

70188

3-P-83-0043

IDRC - Lib
70188

PROGRAMA INTERDISCIPLINARIO DE INVESTIGACIONES EN EDUCACION(PIIE)

Guillermo Briones

EVALUACION DE PROGRAMAS SOCIALES

Teoría y Metodología de la Investigación Evaluativa

I

Santiago, Diciembre, 1984.

ARCHIV
70188
v. 1

Guillermo Briones

EVALUACION DE PROGRAMAS SOCIALES

Teoría y Metodología de la Investigación Evaluativa



SANTIAGO, 1984.

En revisión

ARCHIV
370.001.4
B7
v.1

I N D I C E

pagina

INTRODUCCION

PRIMERA PARTE : TEORIA Y METODOLOGIA

CAPITULO 1: NATURALEZA Y PROPOSITOS DE LA EVALUACION

1

Una definición de trabajo

4

Propósitos de la evaluación

11

Resistencia a la evaluación

15

Niveles de evaluación	17
CAPITULO 2: TIPOS DE EVALUACION	20
Algunos tipos:	
- Evaluación intermedia - evaluación terminal	21
- Evaluación formativa - evaluación sumativa	22
- Evaluaciones descriptivas y explicativas	23
- Evaluación de objetivos internos - evaluación de objetivos externos	27
- Evaluación externa - evaluación interna	28
CAPITULO 3: MODELOS DE EVALUACION	31
Modelos globales:	
- Modelos subjetivistas:	33
- La evaluación "sensitiva"	35
- El modelo "iluminativo"	36
- Modelo de la evaluación focalizada	39
Otros modelos globales:	41
Modelos analíticos:	42
- El modelo contexto-insumo-proceso ^{producto} (modelo CIPP)	44
- El modelo de referentes específicos de evaluación	48
Representación esquemática del modelo	53
CAPITULO 4: VALIDACION DE UN PROGRAMA DE ACCION	55
Programa de desarrollo rural (México)	56
Programa de capacitación campesina (Chile)	57

Programa padres e hijos (Chile)	57
Programa de educación abierta y a distancia (Colombia)	58
Proyecto experimental de educación bilingüe-Puno (Perú)	58
Programa de solidaridad social por cooperación comunitaria (México)	58
Diseño de validación	63
Diseños experimentales:	66
- Diseño experimental clásico	66
Ejemplo de aplicación	68
- Hipótesis general	69
- Hipótesis específicas	69
- Diseño con grupo de control y mediciones "después"	72
Diseño cuasi - experimentales:	73
- Diseño con un grupo de control no equivalente	73
Ejemplo de aplicación	75
- Diseño "parchado" (patch up)	79
Ejemplo de aplicación	80
- Diseño de series de tiempo	84
Validez ecológica del programa	86
CAPITULO 5: DISEÑOS DE EVALUACION I	88
Objetivos de evaluación:	89
- Objetivos descriptivos	91
- Objetivos explicativos	94
Diseño metodológico	99
- Determinación e efectividad	99

- Uso del diseño experimental clásico	100
Un ejemplo de aplicación	101
- Problemas para la aplicación de diseños experimentales	104
- Diseños no experimentales: ^{caracterización} caracterización	
- Diseño con medición solo "después" con grupo de comparación	105
Un ejemplo de aplicación	106
- Diseño con mediciones "antes" y "después" ^{caracterización}	110
- Diseño con medición solo "después" ^{caracterización}	112
Ejemplos de aplicación	113
- Evaluación de servicios de salud mental	114
- Evaluación de un sistema no formal de educación a distancia	117
 CAPITULO 6: DISEÑOS DE EVALUACION	127
Universo de estudio y unidades de análisis	127
Diseños muestrales:	130
- Selección intencional de casos extremos	132
- Selección de algunos casos típicos	132
- Selección de casos en diferentes ubicaciones geográficas o sociales	132
- Selección de casos "críticos"	133
- Diseño mixto	133
Indicadores	134
Recolección de informaciones e instrumentos	139
Trabajo de campo	140
Análisis de los datos	140
El informe de evaluación	141

El diseño en modelos globales. Validación por triangulación	143
La evaluación participativa	152
La investigación acción	154

SEGUNDA PARTE : ANALISIS Y EVALUACION DE COMPONENTES

CAPITULO 7: ANALISIS Y EVALUACION DE OBJETIVOS	158
Naturaleza de los objetivos	160
Evaluación de objetivos:	163
- Identificación	163
- Claridad	166
- Adecuación contextual	166
- Adecuación al campo de dominio	167
- Vigencia de los objetivos	167
- Desplazamiento de los objetivos	168
- Conflicto de objetivos	168
- Grado de generalidad	169
- Secuencia	169
- Jerarquización	171
- Cuantificación	172
- Tiempo de logro	172
- Dimensiones psicológicas	173
- Efectos derivados	174
Resumen	175

