular de una región:
- Participación de la comunidad en la generación del proyecto.
  - Identificación de la comunidad con el proyecto.
  - Interrelaciones comunitarias y orgánicas (de hecho o institucionalizadas) que apoyarían el proyecto.
  - Interrelaciones con las estructuras locales de poder.
- c. Concordancia de objetivos con las estrategias de desarrollo rural municipal, regional y nacional.

## 2. La relación desarrollo rural y mercado en la evaluación Desarrollo Rural Integrado

Hay que partir del hecho de que la concepción de los objetivos del DRI que enfatizan, a la vez, el abastecimiento urbano de alimentos y la elevación del nivel de vida de los campesinos, se plantea implícitamente el reto de armonizar dos lógicas económicas diferentes: la del mercado capitalista y la de la organización de la producción campesina<sup>2</sup>.

Así mismo, intervienen dos concepciones del bienestar también diferentes: la de la elevación de los ingresos monetarios frente al mejoramiento de los elementos naturales, físicos, sociales, culturales y políticos del mundo rural.

---

2. No sin razón se plantea, últimamente, que no existe una lógica económica o una racionalidad campesina. Hay multiplicidad de comportamientos racionales diferentes y aun opuestos entre distintos grupos de campesinos y entre distintos campesinos considerados individualmente "condicionados por las interrelaciones entre factores culturales y actitudes. El ser racional no tiene contenido, éste proviene de los campos de la cultura y de las actitudes..." (John ADAMS, "La vacuidad del concepto de 'racionalidad campesina' en la revista *Coyuntura Agropecuaria*, del CEGA, vol. 5 núm. 2-3, jul – oct de 1988, Bogotá.

Pero de todas formas el hecho de que en la producción campesina la familia sea la base de la organización de la empresa y la subsistencia de la familia su finalidad, le imprime una racionalidad diferente a la de otro tipo de empresas, o mejor un conjunto de comportamientos racionales que aunque diversos entre sí tienen una identidad determinada, a su vez, por la identidad entre familia, producción y reproducción.

De un lado están las presiones alimentarias y macroeconómicas de la sociedad en su conjunto y de otro está el fortalecimiento de uno de los sectores de esta sociedad: la población rural y los campesinos en particular.

Estas consideraciones que normalmente son aceptadas por todo el mundo desaparecen cuando se trata de planificar y de evaluar (o de imponer, en la práctica) el desarrollo rural. Los parámetros utilizados perfilan una concepción unidimensional que desde la lógica del mercado capitalista reproducen los criterios velados de la expropiación y la imposición<sup>3</sup>.

Ante esta situación es necesario adoptar, con toda claridad, nuevos parámetros de evaluación de los programas de desarrollo rural inscritos dentro de la bipolaridad mercado - mundo rural, de tal manera que abarquen indicadores que tengan que ver, al mismo tiempo, con las transferencias económicas con la sociedad mayor y con la elevación de la calidad de la vida de los campesinos desde el punto de vista de las necesidades reales de los campesinos mismos y desde su particular forma de entender y sentir el bienestar.

La evaluación del desarrollo rural debe preguntarse, en consecuencia, si las políticas aplicadas están fortaleciendo o no la economía y la sociedad rural. Este fortalecimiento de la producción rural implica necesariamente, por ejemplo, lograr un relativo grado de seguridad alimentaria. Esto no debe asustar a los planificadores: si los campesinos tienen una mayor disposición de alimentos (producidos por ellos mismos) van a estar en mayor capacidad de producir para el mercado. El hambre y la subalimentación<sup>4</sup> no son un buen aliciente para producir. Al contrario, llevan, en muchos casos, a un estado de postración en el cual las posibilidades productivas languidecen con las personas mismas. A estas alturas del desarrollo de la economía de mercado, el límite aquel en que el campesino al satisfacer sus necesidades básicas deja de producir para dedicarse al ocio pasa por la demanda de educación para los hijos (aun universitaria), de servicios de

---

3. Según el modelo de Bataglia, cuando las comunidades toman «decisiones ajenas» acerca de la utilización de sus propios recursos se configura una situación de expropiación. Si además los recursos son ajenos y las decisiones son tomadas también por personas ajenas a ellas, se conforma un estado de imposición. Este modelo nos ha parecido impresionantemente claro para develar en nuestro propio actuar, y en el ajeno, las pretensiones totalitaristas.

---

4. En el medio rural el 60% de la población, 6 millones y medio de personas, no consumen los alimentos mínimos necesarios. La deficiencia en ingestas calóricas abarca el 60% de las personas y la de proteínas el 53%. Ver Carlos Arturo PATIÑO, Elizabeth CAICEDO DE CARDOZO y Mercedes RANGEL BOHÓRQUEZ, *Pobreza y Desarrollo en Colombia. Su impacto sobre la infancia y la mujer*. UNICEF – DNP – ICBF, Bogotá, 1998.

salud, de vestido, de electricidad y de electrodomésticos, de transporte motorizado y de bicicletas, de nuevas máquinas y herramientas... Además, es claro que si los campesinos se alimentan bien se está contribuyendo a lograr uno de los objetivos centrales del desarrollo: erradicar la pobreza del 35% de los colombianos.

Seguridad alimentaria familiar y regional, salud, educación, conservación de las condiciones naturales de la producción (aguas, tierra y atmósfera, amenazadas por tecnologías aplicadas irracionalmente) autonomía social y política, redistribución de la tierra y otros recursos productivos, reafirmación de la identidad étnica: todos estos son objetivos del desarrollo rural, tan válidos y tan imprescindibles como la elevación de la producción mercadeada.

Visto de otra forma: si alguna conclusión clara ha sacado el análisis de los fracasos de la política de desarrollo rural, impulsada por el Banco Mundial y el BID, en el mundo, es que una condición indispensable para conjurar estos fracasos está en consultar a los campesinos sobre sus necesidades, sus objetivos y sus propuestas. Esta es una consideración que, por supuesto, tiene que ver con la planificación y con la evaluación de tales políticas.

### **3. Algunos aspectos que debe involucrar la evaluación de proyectos del Desarrollo Rural Integrado**

#### **3.1 Sistema permanente de información de precios**

El montaje de un sistema de información permanente de precios –desde el productor hasta los diversos puntos de venta detallista– es una herramienta indispensable para controlar los proyectos de mercadeo. Al mismo tiempo tal insumo sería importantísimo tanto de la gestión institucional como a los campesinos y sus organizaciones<sup>5</sup>.

5. La evaluación que hicimos de los programas de comercialización en el Ariari muestra con claridad que la falta de información de precios no alertó oportunamente a las entidades y a las cooperativas sobre el nocivo rezago de los precios del canal en donde se estaba actuando (Carulla), en relación con otras alternativas (Corabastos). Como se había iniciado el proyecto en una situación de precios altamente favorable por parte de Carulla, se asumió por mucho tiempo que la situación continuaba estable, hasta que los intermediarios reconquistaron el mercado y colocaron a la cooperativa en una situación inmanejable. Con seguridad, se hubiera podido tomar medidas eficaces oportunamente.

Un segundo ejemplo: el DRI montó en Bogotá una de las alternativas de mercadeo más interesantes, exitosas y con grandes potencialidades de impactar el mercado. Nos referimos a los “mercados móviles urbanos”. Hasta donde conocemos, el DRI no tiene ni idea de que está pasando con ellos. Si no se controlan sus precios día a día, esta alternativa pierde su capacidad de impacto en la medida en que se reduce a un negocio más de comerciantes privados que disfrutan de las rentas implícitas que significan operar a bajísimos precios en privilegiados sectores del espacio público.

Es recuperable la trayectoria del ICA, CECORA (y la de nivel piloto la U.E.R de la Universidad Javeriana). Pero, con la infraestructura institucional existente, esta tarea llevaría a generar tan inútil empapelamiento en las gerencias DRI o en las entidades regionales. Habría que crear un soporte institucional. Nuestra propuesta es que el DRI monte, en convenio con algunas entidades, un sistema de toma e información de precios permanentes que debe ser uno de los pilares básicos del sistema de información que requiere la evaluación ex-ante, sobre la marcha y ex-post del programa.

### **3.2 Análisis de proyectos de mercadeo**

- a. Aspectos económicos convencionales.
- b. Impacto en los contextos local y regional.
  - \* Precios sin proyecto frente a precios con proyecto y volumen de los productos afectados por nuevos precios.
  - \* Análisis de la distribución social de los servicios que, a la población rural, le presta el proyecto.
  - \* Grado de concentración en los usuarios de las compras, ventas, colocaciones de crédito, cursos, eventos de capacitación...
  - \* Concentración geográfica (por veredas) del cubrimiento del proyecto.
  - \* Grupos sociales involucrados o marginados del proyecto (de acuerdo con el tipo de producción, tamaño de las fincas, formas de tenencia, asalariamiento, sexo, edad).
  - \* Cantidad de personas que se capacitan en el manejo empírico del proyecto frente a grados de capacitación.
  - \* Consolidación de las organizaciones.
  - \* Socios y no socios participantes contra grado de participación (activa y pasiva en cuerpos directivos, activa y pasiva en las operaciones económicas, activa y pasiva en las asambleas, activa y pasiva en la planeación y discusión de los proyectos).

- c. Consolidación de la gerencia frente al tipo de administración gerencial, (adecuada - inapropiada, participativa - centralizada).
- d. Análisis de los factores determinantes de la viabilidad del proyecto (para la evaluación ex-ante y sobre la marcha)<sup>6</sup>.
- e. Flujo de energía humana alrededor del proyecto o de la idea del proyecto: se trata de establecer, por medio del análisis colectivo (o de entrevistas y observación), el tipo y grado de conocimiento, así como la identificación con la idea (diseño) y el compromiso real de:
  - La base social (socios y no socios de la organización).
  - Los directivos de la organización.
  - Los funcionarios (como personas) y las entidades.
- f. Interrelaciones con las estructuras locales de poder.

### 3.3 Nutrición

Ante una situación nutricional como la existente (más del 50% de la población rural está desnutrida), éste es un parámetro básico de evaluación global que se puede aplicar estratégicamente por pequeños grupos, comunidades veredales, municipios y regiones, para establecer un sistema de diseño, control y ejecución tendiente a que el programa actúe pragmáticamente y en forma integral (producción, autoconsumo, acueductos, campañas educativas, distribución y mercadeo de alimentos por medio de proyectos colectivos como los que ya están funcionando).

Se trataría de aplicar metodologías apropiadas, ya existentes, para medir la desnutrición y sus causas de manera que se cuente con un acervo de información que permita, en el proceso de planificación, ubicar los puntos críticos en que hay que actuar (insumo para

6. Es claro que la evaluación sobre la marcha ex-post y ex-ante se confunden en el caso, absolutamente generalizado en el momento actual, de procesos (o digamos, proyectos) que ya han sido iniciados con o sin apoyo institucional y sobre los cuales se diseñan proyectos para fortalecerlos. La situación inicial corresponde, en estos casos, a la situación de una etapa anterior.

la evaluación ex-ante y sobre la marcha) y valorar periódicamente el estado nutricional (insumo para la evaluación ex-ante, sobre la marcha y ex-post).

### **3.4 Seguridad alimentaria**

Éste es un parámetro asociado con el anterior y que hay que involucrar en los criterios de evaluación de los programas y proyectos de producción y comercialización. La seguridad alimentaria debe considerarse familiar y regional. El sistema de precios puede dar aportes significativos para orientar y evaluar políticas, arrojando señales sobre los puntos neurálgicos en que hay que actuar, al comparar constantemente precios locales al productor con precios locales, regionales y nacionales.

Incremento y diversificación del autoconsumo familiar y regional son parámetros que obviamente hay que tener en cuenta y cuya medición debe abordarse con grados de especificidad que vayan desde los inventarios particulares de familias hasta las condiciones regionales.

### **3.5 Proyectos de asistencia técnica**

- a. Aspectos económicos integrales.
- b. Afectación de las relaciones:
  - Insumos y producción de energía y de biomasa.
  - Costos monetarios - domésticos - precios - dependencia del mercado.
  - Autoconsumo - mercado.
  - Sostenibilidad en términos ecológicos y económicos de las tecnologías.
  - Fortalecimiento de la unidad de explotación teniendo en cuenta la poliaktividad agropecuaria y artesanal.
- c. Impacto en el contexto socioeconómico:
  - Fortalecimiento local de la economía campesina.

- Seguridad alimentaria.
  - Autoconsumo - mercadeo.
  - Generación de trabajo por grupos específicos sociales y demográficos.
  - Impacto ambiental.
- d. Factores determinantes:
- Flujo de energía humana.
  - Disponibilidad de insumos.
  - Mercado
  - Locales
- e. Concordancia con las estrategias regionales de fortalecimiento de la economía campesina y abastecimiento a los mercados.

## Lectura complementaria Nº 4

### Hacia una superación del enfrentamiento entre los métodos cualitativos y los cuantitativos\*

*Por Charles S. Reichardt, Universidad de Denver, y  
Thomas D. Cook, Northwestern University.*

*En Métodos cuantitativos y cualitativos de la investigación evaluativa.  
Madrid: Ediciones Morata, S. A., 1986.*

Es considerable el desacuerdo existente respecto a la adecuación de métodos diversos y posiciones metodológicas para realizar la investigación evaluativa. Uno de los debates actuales, de intensidad creciente, se centra en la diferencia entre métodos cuantitativos y cualitativos. Por métodos cuantitativos los investigadores se refieren a las técnicas experimentales aleatorias, cuasi-experimentales, tests "objetivos" de lápiz y papel, análisis estadísticos multivariados, estudios de muestras, etc. En contraste, y entre los métodos cualitativos, figuran la etnografía, los estudios de caso, las entrevistas en profundidad y la observación participativa. Cada uno de estos tipos metodológicos, es decir el cuantitativo y el cualitativo, tiene un grupo de partidarios quienes afirman que sus métodos preferidos son lo mejor adecuados para la evaluación. Seguidamente se incluye una muestra de las opiniones que sustentan cada una de las partes del debate.

Campbell y Stanley (1966) y Riecken y otros (1974) son citados a menudo como firmes defensores de los métodos cuantitativos. Aunque Campbell y Stanley (1966:2) no se interesaron fundamentalmente por la investigación evaluativa, describen lo experimental como "el único medio de establecer una tradición acumulativa en el que cabe introducir perfeccionamientos sin el riesgo de prescindir caprichosamente de los antiguos conocimientos en favor de novedades inferiores". Riecken y otros (1974:6,12) se muestran tan sólo ligeramente más moderados y no menos entusiastas en sus reivindicaciones acerca de los experimentos: "Los experimentos no sólo conducen a conclusiones causales más claras sino que el mismo proceso del diseño experimental contribuye a aclarar la

---

\* Notas de los autores: El trabajo del que damos cuenta en este capítulo contó en parte con la ayuda de una beca de investigación de la Universidad de Denver, una beca de investigación de la W. T. Grant Foundation y una beca (Dart-78-09368) de la National Science Grant Foundation. Los autores agradecen a Bárbara Minton y a Dafe Schellenger sus valiosos comentarios a un anterior borrador de este texto.



naturaleza del problema social que está siendo estudiado". "Cuando las condiciones no son problemáticas o cuando la creatividad y el ingenio de quien diseña la investigación pueden resolver problemas difíciles, entonces, la experimentación es el método preferible para obtener una información válida y fiable sobre la cual proyectar programas sociales"(cursiva en el original).

Weiss y Rein (1972), Parlett y Hamilton (1976) y Guba (1978) figuran entre quienes se agrupan en favor de los métodos cualitativos. Weiss Rein en particular (1972:243) indica diversas estrategias alternativas de investigación y derivadas de la tradición cualitativa a las que consideran "en general superiores al diseño experimental como metodología para la evaluación de programas de objetivos amplios". Al hablar específicamente de la evaluación educativa, Parlett y Hamilton (1976:141) añaden tajantemente:

De modo característico, los enfoques convencionales han seguido las tradiciones experimentales y sicométricas que predominan en la investigación educativa. Su propósito (irrealizado) de lograr plenamente unos " métodos objetivos" ha conducido a estudios que resultan artificiales y de alcance limitado. Afirmamos que semejantes evaluaciones son inadecuadas para ilustrar las áreas de problemas complejos con las que se enfrentan y, como resultado, suponen una escasa aportación efectiva al proceso de elaboración de decisiones...

La evaluación iluminativa se presenta como perteneciente por contraste a un "paradigma de la investigación antropológica".

De manera similar, Guba (1978: 81) señala que la investigación naturalista (a la que se compara con el trabajo etnográfico de campo y con el periodismo de investigación y se presenta como diametralmente opuesta a la investigación experimental y convencional) brinda "un modo de evaluación más apropiado y más sensible que cualquier otro practicado en la actualidad".

La corriente actual de opinión revela, en realidad, más desacuerdos que los ofrecidos por estos dos grupos de citas. Desde luego, no hay unanimidad sobre si existe o no desacuerdo. Por ejemplo, Rossi y Wright (1977:13) afirma: "Entre los investigadores evaluativos existe una coincidencia casi total en señalar que el experimento aleatorio controlado es el modelo ideal para evaluar la eficacia de la política pública ". Guba (1978) cita esta declaración con obvio desdén.

El propósito del presente capítulo consiste en señalar que parte del debate actual sobre los métodos cualitativos y cuantitativos no se centra en cuestiones productivas y, en consecuencia, no se desarrolla de manera tan lógica como sería deseable. Esto no significa afirmar que sea posible resolver por completo esta disputa metodológica. Como más tarde veremos, del debate surgen cuestiones importantes que permiten unas justas diferencias de opinión y de criterio. Convendría perfilarlas antes de que se manifestaran sentimientos como los expresados al menos en algunas de las citas anteriores. Pero el debate, tal como se está llevando a cabo, oscurece de forma progresiva las cuestiones y crea sin necesidad cismas entre los dos tipos de métodos, cuando debería tender puentes y poner en claro cuáles son los auténticos desacuerdos a los que merece la pena prestar atención.

## **El lenguaje del debate actual**

Para entender algunas de las falacias de las opiniones actuales que están cobrando popularidad, es preciso apreciar más plenamente la forma en que se desarrolla el debate entre los dos tipos de métodos. Recientes comentaristas, críticos y defensores (cf. Guaba, 1978; Parlett y Hamilton, 1976; Patton, 1976, 1978; Rist, 1977 y Wilson, 1977) consideran el debate no sólo como un desacuerdo respecto a las ventajas y desventajas relativas de los métodos cualitativos y cuantitativos, sino también como un choque básico entre paradigmas metodológicos. Según esta concepción, cada tipo de método se halla ligado a una perspectiva paradigmática distinta y única y son estas dos perspectivas las que se encuentran en conflicto. Tal como lo expone Rist (1977: 43): "En definitiva, la cuestión no estriba perse en unas estrategias de investigación: lo que sucede más bien es que la adhesión a un paradigma y su oposición a otro predispone a cada uno a concebir el mundo, y los acontecimientos que en él se desarrollan, de modos profundamente diferentes".

El concepto de paradigma procede de Kuhn (1962, 1970). Basándose en sus trabajos, Patton (1978: 203 ) define un paradigma como:

Una visión del mundo, una perspectiva general, un modo de desmenuzar la complejidad del mundo real. Como tales, los paradigmas se hallan profundamente fijados en la socialización de adictos y profesionales; los paradigmas les dicen lo que es importante, legítimo y razonable. Los paradigmas son también normativos, señalan al profesional lo que ha de hacer sin necesidad de prolongadas consideraciones existenciales o epistemológicas.

Los que ven el debate en términos de un contraste entre paradigmas proporcionan, por lo general, toda una lista de atributos de los que se afirman que permiten distinguir las concepciones globales cualitativa y cuantitativa. Por ejemplo, Rist (1977) brinda tres atributos, Patton (1978) proporciona siete y Guba (1978) aporta catorce. Sin propósito de ser exhaustivos, ofrecemos en la tabla núm. 1 muchos de los atributos más prominentes de cada paradigma. En resumen, del paradigma cuantitativo se dice que posee una concepción global positivista, hipotético-deductiva, particularista, objetiva, orientada a los resultados y propia de las ciencias naturales. En contraste, del paradigma cualitativo se afirma que postula una concepción global fenomenológica, inductiva, estructuralista, subjetiva, orientada al proceso y propia de la antropología social. El estudio que Filstead realiza en este volumen sobre los paradigmas proporciona una descripción más profunda y detallada.

Tales caracterizaciones paradigmáticas se basan en dos suposiciones que tienen una consecuencia directa en el debate acerca de los métodos. En primer lugar se supone que un tipo de método se halla irrevocablemente ligado a un paradigma de manera tal que la adhesión a un paradigma proporciona los medio apropiados y exclusivos de escoger entre los tipos de métodos. Es decir, como conciben el mundo de diferentes maneras, los investigadores han de emplear métodos distintos de investigación. Si la teoría de la evaluación de uno de estos se encuentra relacionada más estrechamente con los atributos del paradigma A que con los atributos del paradigma B, debe automáticamente inclinarse por los medios de investigación que se hallan ligados al paradigma A. En segundo lugar, se supone que los paradigmas cualitativo y cuantitativo son rígidos y fijos y que la elección entre estos es la única posible. O sea que se consideran inmutables a los paradigmas y que no existe la posibilidad de modificaciones ni de otras opciones.

Intencionadamente o no (y en algunas discusiones no resulta claramente evidente la intención), estas dos suposiciones conducen a la conclusión de que nunca cabe emplear juntos los propios sistemas cualitativos y cuantitativos. Como los métodos se hallan ligados a diferentes paradigmas y como hay que escoger entre estas concepciones excluyentes y antagónicas, uno tiene también que elegir entre los diversos tipos de métodos.

Tratar como incompatibles a los tipos de métodos estimula obviamente a los investigadores a emplear sólo uno u otro cuando la combinación de los dos sería la más adecuada para las necesidades de la investigación. Paraliza, asimismo, cualquier tentativa

de superar las diferencias entre las partes enfrentadas en el debate acerca de los tipos de métodos. Por estas razones, la conceptualización de los tipos de métodos como antagónicos puede muy bien estar llevando por mal camino tanto el debate como las prácticas metodológicas actuales. En nuestra opinión constituye un error la perspectiva paradigmática que promueve esta incompatibilidad entre los tipos de métodos. Específicamente ambas suposiciones antes citadas son falsas, así que no se sostiene la conclusión según la cual los investigadores han de elegir entre los tipos de métodos. En el análisis que sigue exponemos la falacia de ambas suposiciones (es decir, el nexo entre paradigma y método y la elección forzada entre paradigma cualitativo y cuantitativo). Tras haber considerado así al conflicto entre los puntos de vista paradigmáticos redefinimos, entonces, las cuestiones suscitadas por el debate acerca de los tipos de métodos y resaltamos algunos de los beneficios potenciales del empleo conjunto de los métodos cualitativos y cuantitativos.

**Tabla 1**  
**Atributos de los paradigmas cualitativo y cuantitativo**

<b>PARADIGMA CUALITATIVO</b>	<b>PARADIGMA CUANTITATIVO</b>
Aboga por el empleo de los métodos cualitativos.	Aboga por el empleo de los métodos cuantitativos.
Fenomenologismo y <i>verstehen</i> (comprensión) "interesado en comprender la conducta humana desde el propio marco de referencia de quien actúa"*.	Positivismo lógico; "busca los hechos o causa de los fenómenos sociales, prestando escasa atención a los estados subjetivos de los individuos."*
Observación naturalista y sin control.	Medición penetrante y controlada.
Subjetivo.	Objetivo.
Próximo a los datos; perspectiva "desde dentro".	Al margen de los datos; perspectiva "desde fuera".

\* Citas de Bogdan T. Taylor (1975: 2). No suscribimos necesariamente estas descripciones de "fenomenologismo" y "positivismo lógico" (cf. Cook y Campbell 1979) aunque semejantes caracterizaciones se hallen muy difundidas.

PARADIGMA CUALITATIVO	PARADIGMA CUANTITATIVO
Fundamentado en la realidad, orientado a los descubrimientos, exploratorio, expansionista, descriptivo e inductivo.	No fundamentado en la realidad, orientado a la comprobación, confirmatorio, reductionista, inferencial e hipotético deductivo.
Orientado al proceso.	Orientado al resultado.
Válido: datos "reales", "ricos" y "profundos".	Fiable: datos "sólidos" y repetible.
No generalizable: estudios de casos aislados.	Generalizable: estudios de casos múltiples.
Holista.	Particularista.
Asume una realidad dinámica.	Asume una realidad estable.

### ¿Determinan lógicamente los paradigmas la elección del método de investigación?

Según el uso actual, un paradigma consta no sólo de una concepción filosófica global, sino también de un nexo con un determinado tipo de método de investigación. En este sentido de la definición, el paradigma determina, entonces, el método. La cuestión que aquí se plantea es la de precisar si este nexo entre paradigma y método resulta necesario e inherente o si sencillamente procede de la definición y de la práctica. En otras palabras: ¿existe una consistencia inherente entre la admisión a la filosofía de un paradigma y el empleo de los métodos de otro? La pregunta se responde con facilidad considerando sucesivamente cada uno de los atributos paradigmáticos para ver si se halla lógicamente ligado tan sólo a uno de los métodos o si puede aplicarse igualmente bien a ambos. Cada uno de los atributos de la tabla núm. 1 es examinado separadamente a continuación.

### Reconsideración de la relación entre paradigma y método

- ¿Es necesariamente un positivista lógico el investigador que emplea procedimientos cuantitativos? Y del mismo modo ¿es necesariamente un fenomenologista el

investigador que emplea procedimientos cualitativos? Ciertamente no, porque, por un lado, muchos investigadores sociales que utilizan métodos cuantitativos suscriben una posición fenomenológica. Por ejemplo, las teorías psicológicas sociales de la atribución son fenomenológicas en cuanto que se proponen comprender conductas y creencias desde la perspectiva de los mismos actores. Sin embargo, la mayor parte, si no toda, de la investigación de la atribución se realiza en el laboratorio con métodos cuantitativos. Pero consideramos la investigación sobre la introspección, un tema que asimismo corresponde claramente al terreno del fenomenologismo. En la revisión sobre la investigación de la introspección, realizada por Niseett y Wilson (1977), la vasta mayoría de los estudios emplearon procedimientos cuantitativos como los experimentos aleatorios y las mediciones "objetivas" de la conducta.

Por otro lado sería posible, aunque quizá resultara improbable, que un etnógrafo realizara investigaciones desde una posición de positivismo lógico. Por ejemplo, imaginemos a un investigador que crea que la categoría socioeconómica se define exclusivamente en términos de bienes materiales como televisores, coches, casas y ropas. Como tales artículos pueden ser observados y contados sin referencia a los significados que tengan para sus propietarios, semejante medida de la categoría socioeconómica corresponde claramente a la tradición del positivismo lógico. Así, un investigador que emplee esta medida y que compruebe las pertenencias de un individuo mediante un trabajo etnográfico de campo estaría suscribiéndose al positivismo lógico al tiempo que empleaba métodos cualitativos.

- ¿Son necesariamente naturalistas las medidas cualitativas y necesariamente penetrantes los procedimientos cuantitativos? Los procedimientos cualitativos, como la observación participativa, pueden resultar penetrantes en algunas situaciones investigadoras. Por ejemplo, la categoría de Margaret Mead como persona ajena fue probable y fácilmente advertida por las personas que decidió estudiar y, desde luego, esta circunstancia influyó en su relación de trabajo con ellas. De la misma manera algunos de sus procedimientos cuantitativos, tales como las experiencias aleatorias, pueden, en ocasiones, ser empleados de un modo completamente discreto (cf. Lofrand y Lejeune, 1960). De hecho, la cuestión del engaño se ha suscitado en muchos experimentos de campo y de laboratorio precisamente porque se consideraba que tanto el investigador como la manipulación se hallaban perfectamente encubiertos (cf. Davis, 1961; Lofrand, 1961; Roth, 1962 y Kelman, 1972).

- ¿Son necesariamente subjetivos los procedimientos cualitativos y necesariamente objetivos los procedimientos cuantitativos? Según Scriven (1972), habría que reconocer que el término subjetivo (o alternatively el término objetivo) ha llegado a tener dos significados diferentes. Con frecuencia subjetivo da a entender "influido por el juicio humano". Conforme a este uso todos los métodos y medidas, tanto cualitativos como cuantitativos, son subjetivos. Desde luego, los modernos filósofos de la ciencia coinciden en gran parte en señalar que todos los hechos se hallan inspirados por la teoría y así, resultan, al menos parcialmente, subjetivos. Por supuesto que la asignación de números de una manera mecánica, como es común en los procedimientos cuantitativos, no garantiza la objetividad. Por ejemplo, Bogdan y Taylor (1975) describen una evaluación de un programa de adiestramiento que demuestra claramente la subjetividad de un indicador nominalmente objetivo. En esta evaluación podía señalarse una tasa de éxito del 12% o del 66%. La discrepancia procede de las diferencias en el modo en que uno (subjetivamente) decide emplear los indicadores objetivos para definir el éxito. Buena parte de este género de subjetividad aparece allí en donde se emplea un diseño y un análisis cuantitativo (cf. Boruch, 1975).

El significado alternativo de subjetivo corresponde a la medición de sentimientos y creencias. Es decir, una medida o un procedimiento son subjetivos si toman en consideración sentimientos humanos, no siendo presumiblemente estos observables de una manera directa. Una vez más hay que señalar que no hay razones para suponer que los procedimientos cualitativos tengan un monopolio de la subjetividad. Los sondeos de opinión de carácter nacional ( por ejemplo, una encuesta sobre la popularidad del actual presidente ) son excelentes ejemplos de medidas cuantitativas que resultan subjetivas. Tal es el caso de las ilustraciones en el análisis del fenomenologismo.

- ¿Aíslan necesariamente los métodos cuantitativos al investigador respecto de los datos? Fienberg (1977: 51) piensa que es "sorprendente que pueda considerarse el acercamiento a los datos como atributo exclusivo del enfoque (cualitativo) ". Como por ejemplo, Fienberg (1977) cita el hecho de enviar a sus alumnos titulados a pasar un par de noches en un coche-patrulla de la policía para ser más capaces de concebir una evaluación cuantitativa de las actividades policiales. De forma semejante, muchos investigadores cuantitativos se aventuran en el campo "para ensuciarse las manos"

y los psicólogos de laboratorio realizarán sus propias manipulaciones y acosarán a preguntas a sus sujetos para averiguar lo que significan las reacciones apreciadas en sus conductas. Tal vez un adicto defensor de lo cualitativo afirmará que todos estos son ejemplos de procedimientos cuantitativos ligados con métodos cuantitativos. De cualquier manera resulta claro que el investigador cuantitativo no debe aislarse de los datos.

- ¿Están necesariamente fundamentados en la realidad, son exploratorios o inductivos los procedimientos cualitativos mientras que los cuantitativos carecen de esa fundamentación y son necesariamente confirmatorios y deductivos? A Glaser y Strauss (1967: 17-18) se les reconoce generalmente como autores de una verdadera Biblia sobre la teoría fundamentada en la realidad. Afirman: "No existe choque fundamental entre los fines y las capacidades de los métodos o datos cualitativos y cuantitativos... Creemos que cada forma de datos resulta útil tanto para la comprobación como para la generación de la teoría, sea cual fuere la primacía del énfasis" (en cursiva en el original). Estamos de acuerdo: los métodos cualitativos no sólo pueden ser empleados para descubrir las preguntas que resulta interesante formular y los procedimientos cuantitativos no sólo pueden ser empleados para responderlas. Por el contrario, cada procedimiento puede servir a cada función. Por ejemplo, Glaser y Strauss (1965, 1967) detallan cómo cabe emplear beneficiosamente los grupos de comparación (al azar o de otro modo) en la generación de una teoría. De la misma manera, los métodos cualitativos poseen reglas de evidencia y de comprobación bien definidas para confirmar teorías (cf. Barton y Lazarsfeld, 1969; Becker, 1958 y McCall, 1969). Campbell (en este volumen) no sólo demuestra cómo el estudio etnográfico de casos puede verificar proposiciones teóricas, sino que brinda sugerencias para hacer más eficaz el procedimiento en la tarea. Lo interesante es que la lógica de la descripción y de la inferencia se extiende a través de los métodos (cf. Fienberg, 1977). El trabajo de Becker en este libro viene a señalar lo mismo cuando aplica a un tipo de datos cualitativos-fotografías criterios de validez concebidos en un principio para métodos cuantitativos.
- ¿Han de emplearse exclusivamente los procedimientos cualitativos para medir el proceso y han de emplearse exclusivamente las técnicas cuantitativas para determinar el resultado? Como en el caso de la distinción anterior entre verificación y descubrimiento, no existe tampoco aquí necesidad de que haya una división estricta



del trabajo entre los métodos cualitativos y cuantitativos. Una vez más la lógica de la tarea (descubrimiento del proceso frente a valoración del resultado) abarca los métodos. Así, por un lado, Hollister y otros ( en este volumen) describen un proceso empleando técnicas cuantitativas y Campbell (1970) llega a sugerir que el diseño experimental resultaría útil en el estudio del proceso porque podría ayudar a descartar las hipótesis alternativas. Por otro lado, el procedimiento del estudio de casos ha sido a menudo empleado con éxito para valorar el resultado. Por ejemplo, un estudio de casos es suficiente para determinar mucho de los efectos de una inundación o de un huracán y la visita al lugar es generalmente respaldada por la profesión de la ciencia social como un modo apropiado de evaluar la eficacia de los programas de adiestramiento.

- ¿Son necesariamente los métodos cualitativos válidos pero no fiables y son necesariamente los métodos cuantitativos fiables pero no válidos? Como contra - ejemplo consideremos la "observación participante " de una ilusión óptica. Por la naturaleza misma de una ilusión, la observación está llamada a originar conclusiones no válidas que sólo pueden ser corregidas mediante el empleo de procedimientos más cuantitativos (véase en este volumen el estudio de Campbell sobre la ilusión de Muller-Lyer). Ni la fiabilidad ni la validez son atributos inherentes a un instrumento de medición (tanto si éste es una regla como si se trata del ojo humano). La precisión depende más bien de la finalidad a la que se hace servir el instrumento de medición y de las circunstancias bajo las que se realiza dicha medición. A veces los ojos y los oídos de una persona constituyen el instrumento más preciso y significativo (cf. Brickell, 1976; Shapiro, 1973) mientras que en otros casos un instrumento más cuantitativo resultaría más válido así como más fiable.
- ¿Se hallan siempre limitados los métodos cualitativos al caso aislado y son por eso no generalizables? Las afirmaciones en este sentido yerran en dos puntos. En primer lugar, los estudios cualitativos no tienen por qué limitarse a casos aislados. Rist (1979), por ejemplo, da cuenta de un estudio etnográfico en el que se examinan sesenta lugares distintos para estar en mejores condiciones de generalizar. En segundo lugar, la posibilidad de generalizar depende, por lo común, de algo más que del tamaño de la muestra. Sólo en muy pocos casos, como cuando se recurre a un sondeo aleatorio, la generalización se basa en un razonamiento estadístico, y se toma de unos datos de muestreo a toda una población. Habitualmente, la

generalización es mucho menos formal y por eso resulta más inductiva y más fiable en potencia. Esto significa que los investigadores normalmente desean generalizar a poblaciones que no han sido muestreadas (por ejemplo, a niños de diferentes distritos escolares, a cabezas de familia en paro de otras ciudades y a distintas épocas y diferentes tipos de tratamientos). Semejantes generalizaciones "nunca se hallan lógicamente justificadas por completo" tanto si están basadas en datos cualitativos como si lo están en datos cuantitativos (Campbell y Stanley, 1966: 171; véase también Cronbach, 1978). Aunque en tales generalizaciones informales puede servir de ayuda una muestra amplia y diversa de casos, también puede contribuir el conocimiento profundo de un solo caso. Así, y en general, no hay razón alguna para que los resultados cuantitativos sean inherentemente más generalizables que los resultados cualitativos.

- ¿Son necesariamente holistas los procedimientos cualitativos y son necesariamente particularistas los procedimientos cuantitativos? Es claro que la respuesta tiene que ser negativa. Un investigador puede estudiar intensamente tan sólo un aspecto circunscrito de la conducta empleando, por ejemplo, el método de la observación participante simplemente porque tal conducta se observa con menos precisión o con una facilidad menor de cualquier otro modo. Por ejemplo, una conducta sutil puede ser sólo detectable para un observador agudo que posee un amplio conocimiento de la cultura local. En este caso, el investigador puede emplear la visión de un observador participante como instrumento de medición sin atender a contexto alguno más amplio que la propia y específica conducta. Del mismo modo, los métodos cuantitativos tales como sondeos o experiencias aleatorios pueden tomar en consideración "toda la imagen" aunque en manera alguna midiendo "todo".
- ¿Han de suponer los procedimientos cuantitativos que la realidad es estable y que no cambia? Reconociendo que algunos diseños son más "rígidos" que otros, los procedimientos cuantitativos en conjunto no obligan al investigador a concebir una realidad inmutable. Desde luego, una de las grandes ventajas de los cuasi experimentos temporalmente seriados es que pueden determinar el cambio temporal en el efecto de un programa contra un pasado de cambios "naturales". En un caso extremo, ninguna estrategia de valoración supone una realidad perfectamente fijada puesto que el propósito mismo de la investigación consiste en advertir el cambio. Esto no quiere decir que unos acontecimientos no planeados e incontrolables no

pueden representar un peligro para una evaluación, eficaz, sino que sólo los cambios en el entorno de la investigación pueden perturbar tanto las evaluaciones cualitativas como las cuantitativas. Por ejemplo, se señala a menudo que las experiencias aleatorias quedan invalidadas cuando se altera en el curso de la investigación el procedimiento del tratamiento. Al mismo tiempo, otros tipos de cambio imprevisto, como el incremento de las tensiones entre el observador y el personal del programa (véase Knapp, en este volumen) pueden poner también en peligro una evaluación etnográfica. En ambos casos es considerable lo que se puede realizar para hacer las evaluaciones más flexibles ante los cambios indeseados.

### **La importancia de la situación**

Partiendo de las observaciones anteriores, llegamos a la conclusión de que los atributos de un paradigma no se hallan inherentemente ligados ni a los métodos cualitativos ni a los cuantitativos. Cabe asociar los dos tipos de métodos tanto con los atributos del paradigma cualitativo como los del cuantitativo. Esto significa que, a la hora de elegir un método, carezca de importancia la posición paradigmática; ni tampoco equivale a negar que ciertos métodos se hallan por lo común unidos a paradigmas específicos. Lo principal es que los paradigmas no constituyen el determinante único de la elección de los métodos.

La elección del método de investigación debe depender también, al menos en parte, de las exigencias de la situación de investigación de que se trate. Por ejemplo, consideramos las evaluaciones del impacto de dos programas educativos. En un caso, el tratamiento consiste en un proyecto nuevo e intensivo destinado a compensar los efectos de una inferioridad económica. La importancia y la intransigencia relativa del problema justifican una inversión sustancial de tiempo y dinero, sobre todo porque un programa que se "demuestre" que es eficaz se puede adoptar como política nacional. En este caso, un experimento aleatorio puede constituir el mejor medio de informar a quienes toman las decisiones.

El otro programa educativo consiste en la concesión de unas becas para graduados que se adiestrarán en un departamento universitario. En este caso no se exige una gran inversión de tiempo y de dinero, tanto porque el programa en sí mismo resulta relativamente pequeño, como porque del resultado de la evaluación no dependerán decisiones de alcance nacional. Entonces el procedimiento habitual es una técnica de estudio de casos, como

la visita al lugar. Así defender que los experimentos aleatorios (o las visitas al lugar) han de ser siempre empleados en las valoraciones de las que se derivan repercusiones resulta tan estúpido como defender que jamás deben ser utilizados. En algunas situaciones, el procedimiento más eficaz de investigación será cuantitativo mientras que en otras se atenderá mejor al mismo fin investigador mediante un método cualitativo.

### **Conocimiento implícito en el nexo real pero imperfecto entre paradigma y método**

Por supuesto que el nexo revela algún conocimiento, en la práctica existente, entre paradigmas y métodos. Los investigadores que utilizan los métodos cualitativos se adhieren más a menudo al paradigma cualitativo que al cuantitativo. De manera similar, existe una correlación entre el uso de métodos cuantitativos y la adhesión al paradigma cuantitativo (aunque estos nexos entre paradigma y método no sean perfectos, como muchos parecen creer). Semejantes nexos pueden ser muy bien el resultado de una evaluación adaptativa que refleja el hecho de que, siendo iguales en todo lo demás, los métodos cualitativos y cuantitativos a menudo resultan los más indicados para los distintos puntos de vista paradigmáticos con los que han llegado a ser asociados.

Sospechamos que la distinción más notable y fundamental entre los paradigmas corresponde a la dimensión de verificación frente a descubrimiento. Parece que los métodos cuantitativos han sido desarrollados más directamente para la tarea de verificar o de confirmar teorías y que, en gran medida, los métodos cualitativos fueron deliberadamente desarrollados para la tarea de descubrir o de generar teorías. No es sorprendente entonces que cada tipo de método haya llegado a ser asociado con estas distintas posiciones paradigmáticas y que los métodos tengan también su mejor rendimiento cuando son empleados para esos fines específicos.

Estos nexos evolutivos podrían explicar las tendencias antiguas y recientes en el empleo de los diferentes métodos en la evaluación. En una primera época se suponía que cabía designar fácilmente unos programas para producir los resultados deseados y que el propósito de la evaluación consiste simplemente en comprobar de modo anticipado estos efectos. De forma así completamente natural la evaluación se orientaba hacia los métodos cuantitativos con su tradicional insistencia en la verificación. Más tarde se descubrió que el mejoramiento no era tan simple y que los programas podían tener una amplia variedad de efectos marginales insospechados. El énfasis de la evaluación comenzó de

ese modo a desplazarse desde la verificación de los presuntos efectos al descubrimiento de la manera en que cabría concebir un programa para que tuviese el efecto deseado y de los efectos tanto sospechados como insospechados que tales programas pudieran tener realmente. En consecuencia, algunos campos de la evaluación (y muy especialmente el de la educación) han revelado un interés creciente por los métodos cualitativos con su énfasis en el descubrimiento.