CAPITULO 8: EVALUACION DE PROCESOS	177
Procesos de mantenimiento	178
Algunos objetos y variables de evaluación	179
Procesos de cambio e innovación	186
- El contexto de interacción	187
- Los procesos de interacción e implementación	187
- La interacción en la sala de clases	189
- Metodologías de innovación y recursos instrumentales	192
CAPITULO 9: EVALUACION DE RESULTADOS	198
Resultados previstos y no previstos	201
Determinación de resultados	203
Una tipología de efectos sociales	205
Locales	206
Efectos regionales	207
Costo - efectividad	207
Cálculo de costo:	209
- Actualización	209
- Niveles de medición	210
Cálculo de la efectividad	211
- Efectividad en subgrupos de la población	214
Costo - calidad	215
Permanencia y autonomía de los programas	219
Evaluación del rendimiento escolar :	220
- Variables y técnicas de análisis	221

- Factores de corrección para la comparación del rendimiento	224
--	-----

TERCERA PARTE : EVALUACION DE SISTEMAS EDUCATIVOS

CAPITULO 10: ANALISIS DE SISTEMAS EDUCATIVOS	231
--	-----

CAPITULO 11: EVALUACION DE LA EFICIENCIA INTERNA DE UN SISTEMA EDUCATIVO	237
--	-----

Cálculo de la retención y de la deserción de una cohorte escolar	238
--	-----

- Método de la cohorte reconstituída	238
--------------------------------------	-----

Ejemplo de aplicación	241
-----------------------	-----

- Método de la cohorte aparente	245
---------------------------------	-----

Indices de eficiencia interna	248
-------------------------------	-----

- Años - alumnos por alumnos graduados	248
--	-----

- Tasa de insumo-producto	249
---------------------------	-----

- Indice de "desperdicio escolar"	250
-----------------------------------	-----

- Costos unitarios como indicadores de eficiencia interna	251
---	-----

CAPITULO 12: EVALUACION ECONOMICA DE LA EDUCACION	253
---	-----

Evaluación del rendimiento de la educación	254
--	-----

Ejemplo de aplicación	256
-----------------------	-----

Evaluación de la rentabilidad de la educación	261
---	-----

Composición del costo educativo	263
---------------------------------	-----

Composición de los beneficios	266
-------------------------------	-----

Perfiles edad-ingresos	268
------------------------	-----

Cálculo de la rentabilidad o tasa interna de retorno	270
--	-----

- Ejemplo de cálculo	272
----------------------	-----

- Cálculo de los beneficios	272
- Cálculo de los costos	275
- Cálculo de la tasa de retorno	277
- Aplicación de la fórmula	278
CAPITULO 13: EVALUACION SOCIAL DE LA EDUCACION	282
Evaluación del proceso - trabajo - vida social	282
Ejemplos de aplicación	284
- Evaluación de los Institutos Nacionales de Educación Media de Colombia	285
- Evaluación de cinco carreras universitarias en Chile	287
CUARTA PARTE: RECOLECCION Y ANALISIS DE DATOS	290
CAPITULO 14 : RECOLECCION DE INFORMACION	291
Información cualitativa	291
La observación etnográfica	293
- Un ejemplo de registro etnográfico	294
Estudios de casos. Historias de vida	297
- Una historia de vida	299
- Historia de vida de Elvira	299
Discusión grupal	302
Información cuantitativa	304
- La técnica Delphi	305

CAPITULO 15: ANALISIS CUALITATIVO	308
Categorización y codificación	309
Elaboración de formatos	313
- Matrices	313
- Ejemplo de matriz acumulativa (Checklistmatrix)	314
Tipificación y clasificación compuesta	316
Determinación de conexiones (linkage analysis)	318
 CAPITULO 16: ANALISIS CUANTITATIVO	 320
Análisis descriptivo	321
Diferencias significativas: Efectividad del programa	323
Asociaciones y correlaciones	326
Efectos diferenciales	329
Determinación de nexos causales	330

I N T R O D U C C I O N

Hace algunos años publicamos un pequeño texto con el propósito central de atraer la atención sobre la importancia de la investigación evaluativa y el poco desarrollo que hasta ese momento ella había tenido en América Latina. Decíamos al respecto " Comparado con el incremento progresivo de las actividades mencionadas, (programas de muy diversos tipo) llama la atención el hecho de que las agencias u organizaciones encargadas de su realización no se hayan preocupado, en grado equivalente, de estimar la eficiencia de las acciones emprendidas. Tal situación contrasta, desde luego, con la magnitud de los recursos financieros y humanos empleados en

muchos programas y con las expectativas que ellos suscitan o promueven" (Metodología de la investigación evaluativa, Bogotá, 1975). Si bien desde entonces ha habido algún aumento en el número de evaluaciones de programas sociales, creemos que todavía mantiene su vigencia el sentido general de ese llamado: por un lado, por la cantidad creciente de programas tanto formales como no-formales que día a día aparecen en nuestros países y por otro, por la mayor disponibilidad que existe actualmente de enfoques teóricos y metodológicos para su uso en estudios evaluativos. En este último aspecto convendría llamar la atención acerca de la utilización de diversos tipos de diseños -no solo experimentales- como también la importancia creciente de la información y del análisis cualitativo en la evaluación de programas de variada naturaleza.