Pero aunque el nexo existente entre paradigma y método puede orientar últimamente la elección de un método de investigación, ese nexo no debe determinar en exclusiva semejante elección. Hemos señalado que la situación de la investigación constituye también un factor importante. Y esto resulta especialmente significativo porque la investigación evaluativa se realiza bajo numerosas circunstancias singulares y exigentes que pueden requerir modificaciones en las prácticas tradicionales. El hecho de que paradigma y método hayan estado ligados en el pasado no significa que en el futuro resulte necesario, o conveniente, que así sea.

### **¿Es preciso escoger entre los paradigmas?**

Como indicamos antes, el debate actual acerca de los métodos crea la impresión de que el investigador no sólo debe escoger un método en razón de su adhesión a un paradigma, sino que también debe elegir entre los paradigmas cualitativo y cuantitativo porque son las únicas opciones disponibles. Ya hemos examinado la primera cuestión; ahora abordaremos la segunda.

Los dos paradigmas a los que nos referimos proceden de dos tradiciones singulares y completamente diferentes. El conglomerado de atributos que integran el paradigma cuantitativo procede de las ciencias naturales y agronómicas, mientras que el paradigma cualitativo tuvo su origen en los trabajos de antropología social y de sociología, sobre todo de la escuela de Chicago. No resulta claro por qué se juzgó que una u otra de estas tradiciones tenían que proporcionar un paradigma adecuado a la investigación evaluativa. Por fortuna los evaluadores no se han limitado a estas dos nociones.

El hecho de que alguien realice una investigación de un modo holista y naturalista, por ejemplo, no significa que tenga que adherirse a los demás atributos del paradigma cualitativo como el de ser exploratorio y hallarse orientado hacia el proceso. Cabría

combinar más bien los atributos naturalista y holista del paradigma cualitativo con otros atributos como el de ser confirmatorio y el de hallarse orientado hacia el resultado, propios del paradigma cuantitativo; o considerar un ejemplo de la práctica general. El positivismo lógico ya no es una postura filosófica comúnmente aceptada para la investigación social. La mayoría de los investigadores han adoptado una posición fenomenológica tanto si la investigación se concentra en el proceso o en el resultado, en el naturalismo o en el control.

De hecho, todos los atributos que se asignan a los paradigmas son lógicamente independientes. Del mismo modo que los métodos no se hallan ligados lógicamente a ninguno de los atributos de los paradigmas, los propios atributos no se encuentran lógicamente ligados entre sí. Podríamos examinar uno por uno los atributos de la lista, como en la sección anterior, proporcionando ejemplos que demostrarán su independencia pero semejante tarea resultaría tediosa. Baste con decir que no existe nada, excepto quizá la tradición, que impida al investigador mezclar y acomodar los atributos de los dos paradigmas para lograr la combinación que resulte más adecuada al problema de la investigación y al medio con que se cuenta.

Suponiendo que se disponga de considerables recursos de evaluación y que se desee una evaluación global, a menudo se tratará de lograr una muestra de atributos de cada paradigma en la misma dimensión. Por ejemplo, las evaluaciones globales deben hallarse orientadas tanto al proceso como al resultado, ser exploratorias tanto como confirmativas, y válidas tanto como fiables. No existe razón para que los investigadores se limiten a uno de los paradigmas tradicionales, sí bien ampliamente arbitrarios, cuando pueden obtener lo mejor de ambos.

Por añadidura, los evaluadores han de sentirse libres de cambiar su postura paradigmática cuando sea preciso. No hay razón para ceñirse en todo momento a una mezcla de atributos. Por el contrario, al pasar de un programa al siguiente o de un estudio a otro (suponiendo que para un sólo programa se proyecte una serie de evaluaciones en lugar de un estudio de una sola vez), lo probable es que cambie la posición paradigmática que resulte más apropiada en la investigación. En consecuencia, la perspectiva paradigmática del investigador ha de ser flexible y capaz de adaptaciones.

## ¿Por qué no emplear tanto los métodos cualitativos como los cuantitativos?

### Redefinición del debate

Ya hemos visto que la elección de métodos no debe hallarse determinada por la adhesión de un paradigma arbitrario. Y ello es así tanto porque un paradigma no se halla inherentemente ligado a una serie de métodos, como porque las características del entorno específico de la investigación cobran la misma importancia que los atributos de un paradigma a la hora de escoger un método. Hemos visto también que un investigador no tiene por qué adherirse ciegamente a uno de los paradigmas polarizados que han recibido las denominaciones de "cualitativo" y "cuantitativo", sino que puede elegir libremente una mezcla de atributos de ambos paradigmas para atender mejor a las exigencias del problema de la investigación con que se enfrenta. Parece, entonces, que no existe tampoco razón para elegir entre métodos cualitativos y cuantitativos. Los evaluadores obrarán sabiamente si emplean cualesquiera métodos que resulten más adecuados a las necesidades de su investigación, sin atender a las afiliaciones tradicionales de los métodos. En el caso de que ello exija una combinación de métodos cualitativos y cuantitativos, hágase así pues.

Por supuesto que aún es posible un debate sincero respecto de cuáles son los métodos mejores, habida cuenta de una perspectiva paradigmática y de una situación de investigación determinadas. Por ejemplo, cierta parte de los textos sobre formación de mano de obra revela una discrepancia acerca de la probabilidad de que cambie la influencia de participantes en estos programas en ausencia de un efecto del tratamiento. Si no se produce semejante consecuencia (tanto de incremento como de déficit) no hay necesidad de un grupo de control para la valoración del impacto; bastaría en tales circunstancias concebir un estudio de casos. Por otra parte, si tiene lugar la consecuencia en ausencia del tratamiento resultaría deseable un grupo de control (quizá, incluso, seleccionado al azar).

Cabe también que tenga lugar un sincero debate sobre la adecuación de una determinada posición paradigmática a una evaluación específica. Como ejemplo al respecto podría citarse el de un desacuerdo acerca de la importancia relativa de la validez interna frente a la generabilidad. Indudablemente, ambos atributos son importantes pero algunos evaluadores juzgan que, a largo plazo, el medio mejor de lograr la generabilidad consiste

en asegurarse de que cada estudio sea internamente válido (cf. Campbell, 1969) mientras que otros prefieren soluciones a corto plazo en donde el objetivo inmediato de la generabilidad se antepone a la validez interna (cf. Cronbach, 1978).

El problema estriba en que en los textos se confunden ambos legítimos debates. Se formulan críticas contra los métodos cuando en realidad lo que pretende atacarse es una concepción global paradigmática y viceversa. Como métodos y paradigmas son lógicamente separables, ésta es una forma equívoca y turbia de debate. El hecho de que haya sido deficientemente empleado un método o utilizado con una finalidad inapropiada no significa que sea, en sí mismo, defectuoso o inadecuado. Los argumentos confusos sobre paradigmas con argumentos sobre métodos sólo han servido para llegar hasta la situación presente en la que los investigadores se alinean tras los tipos de métodos. Si se redefine este debate bajo la forma de dos discusiones diferentes y legítimas resulta patente la falacia del proceder actual: en vez de ser rivales incompatibles, los métodos pueden emplearse conjuntamente según lo exija la investigación.

Cierto número de investigadores habían definido anteriormente esta posición según la cual cabe emplear conjunta y eficazmente los métodos cualitativos y cuantitativos tanto en el contexto de la investigación de evaluación (cf. Britan, 1978; Campbell, 1974; Cook y Cook, 1977 y Stake, 1978 ) como en disciplinas de naturaleza más tradicional (cf. Denzin, 1970; Eisner, 1977; Rist, 1977 y Sieber, 1973 ). Tal vez fue Trow (1957:338 ) quien lo expresó mejor en un debate sobre las ventajas relativas de la observación participativa en comparación con la entrevista.

Cada zapatero piensa que el cuero es lo único que importa. La mayoría de los científicos sociales, incluyendo quien esto escribe, tienen sus métodos favoritos de investigación con los que están familiarizados y en cuya utilización poseen una cierta destreza. Y sospecho que fundamentalmente decidimos investigar aquellos problemas que parecen vulnerables a través de tales métodos. Pero deberíamos, por lo menos, tratar de ser menos localistas que los zapateros. Prescindamos ya de las argumentaciones de la "observación participante " frente a la entrevista –como ya hemos renunciado en buena medida a las discusiones de la psicología frente a la sociología– y prosigamos con la tarea de abordar nuestros problemas con el más amplio despliegue de instrumentos conceptuales y metodológicos que poseamos y que tales problemas exigen. Esto no



excluye la discusión y el debate respecto a la utilidad relativa de los diferentes métodos para el estudio de problemas o tipos específicos de problemas. Pero resulta algo muy distinto de la afirmación de una superioridad general e inherente de un método sobre otro, basándose en algunas cualidades intrínsecas que supuestamente posee (la cursiva es del original).

### **Ventajas potenciales del empleo conjunto de los métodos cualitativos y cuantitativos**

Existen al menos tres razones que respaldan la idea según la cual, cuando se abordan los problemas de evaluación con los instrumentos más apropiados que resulten accesibles, se empleará una combinación de los métodos cualitativos y cuantitativos. En primer lugar, la investigación evaluativa tiene por lo común propósitos múltiples que han de ser atendidos bajo las condiciones más exigentes. Tal variedad de condiciones a menudo exige una variedad de métodos. En segundo lugar, empleados en conjunto y con el mismo propósito, los dos tipos de métodos pueden vigorizarse mutuamente para brindarnos percepciones que ninguno de los dos podría conseguir por separado. Y en tercer lugar, como ningún método está libre de prejuicios, sólo cabe llegar a la verdad subyacente mediante el empleo de múltiples técnicas con las que el investigador efectuará las correspondientes triangulaciones. Ya que los métodos cuantitativos y cualitativos tienen con frecuencia sesgos diferentes, será posible emplear a cada uno para someter al otro a comprobación y aprender de él. Aunque no afirmamos que estos tres puntos se hallen por completo desligados, cada uno será considerado separadamente a continuación.

- 1. Objetivos múltiples.** Como ya se ha advertido antes, una evaluación global ha de interesarse tanto por el proceso como por el resultado. Mediante el análisis del proceso, los investigadores han llegado a distinguir dos objetivos distintos a los que aplicaremos denominaciones diferentes. La finalidad primera del proceso estriba en comprobar: describir el contexto y la población del estudio, descubrir el grado en el que se ha llevado a cabo el tratamiento o programa, proporcionar una retroinformación de carácter inmediato y de tipo informativo, etc. La segunda finalidad del proceso estriba en la explicación causal: descubrir o confirmar el proceso mediante el cual el tratamiento alcanzó el efecto logrado. Naturalmente, la medición del efecto del programa es el resultado total o valoración del impacto.

Para una comprensión completa del programa, una evaluación tendría que realizar al menos tres tareas: comprobación, valoración del impacto y explicación causal. Se trata de una gama muy amplia de tareas que, para ser eficazmente atendidas, requerirán quizá el empleo de métodos tanto cualitativos como cuantitativos. Aunque de un modo inevitable, puede suceder a menudo que la comprobación sea realizada con mayor eficacia conforme a un modo cualitativo, que la valoración del impacto se realice con mayor precisión mediante métodos cuantitativos y que la explicación causal se obtenga de mejor manera a través del empleo conjunto de métodos cualitativos y cuantitativos. Aunque desde luego pretendemos evitar la impresión de que se requiere una división rígida o inherente del trabajo, creemos que un número mayoritario de casos y para conseguir todos los propósitos de una evaluación hará falta una combinación flexible y adaptativa de los tipos y métodos.

En el empleo conjunto de los métodos para atender a las múltiples necesidades de la investigación, la evaluación debe mostrarse receptiva a las formas nuevas y singulares y a la concatenación de los métodos. A menudo se dice que las ciencias sociales cuantitativas han adoptado totalmente sus métodos de las ciencias naturales y agronómicas. Resultaría lastimoso de ser cierto, puesto que difícilmente cabría esperar que los instrumentos de una disciplina fuesen los más indicados para los fines y circunstancias de otra. En realidad, el proceso de adopción de tales métodos ha ido acompañado de diversos perfeccionamientos (cf. Boring, 1954, 1969) y debería estimularse la realización de nuevas adaptaciones.

Algunas adaptaciones útiles pueden suponer un maridaje creativo de los métodos cualitativos y de los cuantitativos, como, por ejemplo, el empleo de experimentos aleatorios con observadores participantes como instrumentos de la medición. Esto podría contribuir a evitar las críticas a menudo suscitadas contra la evaluación cuantitativa y según las cuales las medidas sicométricas más corrientemente empleadas carecen de sensibilidad respecto de la dimensión del interés. También podría acabar con las críticas formuladas contra muchas evaluaciones cualitativas y según las cuales carecen de control los peligros que amenazan su validez interna. Examinando otro ejemplo, consideremos ahora el empleo de los etnógrafos en combinación con el diseño de regresión-discontinuidad. La esencia del diseño de regresión discontinuidad estriba en que el tratamiento es asignado de un modo estricto conforme a una dimensión cuantitativa (cf. Cook y Campbell, 1979). Ésta

podría ser desde una medida de distancia geográfica a un límite como el de una calle o cualquier otra frontera natural que sirva como línea divisoria entre quienes no lo reciben. Para medir en semejante caso el efecto del tratamiento, los etnógrafos pueden ir y cruzar libre y repetidamente la frontera con objeto de determinar si existe una discontinuidad en ese punto en las conductas o actitudes de interés. También son posibles otras combinaciones infrecuentes de métodos cualitativos y cuantitativos.

2. **Vigorización mutua de los tipos de métodos.** En un sentido fundamental, los métodos cualitativos pueden ser definidos como técnicas de comprensión personal, de sentido común y de introspección mientras que los métodos cuantitativos podrían ser definidos como técnicas de contar, de medir, y de razonamiento abstracto. Evidentemente, esto supone un cambio de los significados de los métodos cualitativos y cuantitativos respecto de los empleados en el resto del capítulo, aunque se halla justificado porque algunos autores suscriben estas definiciones. Semejante alteración en el significado resulta útil porque la nueva perspectiva revela claramente cómo se complementa con el otro cada uno de los tipos de métodos. De modo específico, el conocimiento cuantitativo debe basarse en el conocimiento cualitativo pero al proceder así puede superarlo.

Lógicamente, los métodos cuantitativos no pueden sustituir a los cualitativos porque la comprensión cuantitativa presupone un conocimiento cualitativo. Por ejemplo, Campbell (en este volumen) demuestra cómo se emplea la medición cuantitativa de una ilusión óptica para corregir la observación cualitativa, pero procede así apoyándose constantemente en éste y en muchos otros aspectos. La base del argumento es que toda medición se halla fundada en innumerables suposiciones cualitativas acerca de la naturaleza del instrumento de medida y de la realidad evaluada. Por ejemplo, al registrar el movimiento de los frentes de perturbaciones atmosféricas, se da por supuesto que la fuerza gravitatoria de la tierra permanece constante dado que unos cambios en ésta determinarían alteraciones artificiales en la presión barométrica. Al tomar nota del desarrollo de un niño, se supone que quien cambia a lo largo del tiempo es el niño y no el instrumento que se utiliza para medir su crecimiento. La elección de un modelo estadístico que encaje con los datos, la interpretación de los resultados a que dé lugar y la generalización de los descubrimientos a otros entornos se hallan todas basadas en un conocimiento cualitativo. Muy simplemente, los

investigadores no pueden beneficiarse del empleo de los números si no conocen, en términos de sentido común, lo que estos significan.

Del mismo modo, el conocimiento cualitativo puede beneficiarse del conocimiento cuantitativo. Incluso los investigadores más introspectivos y de orientación más subjetiva no pueden prescindir del hecho de contar elementos o de emplear conceptos cuantitativos como " más grande que " y " menos que ". La medición cuantitativa de las ilusiones ópticas puede corregir la observación cualitativa al tiempo que se basa en ésta. Y un hallazgo cuantitativo puede estimular una ulterior indagación cualitativa (Light, 1979; Sieber, 1973 ), como cuando un sorprendente resultado experimental impulsa al investigador a interrogar a los sujetos en busca de indicios introspectivos.

La ciencia habitual emplea conjuntamente el conocimiento cualitativo y el cuantitativo para alcanzar una profundidad de percepción, o visión binocular, que ninguno de los dos podría proporcionar por sí sólo (Eisner, 1977). Lejos de ser antagónicos, los dos tipos de conocimientos resultan complementarios. Eso no significa decir que siempre sea fácil combinarlos. Surgirán a menudo problemas de difícil solución ( cf. Trend, en este volumen ) pero por lo común siempre habrá discrepancias, y de ahí las dificultades cuando se empleen conjuntamente dos métodos cualesquiera. Resolver las diferencias entre el conocimiento cualitativo y cuantitativo no debe ser en principio más difícil que la resolución de otros enigmas de la investigación, aunque sospechamos que quizá pueda resultar a menudo más ilustrativo.

3. **Triangulación a través de operaciones convergentes.** El empleo complementario de métodos cualitativos y cuantitativos, o el uso conjunto de cualesquiera métodos, contribuye a corregir los inevitables sesgos presentes en cualquier método. Con sólo un método es imposible aislar el sesgo del método de la cantidad o de la cualidad subyacentes que se intenta medir. Pero cabe emplear conjuntamente varios métodos para triangular la " verdad " subyacente, separando, por así decirlo, el grano de la paja (cf. Denzin, 1970; Garner y otros, 1956 y Webb y otros, 1966). Aunque cabe utilizar para este fin dos o más, los diferentes métodos que converjan en las mismas operaciones resultan mejores que los semejantes porque probablemente compartirán menos sesgos que estos últimos. Con frecuencia, los métodos cualitativos y cuantitativos operan bien juntos porque son relativamente distintos.

Por añadidura, cada tipo de método puede, en potencia, enseñar al otro nuevos modos de detectar y de disminuir el sesgo. Como estos dos tipos de métodos han existido en tradiciones distintas y en buena medida aisladas, gran parte de su técnica propia ha permanecido también aislada. Al poner juntos los métodos pueden fortalecerse estos dos diferentes depósitos de conocimientos y de experiencia. Es posible, incluso, que a través de su empleo conjunto se descubran nuevas fuentes de sesgos y nuevos medios para disminuirlos, que habían permanecido ignorados de cada una de las dos tradiciones aisladas.

Tradicionalmente, la investigación evaluativa se ha concentrado en los métodos cuantitativos, destacando tanto su empleo como sus sesgos potenciales. Desde luego, el hecho de que muchos de los sesgos probables en los métodos cuantitativos hayan sido tan abiertamente reconocidos ha sido en parte responsable de la creciente insatisfacción respecto de estos métodos y de la progresiva defensa de los métodos cualitativos en algunos sectores. Es indudable que la tradición cuantitativa en evaluación podría aprender mucho de la experiencia acumulada en la eliminación de sesgos que se ha desarrollado dentro de la tradición cualitativa. Por ejemplo, el interés cualitativo por validez descriptiva y los sesgos de muestreos no aleatorios (por ejemplo el sesgo elitista; Vidich y Shapiro, 1955) podrían ilustrar muy bien los procedimientos cuantitativos de muestreo.

Del mismo modo, a pesar de la larga tradición de prevención de los sesgos en la literatura sociológica y antropológica, empiezan tan sólo a aflorar muchas de las dificultades del empleo de métodos cualitativos en el contexto de la investigación evaluativa (cf. Knapp, en este volumen; Ianni y Orr, en este volumen) y este proceso de aprendizaje se podría acelerar gracias a los logros de la tradición cuantitativa. Por ejemplo, la investigación cuantitativa sobre los procesos psicológicos de introspección y juicio se podría muy bien emplear para informar al observador etnográfico. Nisbett y Wilson (1977) han pasado revista a los datos que sugieren que a veces observadores y participantes no pueden describir con precisión su propia conducta a través de la introspección. La investigación de Chapman y Chapman (1967) y Chapman (1967) acerca de las correlaciones ilusorias y el trabajo de Tversky y Kahneman (1974) sobre juicios bajo incertidumbre pueden ayudarnos a apreciar sesgos en la forma en que observadores participantes detectan y registran covariaciones en la conducta que es objeto de estudio. De modo similar Scheirer (1978) se basa en investigaciones

sicológicas de laboratorio para indicar que a menudo tanto participantes como administradores y observadores se exceden al dar cuenta de los aspectos positivos de un programa. Además, muchas de las ideas clásicas de la validez asociadas especialmente a la concepción cuantitativa pueden servir también para el empleo de métodos cualitativos. Ejemplo al respecto son tanto Becker (en este volumen) con su interés por eliminar amenazas a la validez en la interpretación del método cuantitativo de fotografía, como Campbell (en este volumen) con sus sugerencias relativas a grupos adicionales de comparación y observadores múltiples en los estudios de casos.

### **Obstáculos en el empleo conjunto de métodos cualitativos y cuantitativos**

Aunque lógicamente deseable, es posible que surjan cierto número de obstáculos prácticos a la hora de combinar en un estudio de evaluación métodos cualitativos y cuantitativos. El siguiente análisis toma en consideración cuatro de tales obstáculos.

En primer lugar, la combinación de métodos cualitativos y cuantitativos puede resultar prohibitivamente cara. Consideremos que en comparación con un estudio de casos, un experimento aleatorio ofrece el gasto adicional de tener que recoger datos de un grupo de control que no ha sido objeto de tratamiento, mientras que los costos de recogida de datos por cada uno de los que responden es analizado por el agente etnográfico de campo son usualmente mucho más elevados que el del cuestionario tradicional de los procedimientos cuantitativos. La combinación de los métodos puede significar en definitiva la concatenación de los elementos relativamente más costosos con cada uno.

En segundo lugar, el segundo conjunto de métodos cualitativos y cuantitativos pueden suponer demasiado tiempo. A los partidarios de uno y otro grupo en el debate sobre los tipos de métodos les preocupa que sus sistemas preferidos puedan resultar demasiado lentos para cumplir los plazos políticamente fijados. Los experimentos aleatorios, por ejemplo, requieren un tiempo suficiente entre la atribución de sujetos a las condiciones y la recogida de datos tras la prueba para que el tratamiento pueda seguir su curso. De forma similar, los etnógrafos requieren tiempo bastante para desarrollar su comunicación, explorar libremente aquellos aspectos del programa que les parezcan relevantes y seguir

las pistas que de allí partan; luego necesitarán un tiempo parecido para sintetizar sus anotaciones de campo en un informe final. En cualquier caso, esto puede exigir años. A no ser que las actividades, de los dos métodos, puedan tener lugar simultáneamente, en realidad es posible que no haya tiempo suficiente para ambos.

En tercer lugar, cabe que los investigadores carezcan de adiestramiento suficiente en ambos tipos de métodos para utilizar los dos. La mayoría de los estudios que combinan los dos tipos de métodos se basan, por lo común, en equipos interdisciplinarios. Cuando los miembros del equipo proceden de las tradiciones diferentes de los métodos cualitativos y cuantitativos la interacción entre los dos grupos no es siempre fluida (véase Ianni, Orr, y Trend en este volumen).

Finalmente, queda la cuestión de la moda y de la adhesión a la forma dialéctica del debate. La evaluación, como en general la ciencia, es víctima de modas. Con buen motivo, los investigadores a menudo se muestran escasamente dispuestos a no acomodarse a lo que está en boga, sobre todo, cuando los organismos financiadores participan de la tendencia general. Corrientes de este género favorecen con frecuencia a uno u otro de los tipos de métodos, pero rara vez valoran a ambos por igual. Será difícil convencer a los investigadores para que combinen los tipos de métodos hasta que se ponga en claro el carácter ilógico del debate presente con su inclinación separatista.

Por estas razones no nos sentimos optimistas ni creemos que llegará a difundirse ampliamente el empleo conjunto de los métodos cualitativos y cuantitativos. Esto no significa que los evaluadores hayan de volver a escoger entre uno u otro basados en el dogma paradigmático. Los evaluadores han de proseguir acomodando los métodos a las exigencias del problema de la investigación de la mejor manera posible, sin prestar atención a las afiliaciones paradigmáticas tradicionales. Pero ello no quiere decir que tales evaluadores hayan de establecer una prioridad de propósitos y cuestiones ya que, por lo común, serán incapaces de atender a todos los objetivos deseados. A menudo sospechamos que la cuestión del impacto tendrá la prioridad más alta y que se le otorgará prioridad a los procedimientos cuantitativos. Pero en cualquier caso, el empleo exclusivo de uno u otro de los tipos de métodos habitualmente significará aceptar una evaluación no global.

## Conclusión

El reciente debate sobre el empleo de los métodos cualitativos y cuantitativos en la investigación evaluativa ha aportado grandes beneficios. Muchos evaluadores que se formaron en la tradición cuantitativa (incluyendo los presentes autores) se han mostrado manifiestamente apasionados por el empleo de los métodos cuantitativos, en buena parte como el zapatero de Trow que sólo tiene en cuenta el cuero o como el niño que cuando coge un martillo por primera vez cree que es preciso aporrearlo todo. Como respuesta, el debate ha evidenciado que los métodos cuantitativos no son siempre los más indicados para algunos de los objetivos y de los entornos de la investigación en los que se han empleado. El debate ha contribuido adecuadamente a legitimar el incremento del empleo de los métodos cualitativos en la evaluación.

Muchos evaluadores asumieron con excesiva rapidez la serie de atributos que integran el paradigma cuantitativo. Como reacción, el debate ha centrado su atención en el proceso, el descubrimiento, la conexión con los datos, el holismo, el naturalismo y los demás atributos del paradigma cualitativo. En el pasado, el péndulo osciló excesivamente hacia el lado cuantitativo tanto en el paradigma como en el método y el debate permitió que los evaluadores tomaran conciencia de ese desequilibrio.

Pero, aunque el debate haya servido a un propósito muy útil, también ha sido en parte contraproducente, en buena medida la discusión que se desarrolla en la actualidad sirve para polarizar las posiciones cualitativa y cuantitativa y para alentar la creencia de que la única opción posible estriba en elegir entre estos dos extremos. Es como si el péndulo tuviera que oscilar hacia uno u otro lado. El excesivo interés que se percibe ahora por los métodos cuantitativos sólo puede, entonces, ser corregido mediante un interés igual, pero opuesto, por los métodos cualitativos. Naturalmente, una vez que se pongan a prueba los métodos cualitativos del modo tan concienzudo a que se sometió en el pasado a los métodos cuantitativos, los cualitativos se revelarán simplemente tan falibles y débiles como los otros (véase Overholt y Stallings, 1979). Si se mantiene la dicotomía entre los métodos, en una reacción inevitable, el péndulo volverá a desplazarse hacia los procedimientos cuantitativos. Indudablemente también este paso se revelará insatisfactorio y el péndulo seguirá oscilando. De este modo, el debate actual mantiene el desplazamiento del péndulo entre extremos de métodos y extremos de insatisfacción.



La solución, desde luego, estriba en comprender que la discusión se halla planteada inapropiadamente. No hay necesidad de escoger un método de investigación sobre la base de una posición paradigmática tradicional. Ni tampoco hay razón alguna para elegir entre dos paradigmas de polo opuesto. En consecuencia, es innecesaria una dicotomía entre los tipos de métodos y existen todas las razones (al menos en la lógica) para emplearlos conjuntamente con objeto de satisfacer las exigencias de la investigación de la evaluación del modo más eficaz posible.

Cabe desear que la próxima generación de evaluadores sea adiestrada tanto en la tradición cuantitativa como en la cualitativa. Estos investigadores podrán emplear la más amplia gama posible de métodos y sin cicaterías acomodarán las técnicas a los problemas de la investigación. Aprenderán nuevos modos de combinar los métodos y de reconciliar descubrimientos discrepantes sin rechazar arbitrariamente un grupo en favor de otro. Tales investigadores se mostrarán además modestos en sus afirmaciones acerca de los métodos. Comprenderán que todos son falibles y que el descubrimiento de un sesgo no es necesariamente una razón para rechazar un método sino un reto para mejorarlo, de la misma manera en que pugnamos por mejorar una teoría frente a los datos en contra. No será fácil tarea la adquisición de tales destrezas, pero merecerá la pena el esfuerzo.

Existe, sin duda, una cierta ventaja pedagógica en la forma dialéctica de argumentar que polariza los métodos cualitativos y cuantitativos. Por ejemplo, a menudo resulta más fácil formular un caso estableciendo una dicotomía en el continuo entre los polos opuestos para que se revele con mayor claridad la dimensión del interés. Pero ya se ha aprendido la lección de que los métodos cuantitativos se pueden emplear en exceso y ha llegado el momento de impedir que el péndulo oscile de un extremo al otro. El tiempo de dejar de alzar muros entre los métodos y de empezar a tender puentes. Tal vez estemos todavía a tiempo de superar el lenguaje dialéctico de los métodos cualitativos y cuantitativos. El auténtico reto estriba en acomodar sin mezquindades los métodos de la investigación al problema de evaluación. Puede muy bien exigir una combinación de métodos cualitativos y cuantitativos. Distinguir entre los dos mediante el empleo de etiquetas diferentes puede servir sólo para polarizarlos innecesariamente. Si prescindimos de las etiquetas no nos quedará más remedio que superar el debate entre métodos cualitativos y cuantitativos.

## Perspectiva del volumen

Al reunir los trabajos del presente volumen se pretendió deliberadamente obtener una diversidad de opiniones sobre el empleo de los métodos cuantitativos y cualitativos. En parte, la diversidad se logró seleccionando a autores de los que se sabía que representaban concepciones divergentes. Por ejemplo, Becker (Becker y Geer, 1957) y Filstead (1970) son bien conocidos defensores de los métodos cualitativos mientras que Campbell (Campbell y Stanley, 1966) ha sido desde hace tiempo considerado como defensor de los métodos cuantitativos. A la diversidad se llegó también eligiendo autores de una amplia variedad de disciplinas caracterizadas. Los campos de la antropología, la economía, la educación, la sicología y la sociología se hallan, al menos, representados cada uno por dos autores (o coautores).

El resultado es que diferentes capítulos revelan fuertes preferencias por distintos métodos. La disensión presente resulta, sin embargo, menos extremada de lo que cabría esperar. Incluso en aquellos capítulos con las preferencias más acentuadas por uno u otro de los tipos de métodos, aún subsiste la impresión de que para llegar a una evaluación global se requieren tanto los métodos cualitativos como los cuantitativos. En otras palabras, existe coincidencia en admitir que ningún tipo de método es por sí sólo generalmente suficiente para todas las diversas exigencias de la investigación evaluativa. Es posible que el debate entre los métodos esté empezando a experimentar una redefinición.

Los capítulos han sido dispuestos en este volumen conforme a un orden en el que se parte de los más tenaces y se concluye con los más precavidos en su apoyo a los métodos cualitativos. A este respecto, el capítulo de Filstead ocupa merecidamente el primer lugar. Filstead distingue las diferencias tradicionalmente sostenidas entre los paradigmas cualitativo y cuantitativo. Y afirma con vehemencia que el paradigma cualitativo (con su fe en la realidad social como construida por los participantes y acentuando la comprensión de los acontecimientos desde la perspectiva de los propios actores) es el que resulta más adecuado para la investigación evaluativa. Filstead presenta también razones del hecho de que esté cambiando la tendencia entre los evaluadores educativos desde un excesivo énfasis en los métodos cuantitativos a una apreciación de las técnicas cualitativas. Para acelerar este cambio, Filstead sugiere la forma en que cabe utilizar mejor los métodos cualitativos para modelar la investigación evaluativa.

Tal vez sorprendentemente para algunos evaluadores, pero desde luego no para todos, la aportación de Campbell proporciona una explicación teórica en pro del empleo de los métodos cualitativos en el diseño de estudio de casos. Revelando un cambio de sentimiento (y de supuestos de fondo) en relación con trabajos anteriores, Campbell afirma que el estudio de casos puede poseer rigor suficiente (en razón de "grados de libertad procedente de múltiples implicaciones de cualquier teoría") tanto para confirmar como para invalidar hipótesis causales. Campbell demuestra también la complementariedad inherente de los métodos cualitativos y cuantitativos al describir cómo el conocimiento cuantitativo sólo supera al cualitativo cuando se apoya en éste. Sin embargo, Campbell sigue convencido de que el estudio de casos es muy proclive a los sesgos cuando se emplea para valorar el impacto de un programa. Señala también varias medidas que cabe tomar para reducir, pero no eliminar este equívoco.

Del artículo de Trend podría decirse que es un relato policíaco. El autor da cuenta de sus experiencias en una evaluación sobre una gran operación de subvenciones para la vivienda en donde se emplearon tanto observadores participantes como cuestionarios. En un lugar concreto, estas dos fuentes de datos produjeron hallazgos muy discrepantes que determinaron una considerable controversia entre los analistas. Trabajando con ambas fuentes, Trend resuelve el enigma y documenta la forma en que la síntesis resultante supera los informes originales, brindando un entendimiento de los datos. Trend afirma que en un principio deberían estimularse los hallazgos discrepantes para no cerrar prematuramente pistas significativas de la investigación.

La afirmación más sobresaliente de Ianni y Orr es aquella según la cual los evaluadores pueden emplear con proyecto las técnicas etnográficas. Sólo si estos métodos se hallan cuidadosamente adaptados para atender a las demandas específicas de la investigación evaluativa, Ianni y Orr advierten la existencia de cierto número de presiones que inducen a incrementar el uso de los métodos cualitativos en la evaluación educativa pero indican que, por el momento, tales métodos se emplean a menudo de una manera inapropiada. Afirman que el empleo de estos métodos debe ser orientado por la teoría y que la teoría de las disciplinas tradicionales de la ciencia social no resulta adecuada para los fines de la evaluación. Brindan sugerencias para el desarrollo de un marco conceptual apropiado de la evaluación.

Es evidente que la inclusión del artículo de Becker puede parecer un tanto anómala. En este caso, empero, las apariencias son engañosas. Aunque el artículo se refiere ostensiblemente a la fotografía, las ideas de Becker poseen también gran relevancia para el empleo de métodos etnográficos en evaluación. El hecho es que las fotografías son semejantes a cualquier grupo de datos y que, en especial, tienen mucho en común con los datos recogidos por procedimientos cualitativos. De la misma manera en que a Becker le interesa la "verdad" de una fotografía, así también debería interesar a los evaluadores cualitativos la validez de sus datos. Reviste la mayor importancia el hecho de que Becker emplee la lógica de Campbell y Stanley (1966), originariamente desarrollada con referencia a los métodos cuantitativos, para evaluar la validez de inferencias respecto de las fotografías. Así, Becker demuestra que la eliminación de las hipótesis rivales siguen la misma lógica tanto en el procedimiento cualitativo como en el cuantitativo. Becker tiene también mucho que decir acerca de la posibilidad de generalizar, del sesgo del observador, de la censura y de la edición.

Knapp proporciona un examen bien razonado y penetrante del empleo de la etnografía en la evaluación del programa de escuelas experimentales. Al documentar las numerosas dificultades surgidas, Knapp induce al lector a apreciar las presiones tanto sutiles como no sutiles que unas evaluaciones en gran escala ejercen sobre el trabajo etnográfico de campo. Quienes piensan en que las evaluaciones en gran escala se aplica fácilmente y con éxito la etnografía sufrirán una dura sorpresa y quienes emplean la etnografía harán bien en comprender los dilemas de que da cuenta Knapp. De modo semejante, el lector se siente estimulado a considerar las recomendaciones de Knapp respecto de los usos alternativos de la etnografía en la investigación evaluativa.

Finalmente, Hollister y otros ofrecen algunos nuevos atisbos respecto del uso de los métodos cuantitativos. El pensamiento actual en la investigación evaluativa indica que el análisis de proceso requiere procedimientos cualitativos, mientras que la valoración del resultado exige técnicas cuantitativas. Hollister y otros derriban este mito demostrando no sólo que tanto el análisis de proceso como el de resultados pueden realizarse con métodos cuantitativos habituales sino, lo que es más importante, cómo es posible ligar los dos análisis. Aunque tales análisis ligados exigen datos de gran envergadura y calidad, Hollister y otros revelan los beneficios potenciales que pueden ofrecer estos análisis.

## Bibliografía

- Barzon, A. H. y Lazarfeld, P. F. (1969). "Some functions of qualitative analysis in social research" en G. J. McCall y J. L. Simmons (eds.) *Issues in Participant Observation: A Text and Reader*. Reading, MA: Addison. Westei.
- Becker, H. S. (1958). " Problems of interence and proof in paticipant observation". *American Sociological Review*. 169. 652 - 669.
- B. Geer (1957). "Participant Observation and interviewing: a comparison". *Organization* 16: 23-32.
- Boggan, R. y Taylor, S. (1975). *Introduction to Qualitative Research Methods*. New York: John Wiley.
- Boring, E. G. (1969). "Perspective: Artifact and Control". En Rosenthal, R. y Rofnon, R. L. (eds.) *Artifact in Behaxional Research*. Nueva York: Academic Press. (1954). "The nature and history of experimental control". *American Journal of Sychology*, 67:573-589.
- Boruch, R. F. (1975). " On common contations about randomized experiments for evaluating social programs." En R. F. Boruch y H. W. Riecken (eds.) *Experimental Testing of Public Policy*. Boulder Westview.
- Brickell, H. M. (1976). "Neaded: instruments as good as our eyes". *Journal of Curriculum Evaluation*. 2: 56-66.
- Britan, G. M. (1978). "Experimental and contextual models of program evaluations". *Evaluation and Program Planning*. 1: 229-234.
- Campbell, D. T. (1974). "Cualitative Hnowing in research ". Kurt Lewin Adward. *Addres, Society for the Sycological Study of Social Issues*. Reunion de la American Sycological: Association, Nueva Orleans, Septiembre 1, Journal of Social Issues.
- (1970). "Considering the case againts experimental evaluations of social innovations ". *Administrations Science Quarterly*. 15: 110-113.

(1969). "Artifact and control", en R. Rosenthal y R. L. Rosnow (eds.) *Artifact in Behavioral Research*. Nueva York: Academic Press.

J. C. Stanley (1966). "Experimental and Quasi-Experimental Design for Research". Chicago. Rand McNally.

Charman, L. J. (1967). "Illusory correlation in observational report", *Journal of Verbal Learning and Verbal Behavior*, 6: 151-155.

J. P. Chapman (1967). "Genesis of popular but erroneous psychological observations". *Journal of Abnormal Psychology*. 72: 193-204.

Cook, T. D. y Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

Cook, T. D. y Cook, F. L. (1977). "Comprehensive evaluation research and its dependence on both humanistic and empiricist perspectives." En R. S. French (eds.) *Humanists and Policy Studies: relevance Revisited: Curriculum Development in the Humanities*. núm. 3, Washington, D.C.: George Washington University, División of Experimental Programs.

Cronbach, L. J. (1978). "Designing educational evaluations". Stanford: Stanford University.

Davis, F. (1961). "Comment on logical interaction of newcomers in Alcoholics Anonymous". *Social Problems*. 8: 354-365.

Denzin, N. (1970). *The Research Art*. Chicago: Aldine.

Eisner, F. W. (1977). "Critique". *Anthropology and Education Quarterly*. 8: 71-72.

Erickson, F. (1977). "Some Approaches to inquiry school-community ethnography". *Anthropology and Education Quarterly*. 8: 58-69.

Fienberg, S. E. (1977). "The collection and analysis of ethnographic data in Educational Research". *Anthropology and Education Quarterly*. 8: 50-57.

Filstead, W. J. (ed.) (1970). *Qualitative Methodology*. Chicago: markham.

Garner, W. R. Hake, H. W. y Ericksen, C. W. (1958). "Operationism and the concept of percation". *Sychological Review*. 63: 140-179.

Glaser, B. y Strauss, A. L. (1967). "The Discovery of Grounded theory Chicago: Aldine.

(1965). "Discovery of substantive theory: a basic strateny inderying qualitative research". *American Bahavioral Scientist*. 8: 5-12.

Guba, E. G. (1978). *Toward a Methodology of Naturalistic inquiry in educational Evaluation*. Los Angeles: Universidad de California. Los Angeles, Center for the Study of Evaluation.

Kelman, H. C. (1972). "The rigths of the subject in social research,and analysis in terms of relative power and legitimacy". *American Sychologist*, 27: 969-1016.

Kuhn, T. S. (1970). *The Structura of Scientific Revolutions*. Chicago:University of Chicago Press.

(1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Light, R. (1979). "Integrating multiplie empirical studies". Presentado en el Congreso Anual de la American Education Research Association, San Francisco.

Lofrand, J. (1961). "Reply to Davis Comment on 'initial interaction'". *Social Problems*, 8: 365-367.

Lejeune, R. A. (1960). "Initial interaction of newcomers in Alcoholics Anonymous: a field experiment in class symbols and socialization". *Social Problems*. 8: 102-11.

MaCall, G. J. (1969). "The Problems of Indicators in Participant observation research". En G. J. McCall y J. L.. Simmons (eds.) *Issues in Participant Observation: A Text and Reader, Reading MA: Addison Wesley*.

Nisbett, R. E. y Wilson, T. D. (1977). "Tolling more than we can know: verbal recorts on mental processes". *Psychological Review*. 84: 231-259.

Overholt, G. E. y Stallings, W. M. (1979). "Ethnography in evaluation: dangers of methodological transplant". Presentado al congreso anual de la American Educational Reserch Association, San Francisco.

Parlett, M. y D. Hamilton (1976). "Evaluation as illumination: a new approach to the study of innovatory programs". En G. V. Glass (ed.) Evaluation Studies: Review, anual, vol. 1 Beverly Hills: Sage.

Patton, M. C. (1978). Utilization-Focuset Evaluation, Beverly Hills. Sage.

(1975). Alternative Evaluation Research Paradigm. Grand Forks: University of North Dakota Press.



## Lectura complementaria Nº 5

### Diseños de evaluación

***Tomada de: Briones, Guillermo. Evaluación de programas sociales.  
México, Editorial Trillas, 1991.***

Al presentar los modelos de evaluación señalamos que servían al investigador como guías para realizar su trabajo, en cuanto a indicarles los distintos componentes o aspectos del programa de los cuales sería apropiado obtener información y luego proceder a su análisis.

Si bien en un plano general tales esquemas o sus combinaciones pueden cumplir esa primera función, la evaluación de un programa concreto exige mayores especificaciones y precisiones, pues ahora se trata de cumplir con el requisito básico de proporcionar conocimientos que sean directamente pertinentes a la toma de decisiones que puedan mejorar la marcha y los resultados del programa en cuestión. Justamente llamamos diseño de evaluación al plan elaborado para adecuarse a esas exigencias; es decir, al plan que define las diversas estrategias y procedimientos para llevar a cabo la evaluación del programa.

Típicamente, el diseño de evaluación incluye los siguientes componentes:

1. Antecedentes y descripción del programa.
2. Objetivos generales y específicos del mismo.
3. Finalidades de la evaluación.
4. Objetivos de evaluación.
5. Diseño metodológico.
6. Informes.

Puesto que los puntos 1 y 2 de este diseño se limitan a reproducir las características principales del programa, y el tema de finalidades de la evaluación fue suficientemente tratado en el primer capítulo, esta exposición se referirá a los otros componentes del diseño.

## OBJETIVOS DE EVALUACIÓN

De manera general podemos decir que los objetivos de evaluación son los conocimientos o informaciones relativas a los objetos, instancias o características del programa que se requiere recoger y analizar de acuerdo con las finalidades que sirven de justificación al estudio evaluativo. Lo cierto es que no siempre esas necesidades de información son presentadas con claridad por el personal directivo o por el personal que ejecuta el programa en el terreno. Por ejemplo, se puede pedir que la evaluación determine “si el programa funciona bien o mal”, objetivo de conocimiento que puede tener varios significados, como: si se están produciendo los resultados buscados; si el personal técnico cumple adecuadamente sus funciones; si existe desperdicio de recursos; si la población beneficiaria está satisfecha con la atención o servicios que el programa les presta, etcétera.

En el problema reseñado más arriba puede suceder lo que Hyman señalaba años atrás, a propósito de investigaciones encargadas por un cliente o una cierta agencia (situación equivalente, en nuestra exposición, a una evaluación realizada por un investigador externo). Al referirse al trabajo de otros dos autores, el metodólogo citado afirmaba:

*Notan, por ejemplo, que el cliente a menudo no formula su problema en términos que puedan ser introducidos en un diseño de investigación y, consecuentemente, debe haber una clarificación de la formulación inicial. También hacen notar que el cliente puede, con frecuencia, formular su problema de manera que no es ni siquiera apropiada para sus propias necesidades de acción (Survey design and analysis, The Free Press, Glencoe, Ill., págs. 44 a 45).*

Una dificultad adicional en la formulación de los objetivos de evaluación que, reiteremos, son objetivos de conocimientos y, por tanto, en relación con los cuales se recogerá y analizará la información, consiste que tanto estos como los propósitos que están detrás de ellos varían en uno mismo según el nivel o las funciones que desempeñan personas

ubicadas en diferentes niveles de la estructura del programa. Así, las necesidades de información serán distintas para él o los administradores del programa, encargados de la política general que se relaciona con la ampliación, la reducción o la modificación del programa, que para los ejecutores directos de éste, quienes podrán estar interesados en saber qué opinan los beneficiarios de la metodología empleada, del uso de los recursos instrumentales, etcétera.

## OBJETIVOS DESCRIPTIVOS

Cuando la evaluación es de tipo descriptivo los objetivos de evaluación se derivan directamente de los objetivos del programa y de los otros componentes que se han considerado focos u objetos de evaluación. En este caso, los objetivos suelen redactarse de la manera que se detalla a continuación:

Los objetivos de la evaluación son los siguientes (en este caso se refieren directamente a los objetivos del programa):

1. Evaluar el efecto sobre los beneficiarios de las inversiones en términos de aumento de la productividad, ingreso incremental y distribuciones intraaldeanas del ingreso, organización social de las aldeas en materia de actividades productivas y de comercialización, efectos de empleo y perspectivas de crecimiento autónomo de largo plazo.
2. Analizar la participación popular y las cuestiones de gestión y ejecución del PIDER en los planos de la explotación, el ejido y la microrregión, comparando las previsiones con las realizaciones (Michael M. Cernea. *Lo cuantificación de los efectos de los proyectos: la vigilancia y la evaluación en el Proyecto de Desarrollo Rural PIDER*. México, Banco Mundial, Documento de Trabajo núm. 332, Washington, 1979; págs. 64 a 65).

En este caso, los objetivos del programa eran, respectivamente:

1. Elevar los ingresos y niveles de vida del sector rural mediante la introducción de actividades directamente productivas, y

2. Lograr la participación de los beneficiarios en el proceso de planificación de esas actividades.

Otro ejemplo es el siguiente:

La evaluación se propone medir los siguientes componentes y características del programa (se nombran los objetos y variables de evaluación):

- Funcionamiento administrativo: cumplimiento de funciones, calificación del personal, etcétera.
- Resultados: niveles de logro de los objetivos, calidad, distribución entre los usuarios, etcétera.

El término evaluar que se utiliza en la formulación de estos objetivos de evaluación directos significa: 1. realizar descripciones cualitativas; o 2. mediciones de los logros alcanzados (por el «efecto sobre los beneficiarios», del nivel de «participación popular», etc., en el ejemplo dado) así como, en algunos casos, descripción de la calidad y distribución diferencial de esos logros, y 3. comparación de tales logros cuantitativos y cualitativos con alguna norma o meta, criterio deseado, con el fin de pronunciarse respecto del éxito o fracaso del programa.

En evaluaciones que utilizan diseños experimentales o cuasiexperimentales, los objetivos de evaluación se redactan frecuentemente en forma de hipótesis: «Los niños sujetos al programa de estimulación temprana mostrarán ganancias significativas en su desarrollo integral en comparación con los logros obtenidos por los niños de los grupos de control».

Cuando se definen los objetivos de evaluación en algunas de las formas que hemos presentado, lo conveniente sería establecer por anticipado las normas de evaluación o criterios de éxito o fracaso del programa. Así se hizo, por ejemplo, en la evaluación del programa de estimulación con base en la familia. Al referirse a los resultados esperados se dice: «Los niños expuestos al programa mostrarán un porcentaje de logro óptimo del 70% en los curriculares en las tres áreas de desarrollo». Respecto del trabajo de los ejecutores directos del programa se agrega: «Las animadoras tendrán un desempeño del 70% de logro en el trabajo metodológico que realizan con las familias».

En suma, para seguir con el ejemplo, en la misma evaluación se propuso una escala valorativa porcentual para interpretación de porcentajes referidos a las características del personal ejecutor, manejo de instrumentos, etc., con la estructura que se muestra en la tabla 5.1.

**Tabla 5.1.** Escala valorativa porcentual para la interpretación de porcentajes.

Valor	Intervalo
Alto	70 a 100%
Mediano	50 a 69%
Bajo	30 a 49%
Muy bajo	0 a 29%

**Fuente:** *María Alcántara y Cipriano Olivera.* Evaluación formativa del programa 'de estimulación temprana can base en la familia. Lima: Inide, 1983, pág. 31.

## OBJETIVOS EXPLICATIVOS

La formulación de objetivos de evaluación tiende a complicarse cuando la evaluación no se refiere directamente a los objetivos o a los componentes del programa, sino a factores que expliquen los niveles o calidades de logro de los resultados finales o de aspectos instrumentales (funcionamiento, metodología, etc.). Esto obliga, en primer lugar, a pensar en la incidencia de diversas otras variables del programa en tales resultados y, por lo mismo, a formular a continuación hipótesis que conecten esas características con la situación que se desea explicar. Así, si el objetivo de evaluación fuera «evaluar la eficiencia interna del programa y establecer los factores que inciden positiva o negativamente en ella», supuesto el caso que nos hemos puesto de acuerdo con el concepto de eficiencia interna, deberíamos luego buscar factores o variables que podrían estar vinculados en la eficiencia, para formular algunas hipótesis pertinentes. De más está decir que en tal búsqueda se manifiesta la formación, experiencia e imaginación del evaluador.

Algunos ejemplos ayudarán a clarificar la situación que se presenta en la determinación de evaluación y la búsqueda de factores explicativos:

Cabe preguntarse si un mal desarrollo del programa tiene por causa ciertas deficiencias en la organización de los cursos a dificultades originadas en el propio programa... será interesante investigar, por ejemplo, si ciertas decisiones como la elección del lugar y del horario de los cursos, ciertas características del grupo (sexo, edad, nivel de calificación), ciertos detalles de la organización material (ayudas recibidas), ciertas características de los instructores (sexo, edad, homogeneidad socioprofesional) y, finalmente, ciertos aspectos de los gastos (material, mantenimiento, costo medio del grupo) tienen efectos positivos o negativos, estadísticamente comprobados, sobre el desarrollo de los cursos (Roger Couvert. *La evaluación de programas de alfabetización*. París: UNESCO, 1979, pág. 118).

Con el objeto de orientar adecuadamente la evaluación de la eficiencia actual de las Escuelas Radiofónicas, que es el segundo objetivo general, se diseñaron los siguientes particulares (cada uno implica hipótesis sobre factores que inciden sobre las características por evaluar):

- a) Fundamentar posibles recomendaciones sobre la selección y capacitación de los maestros auxiliares.
- b) Fundamentar posibles sugerencias sobre la selección y capacitación del personal empleado en la transmisión.
- c) Fundamentar posibles sugerencias sobre la supervisión de las Escuelas Radiofónicas.
- d) Fundamentar posibles sugerencias acerca de los alicientes que deben usarse para lograr una mayor asiduidad escolar.
- e) Fundamentar posibles sugerencias sobre el contenido de las transmisiones.
- f) Fundamentar posibles sugerencias sobre el uso de ayudas didácticas.

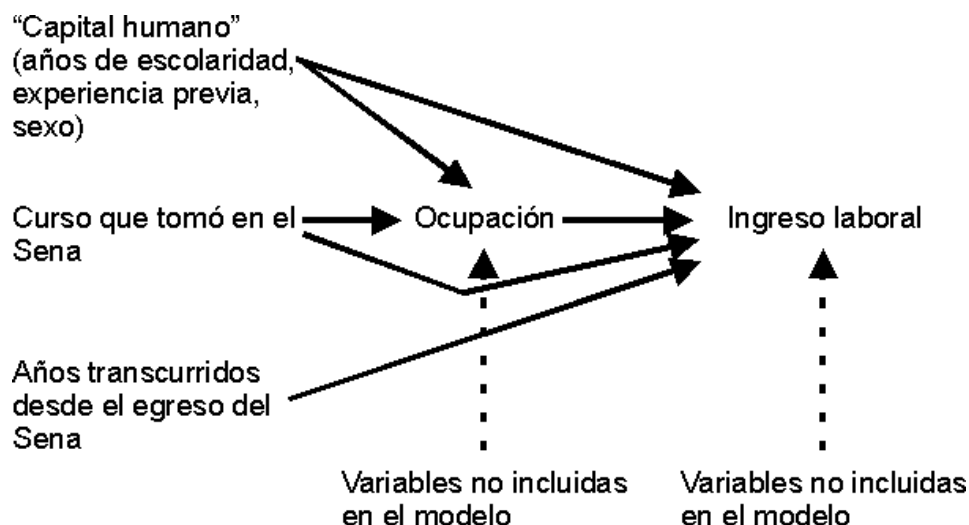
- g) Fundamentar posibles sugerencias por la acción pedagógica del maestro auxiliar. (Sylvia Schmelkes. «Investigaciones sobre las escuelas radiofónicas de la Tarahumara». *Revista del Centro de Estudios Educativos*. México, vol. 1, núm. 3, pág. 128).

Como es fácil observar en los ejemplos anteriores, los objetivos de evaluación contienen relaciones de determinación entre la variable de evaluación (el desarrollo del currículo, en el primer caso y la eficiencia interna del programa, en el segundo) y características correspondientes a otros objetos o componentes del programa (población-objeto, insumos, grupo técnico, etcétera).

En algunas evaluaciones explicativas, los objetivos de evaluación contienen uno o más modelos que pretenden mostrar las diversas interconexiones causales de los factores que se relacionan con los resultados que se desea explicar (resultados de funcionamiento, de objetivos, etc.). Tales modelos, tratados mediante la técnica del análisis de senderos (*path analysis*) permiten establecer las contribuciones diferenciales de las variables consideradas en distintas instancias del programa. Un modelo de este tipo se utilizó en la evaluación de las funciones del Servicio Nacional de Aprendizaje, SENA, en la cual el salario de una persona que tomó cursos en la institución se consideró determinado por tales cursos, su escolaridad, experiencia previa, etc. (Sena, Sistema de planeación de recursos humanos. Bogotá: Sena, 1981, pág. 182) (véase la fig. 5.1).

En suma: la formulación de objetivos en evaluaciones explicativas asume formas como las que se detallan a continuación:

1. La evaluación se propone determinar los factores que facilitan (o dificultan):
  - a) El logro de los objetivos, en cuanto a su nivel y calidad.
  - b) El funcionamiento del programa.
  - c) La aplicación de su metodología, etcétera.



*Figura 5.1 Modelo causal para explicar los ingresos laborales.*

2. La evaluación se propone determinar... (al igual que en la primera forma). Por ello se considerarán las siguientes variables correspondientes a los insumos, procesos, grupo técnico y población de usuarios (aquí se incluiría una lista de esas variables).
3. La evaluación se propone someter a contrastación empírica el siguiente modelo causal (aquí se daría un modelo de relaciones entre variables dependientes e independientes como el que se muestra en la fig. 5.1).

Cuando en las formulaciones anteriores se mencionan «factores que facilitan o dificultan» tales o cuales resultados o la marcha del programa, debe entenderse que dichos factores pueden corresponder a indicadores objetivos o subjetivos (por ejemplo, en este último caso, opiniones de los usuarios, actitudes del grupo técnico, etcétera).

## Un ejemplo de objetivos de evaluación

La transcripción textual que presentamos a continuación es un buen ejemplo de objetivos de evaluación que cubren focos como los objetivos mismos del programa, funcionamiento, usuarios, metodología y recursos instrumentales:



1. Se analizará de qué manera y hasta qué punto se han logrado cada uno de los objetivos específicos del proyecto tal como ellos han sido planteados en la formulación original del mismo. Es decir, los siete aspectos allí enumerados referentes al objetivo de masificar la creación de talleres de aprendizaje, así como el objetivo de efectuar un seguimiento analítico del proceso de difusión.
2. Se pondrá énfasis en el análisis de las diferentes modalidades de funcionamiento que han sido desarrolladas frente a diferentes sectores, centrando el análisis en el tipo de gestión que se lleva a cabo localmente.
3. En el análisis también será destacado hasta qué punto y de qué forma los pobladores están asumiendo la responsabilidad por la puesta en marcha y ejecución de los talleres, y cuál es el aporte a este proceso de la capacitación respectiva ofrecido por el proyecto.
4. Sabiendo que la estrategia establecida en los diferentes talleres está variando de acuerdo con la participación de diferentes sectores con características propias, se analizará qué efectos tienen en los niños estas modalidades de atención.
5. ¿Hasta qué punto estas diferentes modalidades establecidas en la presente etapa del proyecto conducen a ajustes o mejoramientos en el modelo (o «producto») desarrollado en la etapa anterior y descrito en el informe de la primera evaluación?
6. Se analizarán calidad y utilidad de los materiales pedagógicos producidos.
7. Se estudiarán la utilidad y eficiencia de la estructura organizativa, la administración y el manejo financiero del proyecto\*.

## DISEÑO METODOLÓGICO

El diseño metodológico de la evaluación es similar al diseño correspondiente de la investigación social y consiste en la estrategia para reunir, procesar y analizar la información

\* FLACSO. «Informe de evaluación del Proyecto Difusión del Proyecto 'Difusión y Estrategias de Transferencia de los Talleres de Aprendizaje a Comunidades Populares'. ejecutado por PIIE». Autores: Jan Ooigens y Jorge Chateau, con colaboración del profesor Jorge Tagle. Términos de referencia más específicos, noviembre de 1984, págs. 4 a 5.

que se requiere para satisfacer los objetivos de evaluación. De esta manera, en un nivel general, el diseño consiste en los procedimientos que se utilizarán para resolver las siguientes tareas:

1. Determinación de la efectividad del programa.
2. Universo de estudio.
3. Unidades de análisis.
4. Diseño muestral.
5. Instrumentos para la recolección de informaciones.
6. Indicadores y medición de variables.
7. Trabajo de campo.
8. Técnicas de análisis.

Tanto en esta sección como en los capítulos finales desarrollaremos los aspectos principales de estas tareas del diseño metodológico.

### **Determinación de efectividad**

Así como por definición el objetivo básico de todo programa de acción es producir alguna modificación o cambio en una cierta población, la tarea principal de la investigación evaluativa consiste, paralelamente, en establecer si en los hechos el programa, mediante la metodología empleada se alcanzan o no el o los objetivos propuestos. Se trata, como se advierte, de verificar la efectividad del programa.

Ahora bien, puede suceder que tal efectividad haya sido comprobada en la validación del programa. En este caso, la tarea del evaluador consiste en examinar críticamente la validación realizada, con especial atención a la validez ecológica del programa experimental. Sin perjuicio de ello, la evaluación de tipo intermedio o terminal deberá encontrar pruebas de la efectividad de las acciones que se realizan en el programa. Si no ha habido tal

validación, la tarea se presenta con mayor exigencia. La pregunta es: ¿cómo procede la investigación evaluativa cuando el programa ya se encuentra funcionando? La respuesta puede encontrarse en alguno de los tipos de diseño de validación que presentamos antes (experimentales y cuasiexperimentales) o en otros procedimientos distintos, pero que cumplen con la misma finalidad.

### **Uso del diseño experimental clásico**

Es posible emplear el diseño experimental clásico cuando existen poblaciones semejantes a la población en la cual se aplica el programa, de tal modo que, mediante inferencia, el evaluador pueda pronunciarse acerca de la efectividad de éste. En el ejemplo que utilizamos, el programa se realiza en un cierto número de escuelas, pero como no todas las de una región o país participan en el mismo, es posible seleccionar otras que no lo hacen y utilizarlas como grupos experimentales y de control.

### **Un ejemplo de aplicación**

El ejemplo está constituido por la evaluación del programa denominado Círculo Mágico, realizado por Joel M. Moskowitz y otros investigadores entre 1978 y 1979 («Process and outcome evaluation in primary prevention: The Magic Circle Program». *Evaluation Review*, vol. 6, núm. 2, dic. de 1982, págs. 775 a 778).

El Círculo Mágico es el nombre popular del Programa de Desarrollo Humano de California, cuyo objetivo principal es influir en el desarrollo social y emocional de estudiantes del nivel elemental. La estrategia consiste en la constitución de pequeños grupos de discusión en el salón de clases. En una sesión, que dura de 15 a 30 minutos, el profesor guía a un grupo de 5 a 12 estudiantes en la discusión de un tema particular. Y emplea métodos que estimulan la expresión de ideas y sentimientos por parte de los estudiantes. La participación es siempre voluntaria y no se permiten la confrontación ni el debate. El elemento clave del Círculo es que cada alumno tenga una oportunidad de contribuir, aun cuando él o ella prefiera permanecer en silencio; por otro lado, cada participante es escuchado atentamente por los otros miembros del Círculo.

Para la evaluación fueron procesadas 13 escuelas elementales públicas, predominantemente de clase media, del norte de Carolina, sobre la base de las

características de sus estudiantes, profesores, etc. Una por una, las escuelas fueron asignadas al azar a la condición experimental y a la condición de control. En definitiva, el grupo experimental estuvo formado por 102 niños y 115 niñas y el control, por 131 niños y 119 niñas, todos ellos del tercer grado.

A fin de evaluar la puesta en marcha de la estrategia del Círculo Mágico se utilizaron diversos métodos: 1. informes detallados de las sesiones de entrenamiento de los profesores, preparados por un observador; 2. control de asistencia de los profesores a las sesiones de entrenamiento; 3. encuestas a los profesores después de las sesiones de entrenamiento para obtener retroinformación acerca de las sesiones; 4. entrevistas individuales con profesores y directores seleccionados para obtener información más detallada sobre reacciones al entrenamiento; 5. cuestionarios a los profesores después del entrenamiento y nuevamente al final del año escolar, a fin de vigilar la marcha de la estrategia y evaluar la adecuación del entrenamiento; 6. observaciones de clase por parte de profesor especialmente entrenado, mientras los maestros conducían las discusiones con sus alumnos; 7. informes semanales de los profesores acerca de la frecuencia con la cual realizaban sesiones de Círculo en sus salas de clase; y 8. entrevistas con profesores seleccionados del grupo de control con el fin de identificar cualquier difusión del entrenamiento en la estrategia del programa.

La situación de los estudiantes en los grupos experimental y de control, en las variables que habían de evaluarse, fueron medidas antes y después de iniciarse el programa, mediante escalas de autoobservación referidas a relaciones con sus compañeros, sus profesores y sus escuelas. En la prueba «después» se midieron también la creencia en el autocontrol, la responsabilidad en los procesos escolares y la autoestima.

El análisis de los datos reveló que el Círculo Mágico constituyó una experiencia positiva para los profesores experimentales. Les agradó el entrenamiento y realizar las sesiones de discusión con sus estudiantes. Comparados con los profesores de control, los experimentales mostraron mayor incremento en su satisfacción personal al impartir la enseñanza.

En lo que toca a los estudiantes, hubo poca evidencia de los efectos que se suponía debía producir el Círculo Mágico. No se encontraron efectos en las niñas y en los niños se constató sólo uno de los efectos previstos. Comparados con los de control, los niños

del grupo experimental tenían mayor autoestima social después de participar en el Círculo Mágico. Este efecto aislado no se mostró relacionado con el número de sesiones de discusión en las cuales participaron los niños experimentales. En el hecho, ninguno de los resultados de los niños tuvo relación con su exposición al Círculo Mágico.

En vista de los resultados obtenidos por su estudio y por otros realizados con anterioridad, los investigadores sugieren dos posibles conclusiones: o el efecto del Círculo Mágico en los estudiantes es idiosincrático o bien es espurio.

## **Problemas para la aplicación de diseños experimentales**

Al tratar de la validación de programas demostrativos pudimos apreciar las bondades de algunos diseños experimentales y cuasiexperimentales en lo que se refiere al control de diversos factores que pueden reducir o anular su validez interna. Sin embargo, la posibilidad de su uso queda bastante limitada debido a diversos problemas que se presentan para su aplicación a situaciones concretas. Entre ellos podemos destacar los siguientes:

1. Con mucha frecuencia, la evaluación de un programa se inicia cuando éste ya se ha comenzado. No hay posibilidad, por lo tanto, de constituir grupos experimentales y de control por asignación al azar, como tampoco de hacer mediciones «antes».
2. En aquellos casos en que la formación de tales grupos es posible (recuérdese el ejemplo del Círculo Mágico), no siempre resulta fácil controlar la «contaminación» entre las personas del grupo experimental y las del grupo de control (comentan los procedimientos utilizados, los objetivos que perciben del programa, etcétera).
3. En el caso de poder aplicar algún diseño experimental también es difícil mantener constante el número de participantes y la relativa igualdad de los grupos.
4. Existe un problema ético al seleccionar sólo algunas personas para formar el grupo experimental, especialmente cuando las actividades del programa significan algún beneficio para la población: ¿por qué proporcionar servicio a esas personas y no a las otras?

Por éstas y otras razones, la determinación de la efectividad de un programa en marcha debe hacerse por diseños que en general reciben el nombre de no-experimentales (si

bien en circunstancias especiales es posible utilizar algunos tipos cuasiexperimentales como los que vimos anteriormente).

### **Diseño con medición sólo «después» con grupo de comparación**

Este diseño trabaja con dos grupos: uno formado por la población objeto y otro constituido por un grupo externo de comparación. Ello con el fin de lograr un cierto nivel de control de la validez interna. Se supone, desde luego, que las personas a las cuales se aplica el programa mostrarán resultados diferentes en las características o variables de evaluación, que los integrantes del otro grupo. La tarea principal en este diseño es lograr un grupo de comparación lo más semejante posible a la población del programa.

La situación «después» puede ser lograda mediante procedimientos de descripción (como la observación, entrevistas, observación participante, etc.), o bien mediante el uso de algún recurso para la medición de variables, según sean las necesidades de información o los criterios básicos del modelo de evaluación.

La utilización de un grupo de control permite al investigador controlar, en cierta medida, algunos factores de invalidación, como la «historia» y, en menor medida, la «maduración». Sin embargo, como no se puede afirmar que los dos grupos sean equivalentes en cuanto no han sido formados al azar, puede estar presente el factor de «selección», con lo que tampoco se puede tener la certeza de si un grupo era diferente del otro en la variable dependiente antes de aplicar el programa, debido a lo cual se dificulta establecer el verdadero valor de la medición «después».

### **Un ejemplo de aplicación**

El ejemplo está dado por la evaluación del Programa de Perfeccionamiento para Profesores en Servicio -PPS- en la asignatura de ciencias naturales de la enseñanza básica. PPS fue una iniciativa del Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas dependiente del Ministerio de Educación de Chile, con el objetivo –como su nombre lo indica– de perfeccionar de forma masiva a profesores de ciencias naturales, sin que abandonaran sus lugares de trabajo. En 1971, el PPS contaba con 1.200 profesores que atendían 20.000 alumnos de 5 a 8 grado de la enseñanza básica.

La metodología del PPS es un tipo especial de enseñanza por correspondencia y de enseñanza programada. Todos los profesores inscritos en el programa recibían los siguientes materiales, preparados en el Centro por el personal encargado del mismo:

1. *Material de autoinstrucción.* Contiene los aspectos teóricos de cada uno de los procesos científicos que han de desarrollarse en el aula, de acuerdo con el programa de su curso o nivel (observar, medir, etcétera).
2. *Comentarios acerca de los niveles del proceso.* Desarrollan y discuten los distintos niveles jerárquicos dentro de cada uno de los procesos, como también las sugerencias para evaluar a los alumnos en estos aspectos.
3. *Actividades previas al desarrollo del programa.* Comprenden una serie de actividades específicas que el profesor debe realizar antes de plantear una situación nueva de aprendizaje a sus alumnos, con el objeto de conocer y anticipar problemas que puedan plantearse.
4. *Guiones metodológicos.* Plantean sugerencias para organizar el trabajo en el aula en situaciones particulares. Se analizan objetivos, materiales necesarios y secuencias en las actividades de los alumnos.
5. *Material de evaluación.* Contiene diversos materiales que permiten ponderar aspectos cuantitativos del trabajo del profesor, en particular con el material de autoinstrucción. Tales evaluaciones eran remitidas por cada profesor al equipo PPS de Santiago con el fin de realizar su análisis.

Además de estos materiales y otros adicionales, los profesores en servicio recibían visitas periódicas del equipo de Santiago con el fin de resolver en el terreno las dificultades y dudas que pudieran haberse presentado.

La evaluación tuvo como objetivo central medir los resultados logrados en los procesos de observar, medir, comunicar y clasificar tanto en el nivel de los profesores como de los estudiantes que en 1971 cursaron 5º. año básico. El estudio se llevó a cabo durante el mes de mayo de 1972.

Puesto que el programa al comenzar no tomó mediciones de entrada y no formó un grupo de control al azar, se eligió un diseño no experimental con sólo mediciones «después» y un grupo de comparación no equivalente constituido por profesores que se inscribieron en el PPS, pero que lo abandonaron antes de tener alguna participación en él. El grupo experimental final quedó constituido por 52 profesores, y el grupo de composición, por 28. Ambos grupos presentaron características demográficas profesionales muy similares.

La evaluación de los procesos de observar, medir, comunicar y clasificar: como lo que se quería someter a medición era el efecto del PPS en la conducta de los profesores –y no sus conocimientos y actitudes–, los ejercicios para ese objetivo consistieron en realizarlos efectivamente, para lo cual se proporcionó a cada profesor los elementos necesarios (un globo, una balanza, una regla, etc.). El instrumento tenía 100 puntos con ítems como los siguientes: «mida el ancho de esta página y anote el resultado, indicando de alguna forma la incerteza de su medición» (10 puntos).

A los alumnos de los profesores de los grupos experimentales y de control se les aplicó una prueba semejante con el fin de evaluar la posible transferencia del perfeccionamiento docente. También se aplicó un cuestionario a los profesores respecto de su satisfacción en el trabajo, contenidos de los programas de ciencias naturales, opiniones acerca del PPS, etc. Los alumnos, por su parte, respondieron a un cuestionario sobre las ciencias naturales.

Los principales resultados de la evaluación fueron los siguientes:

1. El PPS eleva las habilidades científicas de los profesores en los procedimientos de observar, medir y clasificar, pero no en el de comunicar.
2. Los profesores PPS alcanzan un promedio global superior cuando los resultados obtenidos en la prueba se evalúan mediante una norma deseada.
3. La evaluación del PPS respecto de una norma muestra que los logros totales en los cuatro procesos científicos quedan bajo esa forma: promedio de 49 puntos



contra un valor esperado de 58 a 60 puntos (la norma fue propuesta por el equipo PPS).

4. Los alumnos con profesores PPS logran, en promedio, mayores puntuaciones en los procesos de comunicar y clasificar, pero no en los de observar y medir.
5. El rendimiento global de los alumnos PPS es más bien bajo: 37 puntos de promedio en una prueba de 100 puntos. (Guillermo Briones. *Metodología de la investigación evaluativa*. Bogotá: Universidad Pedagógica Nacional, 1975, págs. 53 a 102).

### **Diseño con mediciones «antes» y «después»**

Este diseño no experimental trabaja con un solo grupo –el formado por la población del programa– y, por lo tanto, sin grupo de control. La medición «antes» puede ser una descripción (observación) o una medición de entrada de las características del programa que se desean modificar para realizar posteriormente una descripción o medición «después», con el fin de comparar ambas situaciones y apreciar el monto de los logros, su calidad y la distribución en subgrupos de la población objeto.

Según el tipo de características, la situación «antes» puede lograrse con fuentes secundarias (estadísticas, documentos, etc.) o mediante preguntas retrospectivas a los usuarios del programa, del tipo “¿qué hacía usted antes que comenzara este programa de salud, cuando sus hijos estaban enfermos?”.

Como justificación del uso de este diseño y del siguiente debe tenerse presente que cuando un programa actúa en un contexto o población en la cual no se aplican programas similares (presenciales o a distancia), los resultados obtenidos, definidos en los objetivos, pueden atribuirse, con mayores o menores niveles de confianza, a las acciones del programa. A este punto se refiere Riecken cuando afirma:

Se puede argumentar que allí donde el programa es suficientemente extensivo y la dirección de su influencia es claramente diferenciada de otras fuerzas concurrentes no se necesita de un grupo de control para demostrar su efectividad (Henry W. Riecken, «Memorandum: Program evaluation», en Carol N. Weiss (ed.). *Evaluating action programs: readings In social action and education*. Boston: Allyn and Bacon, Inc., 1972, pág. 97).

Si no reúne lo anotado antes, el investigador no puede tener la certeza de que el cambio que haya podido mostrar la diferencia entre las mediciones «antes» y «después» haya sido producido efectivamente por el programa, pues el mismo podría deberse a la influencia de otras variables extrañas. Otras debilidades de este diseño, debidas a la falta de un grupo de control, son la incapacidad de controlar la «maduración», la «reactividad» a los instrumentos de medición y la regresión estadística. De ahí que su uso, como en general el uso de diseños no experimentales, deba acompañarse de los mayores cuidados en la recolección de la información, en su análisis y en la búsqueda de elementos de validación distintos de los que proporcionan los diseños experimentales y cuasiexperimentales.

### **Diseño con medición sólo «después»**

Como su nombre lo indica, es un diseño que, igual que el anterior, también no experimental, no utiliza grupo de control, y el grupo experimental está formado por los usuarios del programa; a diferencia de él, sin embargo, únicamente puede disponer de una sola medición: la medición «después», realizada en algún momento de la marcha del programa o justamente a su término. Debido a estas características, el diseño está abierto a la mayoría o a todos los factores que actúan contra la validez interna, en especial a la intervención de factores externos al programa que pudieron producir los cambios detectados por medición «después», si es que realmente se produjeron, pues no se dispone de información respecto de la situación de entrada de los usuarios; o bien esos cambios se produjeron por la selección sesgada de tales usuarios o por la maduración biopsicológica de los mismos.

Por estas razones, la evaluación de los resultados con este diseño debe apelar a todos los recursos que sea posible utilizar para probar que fue el programa el que produjo los cambios (exclusividad del programa, características de los usuarios, etc.).

El más usado del diseño de un solo grupo y medición «después» es la encuesta social, ya sea de cobertura total o por muestreo. Los datos recogidos mediante esta técnica pueden ser cuantificados y empleados en la búsqueda de conexiones causales entre los resultados del programa y las características de sus componentes, por medio del análisis de senderos (*path analysis*), por ejemplo. Por otro lado, la información que ha de recogerse puede cualificarse y ampliarse mediante entrevistas en profundidad, preguntas abiertas, etcétera.

La ausencia de grupos de control y de mediciones «antes» imposibilitan la apelación a las características de los diseños experimentales para asegurar la validez interna del programa y de manera particular confiar en la efectividad de las acciones instrumentales para producir cambios en los usuarios, de acuerdo con la intención contenida en los objetivos. Por ello, además de comprobar que el programa es típico en la zona de aplicación; es decir, que no existan otros programas que puedan producir los mismos resultados del programa en evaluación, es necesario recurrir a técnicas de análisis para poder verificar los efectos de este último.

Una manera de proceder es averiguar en la población en referencia qué personas han sido afectadas por otros programas con objetivos similares. Si existen estas personas, se les trata separadamente. En los otros, luego de establecer los niveles de resultados, se buscan relaciones con modalidades de funcionamiento propios del programa (asistencia a reuniones, exposición a diversos medios instrumentales, etc.), de la forma que se muestra en los ejemplos de los cuadros 5.1 y 5.2.

		Alta	Baja
Logros:	Altos	$P_1$	$P_2$
	Bajos	$q_1$	$q_2$
		100%	100%

**Cuadro 5.1.** Asistencia a reuniones del programa.

		3 y +	– de 3
Logros:	Altos	$P_1$	$P_2$
	Bajos	$q_1$	$q_2$
		100%	100%

**Cuadro 5.2.** Exposición a medios instrumentales (folletos, conferencias, audiciones de radio, etcétera).

Si el programa ha sido efectivo, entonces, en el primer caso,  $P_1 > P_2$  y, en el segundo,  $P'_1 > P'_2$ . Sobre estos y otros procedimientos de validación y de comprobación de la efectividad de un programa volveremos en los capítulos destinados a presentar algunas técnicas de análisis.

## Ejemplos de aplicación

Ejemplificaremos el uso del anterior diseño con dos investigaciones evaluativas de especial interés. La primera, en cuanto utiliza como unidades de análisis conglomerados de personas, con fuentes secundarias de información; la segunda, porque se refiere a un programa de educación a distancia de gran magnitud y en la cual es posible apreciar la utilización de un diseño de evaluación que es, en varios aspectos, similar al que estamos exponiendo en este capítulo. Así se justifica, además, la extensión de dicho ejemplo.

## Evaluación de servicios de salud mental

El estudio que sirve de ejemplo –de alta elaboración metodológica– explora el uso del análisis de áreas sociales (conglomerados) como un medio para evaluar la repercusión de los Centros de Salud Mental Comunitaria en Nueva York. La ley federal de 1963,

creadora del programa, contenía varios objetivos de la profesión psiquiátrica: proporcionar opciones para atender en hospitales grandes y distantes; brindar atención en las comunidades de residencia con algunos recursos; dar un amplio conjunto de probabilidades de hospitalización: clínicas, hospitalización parcial, rehabilitación, consultas y servicios educacionales; proporcionar continuidad en la atención entre esos componentes, etc. Luego se agregó una función de especial importancia: orientar los servicios hacia las comunidades en lugar de centrarlos en los individuos, con la participación de los grupos comunitarios en las decisiones.

Algunos estudios previos habían explorado extensamente la repercusión de las características individuales, tanto en el número de personas que recibían los servicios como en los resultados obtenidos. Según los autores de la evaluación, era conveniente un enfoque diferente para medir los efectos en la comunidad. Dicho enfoque significó, entre otros aspectos, medir las relaciones entre las tasas de utilización por cada tipo de servicio y un conjunto de características de las comunidades, tales como estatus socioeconómico, condiciones de la vivienda, organización (desorganización de la familia). En el estudio de esas relaciones se mantuvo constante la accesibilidad geográfica a los centros de salud.

Para efectos de evaluación, se derivó un conjunto de hipótesis de los objetivos del programa, de tal modo que pudieran contrastarse empíricamente. Por ejemplo, «cuando las características sociales del área se controlan, las tasas totales de utilización serán diferentes en áreas cubiertas por centros de salud fiscales, centros municipales y centros voluntarios». El conjunto de hipótesis se combinó en un modelo causal (path diagram).

De acuerdo con el enfoque distintivo del diseño, la unidad de análisis estuvo constituida por «un área de salud». De las fichas del censo de población de 1970 se tomó una muestra de 333 de tales unidades, de cada una de las cuales se extrajo la información requerida para la prueba de las hipótesis (variables colectivas: mediana de los ingresos, presupuesto de todas las clínicas localizadas en una misma área de influencia, porcentaje de madres en la fuerza de trabajo, promedio de escolaridad, etc.). La técnica principal de análisis fue la de regresión múltiple, pero también se introdujeron modificaciones para incorporar relaciones que no reunían los requisitos de ser lineales, aditivas y continuas.

Los principales hallazgos de la evaluación fueron los siguientes:

1. Las características sociales del área parecen constituir el mayor factor causal para las tasas de utilización de los servicios de salud.
2. Existe un considerable sobrecubrimiento en los efectos de las características sociales de las áreas y las facilidades de cobertura; parecen reforzarse mutuamente.
3. No se observaron efectos de prevención en el nivel de análisis de la comunidad; esto es verdad para los centros de salud mental, clínicas, hospitales o el total de servicios municipales y voluntarios.
4. Todo el sistema de servicios está estratificado por niveles de ingreso, con diferentes conjuntos de relaciones para los grupos de ingresos bajos y medios (Richard N. Brandon, «Differential use of mental health services: social pathology or class victimization?», en M. Guttentag y E. Struening, (eds.). *Handboock of Evaluation Hesearch*. Beverly Hills: Sage Publication, CA., 1975, págs. 341 a 429).

## **Evaluación de un sistema no formal de educación a distancia**

El sistema en referencia está constituido por el programa de educación no formal a distancia, en el nivel nacional, denominado Acción Cultural Popular, ACPO, que trabaja en Colombia desde 1947, y se ha orientado a la población rural del país.

A través de los años, el objetivo central de ACPO ha sido fundamentalmente el mismo: proporcionar educación básica para el desarrollo social y personal a través de un sistema complejo de medios masivos, apoyados por la comunicación grupal e interpersonal. Ese sistema incluye los siguientes componentes:

1. Radio Sutatenza. Es una radio del programa que transmite cursos de educación diseñados para lograr la alfabetización básica de los usuarios. Para los campesinos que no pueden escucharlo con regularidad, el curso completo ha sido reproducido en una serie de discos y está disponible junto con un tocadiscos.

Los otros cursos se centran en las siguientes áreas: salud y trabajo, números y comunicaciones, y la vida en la comunidad y sus problemas. En 1979, el programa

de radio educativa ofrecía seis horas diarias de transmisión con cuatro cursos que, en conjunto, desarrollan los objetivos de ACPO.

2. La escuela radiofónica no es una escuela en el sentido formal del término, sino un ambiente educativo –por lo general, la casa de un campesino– en donde se reúnen hasta diez participantes, miembros o no de una misma familia, niños y adultos. En este ambiente se escuchan transmisiones de la radio de tal modo que en el proceso de aprendizaje son ayudados por un auxiliar voluntario. Se utilizan además las cartillas, los libros de la biblioteca y el periódico.
3. Las cartillas y la Biblioteca del Campesino. Las cartillas tratan los contenidos básicos del programa ACPO. Esas cartillas son:
  - *Cartilla básica*
  - *Nuestro bienestar*
  - *Hablemos bien*
  - *Cuentas claras*
  - *Suelo productivo*
  - *Comunidad cristiana*

La Biblioteca del Campesino, con más de 50 títulos, complementa los programas radiales y las cartillas. Incluye temas como el cultivo de árboles frutales, piscicultura y pesca, cría de cerdos, apicultura, ganadería, primeros auxilios, salud y nutrición, enfermedades contagiosas, sexo y matrimonio, cuentos para niños, cómo tocar la guitarra, derechos cívicos, etcétera.

En la práctica es común que la colección de libros de la biblioteca se localice en una o más casas de la comunidad a fin de propiciar su mejor utilización por parte de los campesinos.

4. El periódico *El Campesino*. Este medio incluye el contenido normal de un periódico, pero con énfasis en temas que sean importantes para el desarrollo rural y para los intereses del hombre del campo.
5. La correspondencia. Son cartas que los campesinos envían a las oficinas centrales de ACPO y en las cuales expresan sus problemas y necesidades y solicitan ayuda. En 1978 se recibieron y fueron contestadas más de 28.000 cartas de este tipo.
6. Los cursos de extensión. Son cursos de corta duración, a veces de un día, desarrollados por un agente interpersonal de ACPO. Tratan problemas y necesidades locales específicos (cooperativas, utilización correcta del crédito, etcétera).
7. Las campañas. Son estrategias para estimular y ayudar a los campesinos a hacer aplicaciones prácticas de aquello que han aprendido. Estas campañas se anuncian en *El Campesino*, se promueven en la radio Sutatenza y son manejadas localmente por agentes interpersonales de ACPO. Entre esas campañas se cuentan las de nutrición, vivienda, suelo y recreación.
8. Los agentes interpersonales. Son las personas que ayudan a guiar el proceso de aprendizaje de los participantes en los ambientes educativos no formales y en otras actividades del programa; desempeñan un papel fundamental dentro del sistema ACPO.
9. Los profesores como agentes educativos. Los profesores que trabajan en el centro de ACPO son los encargados de planear, desarrollar y producir los programas educativos, con sus respectivas metodologías. Escriben también artículos para el periódico *El Campesino* y preparan instrumentos para evaluar los progresos de los estudiantes de las escuelas radiofónicas.
10. Agentes educativos de campo. Son las personas (más de 20.000 en 1979) que trabajan de diversas formas en la enseñanza, casa por casa, con los campesinos. Están divididos en a) auxiliares inmediatos o monitores de las escuelas radiofónicas; b) dirigentes campesinos; c) líderes (campesinos adultos que han terminado dos



ciclos de cursos en uno de los institutos campesinos de ACPO); d) supervisores (personas que tienen responsabilidades de administración y supervisión en relación con los demás trabajadores de campo, en el nivel regional).

11. Los educadores. Este grupo está constituido por los profesores de los Institutos Campesinos de ACPO en los cuales se entrena a los dirigentes, a los líderes y a los supervisores. Comprende los siguientes tipos de profesionales:
  - a) trabajadores del servicio de extensión agrícola;
  - b) enfermeras y expertos en salud;
  - c) normalistas, y
  - d) campesinos que han terminado los cursos para líderes y dirigentes y han tenido una buena experiencia de campo.

Dentro del objetivo general que enunciamos al comienzo de este resumen, ACPO tiene los siguientes objetivos más específicos:

1. El desarrollo de las habilidades básicas de alfabetización: lectura, escritura y números.
2. El mejoramiento de la salud individual y familiar por medio de prácticas higiénicas, nutrición adecuada y utilización de los servicios médicos disponibles.
3. El aumento de producción en la finca por medio de conocimientos y prácticas de técnicas modernas de agricultura, cuidado de los animales y conservación del suelo.
4. La adquisición de conocimientos y habilidades económicas y su aplicación a la administración de empresas familiares, agrícolas y otros.
5. El mejoramiento o construcción de viviendas que satisfagan las normas mínimas de salud, higiene y bienestar, y que contribuyan a una mejor relación familiar.

6. La utilización del tiempo libre para la participación en deportes u otras actividades de recreo.
7. El desarrollo de una capacidad crítica y analítica para confrontar problemas y necesidades personales y sociales, y para tomar decisiones relativas a los mismos.
8. Vivir de acuerdo con principios éticos y religiosos y servir como modelo para otros.
9. La participación en actividades destinadas a la continuación de la educación y el automejoramiento.
10. El desarrollo de estabilidad, entendimiento y mejoramiento de las relaciones humanas dentro de la familia.
11. El reconocimiento de las responsabilidades de la paternidad y la práctica de la procreación responsable.
12. La participación en actividades de la comunidad, asociación con grupos organizados, y cooperación para el mejoramiento de la comunidad.
13. La adquisición de conocimientos acerca de los servicios de otras organizaciones de la comunidad que se interesen en el mejoramiento individual y comunitario, y la utilización de los mismos.
14. El desarrollo de una conciencia política que se centre en la protección de los derechos humanos y en la importancia de que todas las personas asumen sus responsabilidades en la sociedad.

Frente a estos objetivos del programa, la evaluación se dirigió a obtener respuesta a las siguientes preguntas u objetivos de evaluación:

1. ¿En qué medida han sido efectivos los programas de ACPO en mejorar el nivel de conocimientos en salud, actividades agropecuarias y educación básica de las familias campesinas?

2. ¿En qué medida y hasta qué punto fueron capaces las familias campesinas de llevar a prácticas concretas los conocimientos adquiridos en salud, agropecuarios y educación básica?
3. ¿Resultó el aprendizaje en un cambio de actitudes del campesino hacia la salud, la actividad agropecuaria y la educación básica?
4. ¿Tuvieron estos cambios, en conocimientos, prácticas y actitudes, consecuencias positivas o negativas en el nivel de vida del campesino?
5. ¿En qué medida y en qué condiciones han estado las familias campesinas utilizando los medios de ACPO?
6. ¿Qué elemento o combinación de elementos ha tenido mayor repercusión en los resultados observados?
7. ¿En qué medida podría atribuirse a los programas de ACPO los resultados observados? ¿Cuál ha sido la contribución total y la contribución relativa de las variables de proceso, en lo que toca a explicar las diferencias en conocimientos, prácticas y actitudes entre las familias campesinas?
8. ¿En qué medida podía atribuirse a las condiciones en que vivían las familias, los resultados observados? ¿Cuál ha sido la contribución total y la contribución relativa de las variables de contexto en lo que atañe a explicar esas diferencias?

Con el fin de obtener la información para cumplir estos objetivos de evaluación se tomó una muestra aleatoria de 600 familias campesinas de 60 veredas (aldeas) de 30 municipios, de tal modo que la misma fuera representativa tanto de 22 de los 27 departamentos de Colombia como de las actividades de ACPO. En el trabajo de campo se utilizaron diversos conjuntos de instrumentos para la entrevista de las familias seleccionadas en la muestra. Los objetivos de evaluación dieron lugar a dos tipos de análisis: un análisis descriptivo de los programas de educación de ACPO, sus objetivos, las metodologías utilizadas y las características de cada uno de los componentes, con especial énfasis en la descripción de la actividad radial, el material impreso (cartilla, periódico, libros) y los agentes interpersonales. Se incluyó, además, una descripción de las regiones geoculturales en las cuales se realizó el estudio. En segundo término se hizo un análisis de tipo correlacional,

mediante el uso de técnicas de correlación, regresión múltiple, correlación canónica, análisis de redundancia y regresión múltiple multivariada.

En general, los resultados obtenidos por la investigación evaluativa fueron consistentemente positivos y apoyan la conclusión de que los programas de ACPO han tenido efectos benéficos en la vida de las familias campesinas colombianas. Fue posible conocer la contribución de otras fuentes de aprendizaje en el sector rural colombiano y examinar los efectos residuales atribuibles exclusivamente a ACPO (Informe final).

En el área de salud hay una interdependencia entre los procesos de ACPO y otros servicios de salud de la comunidad. Cada uno contribuye en proporciones aproximadamente iguales a explicar la varianza en las prácticas, los conocimientos, las actitudes y los efectos generales en dicha área.

En el área de agropecuarios se detectó una relación significativa entre la participación en los programas de ACPO y una positiva conducta agropecuaria. En esta área de variables de contexto, otras fuentes de apoyo a la actividad agropecuaria no contribuyeron significativamente en los efectos observados. En educación básica, en relación con los conocimientos y destrezas en lectura, escritura y matemáticas, se encontró que las familias que tenían un alto grado de participación en los programas ACPO tendían a mostrar un desempeño relativamente alto, y que esta relación persistía cuando los efectos de otras variables educativas del contexto eran controladas. Un resultado paradójico fue la baja relación entre las actitudes acerca de la educación y las variables de proceso de ACPO, y las variables de contexto. En el estudio se descubrió que aunque buena parte de quienes respondieron la encuesta habían recibido su educación básica en la escuela formal, la mayoría de los usuarios de ACPO tendían a obtener puntuaciones altas en la prueba utilizada para medir el desempeño de destrezas. Esto sugiere que aun en los casos en que ACPO no sea la fuente primaria para el aprendizaje de las destrezas en educación básica, sí desempeña un papel fundamental en el mejoramiento y mantenimiento de esas destrezas.

Se detectó que los cinco componentes de ACPO estaban altamente interrelacionados, de tal forma que fue imposible concluir que los componentes contribuían diferencialmente en los efectos observados. Un alto uso de los componentes de ACPO en conjunto, como

se refleja en el índice de utilización de las familias, fue más importante para explicar los resultados que una combinación particular de componentes. Sin embargo, algunos juicios emitidos por las familias acerca de la relativa importancia de los componentes sugieren que los agentes interpersonales son el componente clave. Aunque esta afirmación no es confirmada por el análisis estadístico, es consistente con el juicio del equipo del proyecto y los entrevistadores de campo (Robert M. Morgan, Liliana Muhlman y Paul H. Mahoner. *Evaluación de un sistema de educación no formal a nivel nacional, Acción Cultural Popular*. Bogotá, 1980, págs. 10 y 11).

## **Diseños de evaluación II**

En páginas anteriores nos referimos a los distintos procedimientos que se utilizan en la verificación de la efectividad de un programa y proporcionamos algunos ejemplos. Desde luego que al preparar el diseño metodológico sólo hacemos una descripción general del procedimiento elegido, el cual se desarrolla en el transcurso de la investigación.

Algo similar ocurre con los otros componentes del mismo en el sentido de que durante la etapa preparatoria basta con referirse a ellos de forma general. Así lo haremos también ahora, con ejemplificaciones para aclarar los aspectos tratados.

## **UNIVERSO DE ESTUDIO Y UNIDADES DE ANÁLISIS**

En la investigación evaluativa, el universo de estudio puede o no coincidir con el universo de usuarios. El último caso ocurre cuando el o los focos de evaluación caen en solo algunos subgrupos de los usuarios del programa. Desde luego, esta delimitación queda fijada en general por los objetivos de evaluación, pero es necesario tener suficiente claridad en este punto, ya que de él dependen otras operaciones técnicas posteriores.

La búsqueda de la respuesta a la pregunta básica de la evaluación: ¿se produjeron o no los cambios, modificaciones o efectos anunciados en los objetivos del programa? lleva, a su vez, a inquirir en qué niveles se tratará de encontrar las confirmaciones del caso, pues como lo muestran objetivos de diversos o de un mismo programa, la orientación de las actividades instrumentales dirige algunas veces a personas consideradas individualmente, y otras veces a grupos naturales (familias, cursos de una escuela, etc.), a categorías de

personas (jóvenes, personas de bajos recursos económicos, etc.), a regiones geográficas o a varios de esos tipos de referentes. La decisión para elegir las unidades de análisis debe estar de acuerdo con preguntas como éstas: ¿a quiénes o a qué me referiré cuando termine la evaluación?, ¿a quién imputaré los resultados: a personas consideradas individualmente o a grupos de sujetos tomados justamente como grupos? Es fácil observar que la información que se requiere para uno y otro nivel de análisis es diferente, como también lo son las técnicas de recolección y tratamiento de los datos obtenidos en la evaluación.

Un par de ejemplos nos ayudará a comprender mejor la importancia que tiene la adecuada elección de la unidad de análisis o el uso de más de una de tales unidades. Veamos el primer caso: en la evaluación del Programa de Desarrollo Rural de México se recomienda que en investigaciones futuras no se tome a la empresa agrícola como la unidad fundamental de la agricultura campesina, sino la unidad familiar:

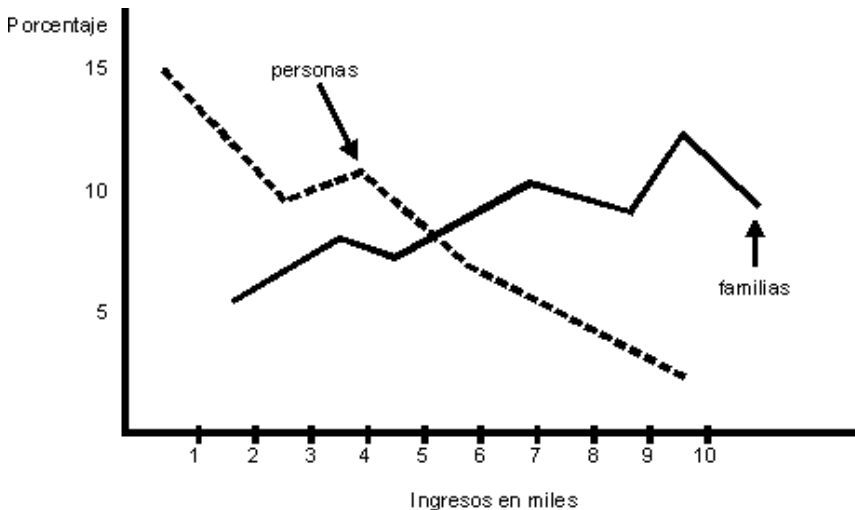
Sólo podemos tener la esperanza de decir si el programa de Desarrollo Rural ha tenido o no efectos de justicia y distribución de ingreso para el pequeño productor agrícola, si esta unidad pequeña, pero compleja y dotada de un delicado equilibrio socioeconómico, que es la unidad familiar, se constituye en la unidad de interés analítico en la evaluación de proyectos (Cernea, op. cit. pág. 113).

El segundo ejemplo se refiere a los distintos tipos de información que pueden proporcionar dos unidades de análisis consideradas separadamente; también nos hace ver los posibles errores que podrían cometerse al hacer referencias de un nivel a otro (errores que caen en la llamada falacia ecológica). El ejemplo es éste: si tomamos información de los ingresos individuales de las personas y de la mediana de los ingresos de las familias en una misma comunidad y si hacemos la representación gráfica de tal modo que en el eje vertical coloquemos los porcentajes de casos en cada categoría de ingresos (individuales y colectivos), podemos obtener una figura como la siguiente (fig. 6.1):

Como se advierte, los resultados son muy diferentes, pero, por lo mismo, usados conjuntamente pueden proporcionar informaciones muy útiles en la evaluación.

En suma: la elección de las unidades de análisis y la recolección de la información pertinente deben permitir formular proposiciones de los siguientes tipos: a) sobre la totalidad

del programa: «el ingreso promedio de la población subió en un X por ciento»; b) sobre los ambientes de interacción de la población: «el ambiente social de la escuela X es de tipo democrático»; y c) sobre las personas consideradas individualmente, con el fin de establecer relaciones como: «la mayoría de las personas de más edad tuvieron una alta participación en actividades comunitarias».



**Figura 6.1.** Ingresos individuales y familiares

## DISEÑOS MUESTRALES

Una vez que se ha tomado una decisión acerca de la o de las unidades de análisis hacia las cuales se referirá empíricamente la evaluación, deben tomarse otras nuevas si existen varias de tales unidades: tomar una, tomarlas todas o seleccionar algunas del total. Esta última decisión envuelve dos posibilidades principales: a) utilizar un criterio probabilístico; b) seleccionar las unidades a partir de algunos criterios establecidos por el evaluador. La opción a) equivale a decir que se utilizará un diseño muestral probabilístico; la opción b), que se utilizará un diseño muestral intencional o por cuotas, no probabilístico.

Como sabemos, los diseños probabilísticos dan una probabilidad conocida, distinta de cero, a cada una de las personas del universo de estudio, de entrar en la muestra. En algunos de ellos, la probabilidad es igual para cada unidad. Estos diseños son de uso

obligado en las investigaciones experimentales. Sus principales ventajas son: 1. representatividad de la muestra respecto del universo; 2. posibilidad de calcular por anticipado la magnitud y la probabilidad de que se dé el error de muestreo; 3. posibilidad de generalizar el universo los resultados obtenidos en el análisis de la muestra (para un tratamiento detallado de los diversos diseños probabilísticos, consúltese a G. Briones, *op. cit.*, cap. V.).

Las muestras intencionales se seleccionan según ciertos criterios definidos por el evaluador con la finalidad principal de ubicar los focos de evaluación en grupos a los cuales le asigna especial importancia. Como el procedimiento empleado no es probabilístico en su conformación, no es posible obtener las propiedades de los diseños probabilísticos que señalamos más arriba. Sin embargo, en algunas situaciones es posible hacer inferencias lógicas (pero no inferencias estadísticas, en sentido estricto) hacia poblaciones que tienen características similares a las de la muestra generadora. Pero su principal característica consiste en posibilitar los estudios en profundidad de componentes o procesos del programa.

A continuación se presentan los tipos de muestras intencionales de mayor uso en la investigación evaluativa con diseños básicos cuasiexperimentales y no experimentales.

1. Selección intencional de casos extremos. Quizás sea posible hacer una clasificación aproximada de las unidades que componen el programa en dos categorías, por ejemplo, escuelas que funcionan bien y escuelas que funcionan mal. Si así fuera, podrían tomarse unas pocas escuelas de ambas categorías para estudiar otras características diversas relacionadas con su nivel de funcionamiento.
2. Selección de algunos casos típicos. Si existe una mayoría considerable que puede ser considerada como típica o modelo, entonces podría bastar la selección de algunos casos para hacer estudios en profundidad de los aspectos que interesan a la evaluación, pues esos estudios representarían a la mayoría de casos similares.
3. Selección de casos en diferentes ubicaciones geográficas o sociales. Este tipo queda representado por la selección de casos que cubren las diversas zonas de un espacio geográfico (escuelas urbanas - escuelas rurales; zonas centro norte-sur-oriente-occidente) o de espacios sociales pertinentes (hogares populares, de



clase media, de la burguesía). Los diversos subgrupos de la muestra pueden ser proporcionales a sus respectivos tamaños en el universo.

4. Selección de casos críticos. La selección recae aquí en casos que presenten algunas características críticas: alta deserción, altos costos, baja eficiencia administrativa, etc. El análisis de situaciones como éstas puede llevar a hacer recomendaciones a todos los grupos con el fin de introducir en ellos modificaciones que se consideren útiles o necesarias.
5. Diseño mixto. En algunas oportunidades es posible combinar una muestra intencional con una submuestra probabilística. Por ejemplo, si se toma cierto número de escuelas en áreas rurales y otro número en áreas urbanas, entonces, en una segunda etapa es posible tomar una submuestra probabilística de estudiantes o profesores en cada una de las escuelas seleccionadas intencionalmente. El diseño probabilístico puede ser del tipo aleatorio simple, sistemático, estratificado, etcétera.

Cualquiera sea el diseño que se elija, en el diseño metodológico de la evaluación se hará una descripción general de sus principales características (tamaño, selección probabilística, selección intencional, etcétera).

## Indicadores

Es frecuente que las variables o características que aparecen en los objetivos de evaluación no ofrezcan mayores dificultades para su descripción o su medición directa (por ejemplo logros en el aprendizaje, elevación del ingreso familiar, etc.). También es frecuente, sin embargo, que otras variables requieran definiciones para comprender su significado; de indicadores que permitan su descripción y, cuando sea posible o necesario, su medición. En este último caso, es indispensable utilizar técnicas ya preparadas o que el evaluador debe desarrollar por su cuenta (pruebas, escalas, etcétera).

En ocasiones una propiedad incluida en los objetivos de evaluación queda bien representada por uno o dos indicadores (por ejemplo, estatus socioeconómico, por la educación y los ingresos); otras veces la solución es más difícil tanto por el número de indicadores que es necesario utilizar como por la forma en la cual deberían ser combinados para obtener una representación válida y confiable. La tarea suele ser aun más complicada cuando la situación

que ha de evaluarse es de naturaleza estrictamente cualitativa («evaluar el aumento de la capacidad crítica»), o bien se trata de evaluar la calidad de algunos de los elementos, procesos o resultados del programa. En tales casos se recurre a procedimientos cualitativos como entrevistas en profundidad, evaluaciones grupales, etc., o se utilizan características, que implican por ejemplo, variedad de recursos instrumentales empleados, contenidos de aprendizaje, distribución homogénea de los logros obtenidos, etcétera.

Los indicadores pueden referirse a variables individuales (actitudes, nivel socioeconómico, etc.) o a variables colectivas, es decir, de los grupos (homogeneidad, igualdad de oportunidades de clases sociales, eficiencia del programa, etc.). Pueden ser, por otro lado, indicadores objetivos o bien indicadores subjetivos constituidos por juicios, opiniones o estimaciones hechas por los evaluadores o las personas sometidas al proceso de evaluación.

En lo que se refiere a programas sociales (salud, educación, vivienda, trabajo, desarrollo de la comunidad, etc.), hay listas de indicadores aceptados internacionalmente, y que son fáciles de conseguir. Sin perjuicio de su uso, de acuerdo con los objetivos de una evaluación determinada, la especificidad de algunos programas o las necesidades mismas de la evaluación pueden requerir el uso de indicadores especiales propuestos por los evaluadores. Un ejemplo que cae en esta solución lo encontraremos en la evaluación del Servicio Nacional de Aprendizaje, SENA, realizada en Colombia en 1982. Al referirse a la evaluación del objetivo «contribuir al crecimiento dinámico del país mediante la preparación de recursos humanos», se propone hacerlo de la forma siguiente:

Al nivel analítico de la sociedad en su conjunto, se tomará como criterio de evaluación la «rentabilidad social» de cada programa de adiestramiento, o valor actualizado del aumento en el producto nacional por unidad de costo, o sea el aumento directo o el aumento total. Este indicador, que sintetiza el aporte del curso a la función crecimiento, ha sido elegido en atención a sus consideraciones principales. (Sena. *Sistema de Planeación de Recursos Humanos*. Bogotá, 1982, págs 140 a 141).

También en esta evaluación se utilizan diversos indicadores subjetivos para evaluar otras características del sistema de aprendizaje, como el siguiente que ilustra la clasificación que acabamos de presentar: «Percepción sobre la utilidad ocupacional del curso para el egresado o desertor» (pág. 145).

Otro ejemplo de un indicador complejo para evaluar la equidad social es el utilizado en una investigación acerca del funcionamiento de la educación superior en Chile (1981). En este caso, el indicador o índice se definió de la siguiente manera:

$$\frac{\text{Porcentaje de alumnos matriculados en la educación superior cuyos padres tienen las Ocupaciones X}}{\text{Porcentaje de personas en la fuerza de trabajo con ocupaciones X}}$$

La X puede representar «obreros» y luego «profesionales», por ejemplo. La comparación de los dos índices permite evaluar la equidad social. (Guillermo Briones. *La distribución de la educación en el modelo de economía neoliberal: 1974-1982*. PIIE. Santiago de Chile, 1983.)

En la evaluación de efectos de la educación funcional de adultos es interesante destacar algunos indicadores que utilizan diversos criterios o dimensiones, de los cuales proporcionamos otros ejemplos:

Efecto de la alfabetización: querer informarse y comunicar.

Criterio 1: Utilización de los medios de comunicación de masas.

4.1.1. Número de periódicos y revistas –diarios, semanarios, mensuales– comprados o leídos, por unidad estadística (individuo, familia, comunidad) y por unidad de tiempo.

4.1.2. Número de receptores de radio o televisión poseídos o escuchados.

4.1.3. Número de escuchas, por unidad de tiempo y por tipo de emisión (informaciones, emisiones profesionales, música, etcétera).

4.1.4. Número de magnetófonos y número de casetes poseídos o escuchados por unidad de tiempo de tipo de contenido.

Criterio 2: Utilización de otras fuentes informativas y de las fuentes de ayuda más adecuadas.

Por unidad de tiempo y por unidad estadística.

4.2.1. Número de consejos pedidos a las organizaciones profesionales (sindicales, etc.), cívicas, administrativas (ayuntamientos, concejos, etcétera).

4.2.2. Número de consejos pedidos a organismos de ayuda social (seguridad social, mutualidades, compañías de seguro, etcétera).

4.2.6. Número de viajes emprendidos por duración y naturaleza: razones familiares, emigración estacional, búsqueda de trabajo, razones religiosas, etcétera.

Criterio 3: Utilización social e informativa de la lectura y de la escritura, por unidad estadística y por unidad de tiempo.

4.3.1. Número de cartas recibidas o enviadas, por tipo de correspondencia: familiar, administrativa, informativa, diversa.

4.3.2. Número de agendas, carnés, almanaques, anuarios, etc., utilizados (Couvert, op. cit., págs. 16 a 161. En el texto se dan otros indicadores en la línea de los tomados como ejemplo).

Es importante destacar que en algunas evaluaciones es necesario tener indicadores de conductas reales o prácticas, de conocimientos y de actitudes. Así podemos comprobarlo en la evaluación del programa de Acción Cultural Popular, ACPO, que presentamos páginas atrás. Debido a que dan una idea de los elementos, presentaremos algunos de los indicadores seleccionados en el área de salud:

## 1. Prácticas

### A. Cuidados personales

#### 1. Nutrición

- 1.1. Frecuencia de consumo de vegetales
- 1.2. Frecuencia de consumo de leche o productos lácteos por los niños
- 1.3. Frecuencia de consumo de frutas

## 2. La madre y el niño

- 2.1. Dieta de la mujer embarazada
- 2.2. Control médico de la mujer embarazada

## 3. Higiene

- 3.1. Frecuencia con que se bañan los niños

## B. Cuidado del ambiente

- 1. Huerto casero
  - 1.1. Número de vegetales que se cultivan
- 2. Agua
  - 2.1. Utilización del filtro casero

## II. Conocimientos

- 1. Ventajas de la huerta casera
  - 1.1. Razones económicas: como ingreso
  - 1.2. Razones económicas: como ahorro
  - 1.3. Razones de nutrición

2. Dieta familiar
  - 2.1. Inclusión de vegetales
  - 2.2. Inclusión de frutas
3. Primeros auxilios
4. Purificación del agua

### III. Actitudes

1. Consumo de vegetales
2. Cultivo de la huerta casera
3. Control médico de la mujer embarazada

(Robert M. Morgan et al., op. cit., págs. 197 a 201).

Una vez que se definen o eligen los indicadores, se traducen en preguntas o en indicaciones para la observación o su comprobación mediante entrevistas o por su uso en otros procedimientos para la recolección de información.

## Recolección de información e instrumentos

En el diseño metodológico de una evaluación deben indicarse los procedimientos que se utilizarán para la recolección de informaciones y el o los instrumentos que se destinan a tal fin. En lo que toca a procedimientos, podrían utilizarse observaciones, entrevistas de todo tipo, discusiones grupales, etc. Los instrumentos, según el caso, pueden ser protocolos de observación (estructurados o no), guías de entrevistas, cuestionarios, registros de discusiones, etcétera.

En algunos casos existen instrumentos contruidos y validados por otros investigadores. Un caso así vimos en la evaluación del programa de estimulación temprana con base en la familia, en la cual los investigador es utilizaron la prueba de Denver, de tipo cualitativo

con los ítems para evaluar el desarrollo psicológico de los niños de cero meses a seis años de edad. En otros casos, el investigador debe construir el o los instrumentos adecuados, sea para describir o medir conductas, conocimientos y actitudes, tarea en la cual suele ayudarse con las indicaciones pertinentes que se dan en textos de técnicas de investigación social.

## **Trabajo de campo**

Aquí, por lo general, se indica quiénes recogerán la información en el terreno, cómo serán preparadas esas personas, cuál será la distribución del trabajo, de qué forma se llevará a cabo la supervisión, etcétera.

## **Análisis de los datos**

En el diseño deben mencionarse los tipos de análisis a los cuales serán sometidos los datos y los procedimientos que se utilizarán, sean de tipo cualitativo o cuantitativo. Por ejemplo, en el primer caso se mencionan tipologías cualitativas, conexiones entre procesos y resultados, etc. En el segundo, el evaluador podrá referirse a análisis relacional y enumerar técnicas como las de regresión, análisis de senderos y de redundancia, entre otras. O en diseños experimentales, técnicas como análisis de la varianza, covarianza, etcétera.

## **El informe de evaluación**

El informe de evaluación describe la investigación realizada y en él se puntualizan con mayores detalles los componentes del diseño, como también las dificultades que se encontraron en su aplicación. Pero el núcleo del informe está constituido por los resultados obtenidos, sea en la evaluación del contexto del programa, de su estructura, de su funcionamiento y de sus logros y efectos; con el análisis de los factores que pudieron facilitar o dificultar la marcha del programa y la consecución de sus objetivos.

El informe de evaluación culmina con sugerencias o recomendaciones basadas en los resultados obtenidos.

La aceptación de las recomendaciones puede verse afectada por diversos factores entre los cuales se destacan problemas como los citados por dos autores preocupados de esta situación:

- Diferencias entre los roles y estilos del personal del programa y los evaluadores.
- Estudios que son insuficientemente rigurosos para ser convincentes.
- Demasiada sofisticación metodológica y poca preocupación con los aspectos concretos del programa.
- Resultados que no están disponibles a tiempo para la toma de decisiones.
- Respuestas a preguntas en las cuales los administradores no tienen interés.
- Los asuntos de la marcha y financiamiento son a menudo más importantes que los resultados de la evaluación.
- Los resultados de la evaluación se comunican al personal del programa, de manera que ellos no pueden ser usados (Ray C. Oman y Stephen R. Clitwood. «Management evaluation studies: Factors affecting the recommendations». *Evolution Review*. vol. 8, núm 3, junio de 1984, pág. 286).

De manera típica, el informe de evaluación contiene secciones como las siguientes:

1. Descripción del programa (estructura, objetivos, ubicación, etcétera) .
2. Contexto socioeconómico
3. La evaluación:
  - a) origen
  - b) propósitos
  - c) objetivos de evaluación



- d) metodología
  - e) problemas en la marcha del programa
4. Resultados
  5. Recomendaciones

## **EL DISEÑO EN MODELOS GLOBALES, EVALUACIÓN PARTICIPATIVA**

El diseño de evaluación de programas orientado por algunos de los modelos globales que presentamos en el capítulo 3 o por otros modelos similares tiene importantes diferencias con el que hemos desarrollado en las páginas anteriores. Para comenzar, debemos decir que más que una descripción específica de las diversas fases y tareas que deberían desarrollarse en la investigación evaluativa, el diseño de los modelos globales consiste en indicaciones generales acerca de cómo conducir el estudio, ya que varias de las acciones particulares se definirán sobre la marcha.

Por otro lado, como características que se dan en muchos de los diseños figuran la participación de los usuarios del programa en todas o en varias etapas de la evaluación, y la utilización de indicadores subjetivos para la valoración del desarrollo del programa y de los resultados obtenidos (autoevaluación, evaluación grupal, etc.), todo ello mediante el uso de técnicas cualitativas, preferentemente para la recolección y el análisis de la información. Algunos de estos modelos ponen en el centro de la evaluación la descripción de los procesos que se desarrollan en el programa, antes que la determinación de resultados.

Sin perjuicio de diseños mixtos que incorporan componentes de los diseños analíticos, el global contiene orientaciones en los siguientes aspectos:

1. Consulta a los usuarios del programa para establecer qué les interesa evaluar.
2. «Negociación» entre los diferentes puntos de vista aparecidos en la consulta para llegar a un acuerdo. Pueden incluirse aquí los intereses de los administradores del programa.

3. Preparación de un esquema para el trabajo de campo.
4. Trabajo de campo:
  - a) Observación participante del personal encargado de la evaluación.
  - b) Reuniones grupales para evaluar el funcionamiento del programa y resultados mediante discusiones y juicios valorativos.
  - c) Entrevistas en profundidad a informantes seleccionados para profundizar temas importantes detectados en a) y b).
  - d) Registro de experiencias personales de algunos usuarios, e historias de vida.
5. Análisis de la información recogida:
  - a) Imágenes verbales sobre el programa y sus componentes: objetivos, necesidades, relaciones interpersonales, etcétera.
  - b) Determinación de la significación cultural de las expresiones verbales de los usuarios.
  - c) Conexiones entre procesos.
  - d) Tipología de comportamientos.
6. Determinación de la efectividad del programa. Esquema de "triangulación".

En el hecho, tal procedimiento consta de las siguientes tres fuentes de validación:

1. Confrontación de información cualitativa y cuantitativa (triangulación metodológica);
2. Comparación de diversas fuentes de información cualitativa (triangulación de fuentes de información); y

### 3. Comparación de las informaciones de diversos observadores (triangulación de investigadores).

Un ejemplo de triangulación lo proporciona la evaluación del Programa padres e hijos al cual ya nos referimos en este trabajo y que, de la parte pertinente, transcribimos textualmente:

Como hemos mencionado en el capítulo anterior, la triangulación se realizó con el fin de comprobar, a través de procedimientos diferentes y desde distintos ángulos, la coherencia de la idea que tenemos acerca de lo que el programa es. Se investiga si la evidencia con que contamos es consistente con el concepto que tenemos, como evaluadores, de cómo funciona, en este caso, el mecanismo del Programa padres e hijos. Se procedió, entonces, a triangular la imagen verbal (contestaciones de los usuarios). Hubo varias afirmaciones entregadas por los primeros informantes y luego corroboradas por los grupos, que se intentaron precisar a través de este proceso de triangulación. Entre éstas se destacaron, por ejemplo, el mejor rendimiento de los niños que asistían al PPH en el colegio; el mejoramiento de la nutrición de los niños; la disminución del alcoholismo entre los participantes...

Veamos cómo se hizo la triangulación de uno de estos aspectos. Tomaremos, a modo de ejemplo, el efecto del programa en los niños en edad preescolar inscritos en el PPH. Se realizaron distintos estudios que «apuntaban» al problema desde perspectivas distintas. Por una parte, se entrevistaron tres profesores de escuelas básicas que tenían niños que habían participado en el PPH y niños que no. Se les pidió a los profesores que clasificaron como excelente, bueno, aceptable o deficiente a los niños de primero básico según su habilidad para leer. Los profesores no estaban en conocimiento de los niños que habían participado en el programa. El resultado fue el siguiente:

	Con PPH			Sin PPH			Total
Excelente	5	niños	19%	2	niños	4%	7 a 9%
Bueno	15	niños	52%	17	niños	35%	32 a 42%
Aceptable	9	niños	31%	24	niños	49%	33 a 41%
Deficiente	-.-		0%	6	niños	12%	12 a 6%
Total	29	niños	100%	49	niños	100%	78 a 100%

El cuadro muestra que la clasificación destaca favorablemente la habilidad de lectura de los niños con PPH. Lo confirma la distribución porcentual de los niños con y sin PPH en cada uno de los rubros.

Se aplicó también el test de la figura humana (Goodenough y Harris) a un grupo de 26 niños con PPH y a un grupo de 24 sin PPH. La muestra fue elegida al azar. La evaluación de los resultados del test fue encomendada a un experto en psicología del desarrollo infantil que no tenía conocimiento previo acerca del programa. El resultado de la evaluación de los dibujos aparece en el siguiente cuadro:

#### Clasificación de dibujos

*Mejor dibujo: niño del grupo de control*

	Grupo control	Grupo PPH	Total
Superior	3	2	5
Bueno	3 25%	9 42%	12
Promedio	6	10	16
Pobre	9	3	12
Inferior	3	2	5
	24 niños	26 niños	50 niños

*Peor dibujo: niño del grupo de control*

La evidencia que muestra el cuadro complementó los resultados de los otros estudios que se fueron realizando en la triangulación.

Es importante destacar que en la triangulación se utilizan técnicas de distinta índole. Algunas de carácter etnográfico como las entrevistas y las observaciones; otras de carácter cuantitativo. Todas cumplieron la función de investigar si las evidencias con que se contaba eran consistentes con la imagen que se tenía de lo que el programa era (Horacio Walker. *Una evaluación para los proyectos de educación popular*. Santiago de Chile: Cide, 1982, págs. 52 a 57).

De manera característica, el informe de la evaluación no formal se redacta en términos sencillos, con amplio uso de opiniones y relatos en las propias palabras de los participantes. El informe puede o no terminar con recomendaciones para contribuir al mejoramiento del programa. Convendría también decir que en algunas oportunidades se acompaña de gráficos, dibujos de los usuarios, diapositivas y grabaciones (diapofilm) de diversas instancias de desarrollo del programa, etcétera. (Para algunos otros detalles de la evaluación de programas globales se puede consultar a Thomas D. Cook and Charles S. Reíhardt (eds.). *Qualitative and quantitative methods in evaluation research*. Beverly Hills: Sage Publications, s./f.)

No todas las evaluaciones de programas no formales trabajan con el enfoque presentado más arriba, que pretende lograr una alta participación de los usuarios. Con frecuencia se encuentran ejemplos en los cuales esa función es realizada principalmente por el personal encargado del programa, en especial por el equipo técnico de campo. Así sucede en proyectos de desarrollo rural, área en la cual debemos destacar el enfoque de Oakley con las características que este investigador señala. En resumen –y de forma muy apegada a su propia descripción– el esquema utilizado comienza por identificar los criterios e indicadores de acuerdo con los cuales se puede apreciar el progreso del programa. Para ello se elabora, conjuntamente con los técnicos o monitores, una lista de características (criterios) que los usuarios poseían antes de comenzar las actividades innovadoras. Luego se les pide caracterizar al mismo grupo de usuarios después de cierto tiempo de la puesta en práctica de tales acciones. Las características son las siguientes:

a) Características (criterios de los usuarios antes del desarrollo del programa):

- Individualismo, que se ve reflejado en el predominio de la propiedad privada, en el poco interés en proyectos comunitarios, así como también en la escasa participación en las decisiones.
- Falta de análisis crítico de su situación: incapacidad para identificar las causas de problemas estructurales y buscarles solución.
- Dependencia económica, social y política del patrón y explotación por parte del mismo.

- Falta de confianza en su propia capacidad para cambiar la situación.
- Ausencia de organizaciones que representen los intereses del grupo.
- Falta de cooperación.
- Ignorancia, sospecha y aislamiento. La gente tiene temor de conversar y discutir y verse involucrada.

b) Características del grupo después de un período de desarrollo del programa:

- Cohesión interna.
- Sentido de la solidaridad.
- Conciencia crítica. Facultad crítica.
- Participación activa y crítica.
- Menor dependencia. Aumento de la autoconfianza.
- Autogestión, autosuficiencia y autonomía del grupo.
- Recursos colectivos. Reservas de capital.
- Capacidad para la administración del programa.
- Demostración del poder. Responsabilidad colectiva.
- Articulación con otras instituciones controladas por campesinos.
- Interés en la creación de grupos similares.
- Capacidad para tratar con funcionarios oficiales.

En situaciones concretas, un grupo en evaluación puede no presentar todas las características indicadas; más bien sirven de ejemplos de las que pueden darse en la realidad. El trabajo, en todo caso, continúa con la búsqueda de indicadores para tales características. Así, en relación con los criterios del ejemplo, se señalan indicadores como los siguientes, que el personal técnico de programas de desarrollo rural cree pueden utilizarse para pronunciarse acerca de los cambios que se han producido o se están produciendo en los usuarios:

- Actividades realizadas por el grupo: obras de construcción, organizaciones, etcétera.
- Acciones específicas del grupo con la ayuda del monitor o sin ella. Por ejemplo: acciones para obtener beneficios de seguridad social o para demandar el cumplimiento de la legislación existente.
- Cambios en la conducta del grupo: tipo de reuniones, uso del lenguaje, naturaleza de la participación, etcétera.
- Naturaleza de la intervención: cambios en las relaciones del monitor con el grupo como resultado de que este último ha asumido la responsabilidad de su propio desarrollo.
- Relaciones con otros grupos: naturaleza de las relaciones con otros grupos y acciones conjuntas, empresas y organizaciones comunes, etcétera.

En términos operativos se sugiere que el personal del proyecto encargado de la evaluación o de la recolección de información para el trabajo que otro grupo ha de realizar registre continuamente sus observaciones o reúna los datos pertinentes. Para ello pueden emplearse los siguientes métodos:

1. Encuesta para el perfil del grupo, es decir, para establecer la situación cuando comienza el programa.
2. Recolección de datos cuantificables. Tales datos deben incluir la cantidad de reuniones, de participantes, el número involucrado en actividades particulares y las cifras básicas relativas a cualquier actividad económica que se lleve a cabo.

3. Descripciones continuas y observaciones del grupo. Éstas implican llevar un registro periódico basado en las cinco áreas de indicadores. Los monitores deben escribir libremente sus apreciaciones subjetivas, sin prestar mayor atención a la estructura o forma del informe. El énfasis debe estar en el registro de los cambios que se observan.
4. Minutas de otras formas de registros de las reuniones de los monitores.

Una vez recolectada toda la información se procede a ordenarla y a interpretarla en términos de los cambios que se hayan producido en relación con las condiciones iniciales, las cuales, a su vez, ayudaron a definir los objetivos del programa. La última tarea consiste en transferir los resultados y conocimientos obtenidos a todo el personal técnico para mejorar la marcha del proyecto (Peter Oakley. «Evaluating social development: how much or how good», Ideo RRDC Bulletin, marzo de 1982, 14, págs. 12 a 17).



## Los autores

El doctor *José Santiago Correa Uribe*, ha sido profesor de la Universidad de Antioquia desde el año 1979 hasta la fecha. Se ha desempeñado además, como profesor de la Universidad Pontificia Bolivariana y Universidad de Medellín y como profesor invitado de las principales universidades del país.

Ha publicado, entre otras, las siguientes obras: *Introducción a la investigación* (Medellín, Universidad de Antioquia 1981), *Currículo y calidad de la educación superior en Colombia* (Santafé de Bogotá, ICFES-COLCIENCIAS, 1987), *Metodología de la investigación científica* (Medellín, Universidad de Medellín, 1988), *La innovación en educación* (Medellín, Universidad de Antioquia-COLCIENCIAS, 1988), *Evaluación institucional aplicada a la educación* (Medellín, Editorial Sagrada Familia, 1982), *Metodología de la investigación aplicada a la gestión bancaria* (Ibagué, Universidad del Tolima, 1994).

Actualmente se desempeña como asesor de la Vicerrectoría de Docencia de la Universidad de Antioquia, en procesos de autoevaluación y acreditación.

El Doctor *Antonio Puerta Zapata*, hizo sus estudios de Filosofía y Letras, y Licenciatura en Trabajo Social en la Universidad Pontificia Bolivariana. Cursó su Maestría en Sociología de la Educación en la Universidad de Antioquia y Maestría en Trabajo Social (Área de Investigación) de la Universidad de Minnesota.

Se ha desempeñado como profesor visitante de varias universidades colombianas y de la Universidad Católica de Guayaquil y es Miembro del Consejo Directivo de la Federación Antioqueña de Organizaciones no Gubernamentales.

Ha sido asesor de: Evaluación de programas en salud Mental y prevención de la violencia para Metrosalud. Medellín Participante del Grupo de Métricas Sociales de la Federación

antioqueña de ONG y de varias investigaciones en Basf Química Colombiana y empresas Varias de Medellín y ha actuado como par Académico del Consejo Nacional de Acreditación.

Actualmente es Profesor Titular de tiempo completo de la Universidad de Antioquia. Jefe del Departamento de Trabajo Social y Asesor de la Vicerrectoría de Docencia en los Procesos de Autoevaluación y Acreditación de la misma Universidad.

Publicó el Libro sobre Evaluación y sistematización de Proyectos Sociales; ha escrito artículos en Revistas Nacionales de Trabajo Social, y ha publicado escritos sobre Acreditación.

El Doctor *Bernardo Restrepo Gómez* es profesor emérito de la Universidad de Antioquia, institución en la ha trabajado durante 36 años en la docencia y la investigación de la educación. Aunque esta universidad ha sido el centro de sus actividades, ha sido profesor visitante y conferencista en varias universidades del país. Ha actuado como consultor de la UNESCO y del Convenio Andrés Bello y como asesor de la OEA en programas educativos; fue asesor del Ministro de Educación, consejero del Programa de Estudios Científicos de la Educación de COLCIENCIAS y se desempeña actualmente como consejero del CNA, Consejo Nacional de Acreditación de Colombia. Ha realizado un buen número de investigaciones y evaluaciones de programas educativos y escrito libros y artículos en educación y pedagogía. Sus publicaciones más recientes son: "ABP, Aprendizaje Basado en Problemas, en la Formación de Profesionales de la Salud" (2000), "La Educación y el Desarrollo Sostenible en la Región de Planificación del Occidente Colombiano" (2000), "La Investigación Formativa en la Universidad: concepto, relación con el pregrado y formas de implementación" (1999), "Maestro Investigador, Escuela Investigadora e Investigación de Aula" (1999); "Calidad y Eficiencia Interna de la Educación en el Occidente Colombiano" (1998).