Nuestro texto tiene la finalidad pragmática de presentar elementos conceptuales y técnicas de análisis actuales para ayudar a investigadores y a personas encargadas de la administración y funcionamiento de programas a realizar evaluaciones del mismo, sea durante su desarrollo o a su finalización. Como se podrá apreciar, no se toma posición respecto a diversas dicotomías en contraste que suelen encontrarse con frecuencia tanto en la epistemología como en la metodología de la investigación social. De esta manera, el lector podrá elegir o combinar, de acuerdo con sus propias exigencias teóricas y las necesidades de la evaluación propiamente tal, las diversas opciones que se presentan en nuestra exposición.

La primera parte del libro contiene lo que podríamos llamar la teoría y la metodología general de la evaluación para continuar, en los tres siguientes, con temas más específicos de la investigación evaluativa. En ellos, se encontraran numerosos ejemplos de programas y de evaluaciones desarrolladas en América Latina en los últimos años. Al respecto, recomendamos con mucho énfasis el estudio de esos casos no solo porque ilustran nuestro tratamiento sino porque de ellos se podrán sacar lecciones útiles para el propio trabajo profesional. La última parte, dedicada a la recolección y análisis de datos, debe considerarse como una introducción al tema, en atención a que su exposición detallada habría duplicado nuestro trabajo y, especialmente, por la circunstancia de que existen numerosos libros especializados en esas áreas.

Un libro de esta naturaleza siempre deja deudas con muchas personas. De manera especial queremos agradecer, por lo tanto, a numerosos investigadores y directores de programas de varios países de América Latina que nos hicieron participar de sus experiencias, nos proporcionaron informaciones y materiales como también nos animaron en nuestro trabajo.

En forma que pudimos colocar al comienzo de esta introducción, queremos finalmente agradecer al International Development Research Centre de Ottawa, Canadá, cuya contribución a las actividades docentes y de entrenamiento del Programa Interdisciplinario de Investigaciones en Educación hizo posible la preparación de este libro.

Santiago, diciembre de 1984.

PRIMERA PARTE
TEORIA Y METODOLOGIA

CAPITULO 1: NATURALEZA Y PROPOSITOS DE LA EVALUACION

En su acepción corriente, el término evaluación se utiliza para referirse al acto de juzgar o apreciar la importancia de un determinado objeto, situación o proceso en relación con ciertas funciones que deberían cumplirse o con ciertos criterios o valores, explicitados o no. Es así como en la vida cotidiana las personas pronuncian frecuentemente juicios sobre la utilidad, conveniencia, eficacia, compatibilidad, cumplimiento, ajuste, etc. de otras personas, recursos, procedimientos o productos, juicios que en su esencia constituyen acciones de evaluación con muy distintos supuestos, normas o contextos de referencia, en su mayoría, subjetivamente elegidos.

En el desarrollo mismo de la investigación evaluativa dirigida, en los primeros momentos, casi exclusivamente hacia la evaluación del aprendizaje es posible distinguir dos direcciones o énfasis principales, aun cuando relacionados entre sí. Por un lado, la evaluación concebida y practicada como medición; por otro, la evaluación definida y también practicada como determinación de logros o resultados en comparación con objetivos propuestos en el programa. Ambos enfoques han tenido importantes consecuencias en el desarrollo de la investigación evaluativa como lo señalamos a continuación.

La evaluación como medición equivale a un proceso de cuantificación que permite asignar números o puntajes a los objetos o personas evaluadas. Es lógico que tal conceptualización llevará a un gran desarrollo de pruebas y técnicas posibles de proporcionar información estadísticamente manejable y de tal modo que los instrumentos utilizados cumplieran con los criterios de validez y confiabilidad. Las desventajas de este enfoque son evidentes. Desde luego, la acentuación del carácter instrumental de la evaluación actuó en contra del desarrollo de lo que podríamos llamar "la teoría de la evaluación". Por otro lado, la evaluación así entendida, estrechó el campo de su aplicación a aquellas propiedades que pudieran ser expresadas en términos de medidas o puntajes, en perjuicio de otras características tan importantes como aquéllas, pero que al no poder ser cuantificadas fueron consideradas como "no relevantes" o "no significativas". Una tercera desventaja consistió en limitar la medición a variables individuales, es decir, a propiedades que se dan en las personas, con lo cual no se tomaron en cuenta aspectos de la estructura y funcionamiento de un programa que corresponden a su totalidad o

que identifican partes o componentes del mismo.

La segunda concepción de la evaluación puso en el centro de la tarea investigativa la determinación de los logros que pudo alcanzar un programa o un curso de acción, en relación con los objetivos o metas del mismo. Si bien es necesario reconocer que este enfoque llevó, en el campo de la educación, a preocuparse por los procesos y materiales curriculares que pudieran dar cuenta del rendimiento escolar y de otros resultados instruccionales, no es me-
nos cierto que el propósito básico de encontrar la correspondencia entre ob-
jetivos y logros obligó a definir los primeros en términos operacionales, medibles, con la consiguiente pérdida de contenidos de aprendizaje que no pue-
den ser sometidos a esta forma de expresión cuantitativa. Una desventaja
adicional de esta concepción fue la de hacer casi sinónimos los términos de
"evaluación" y "determinación de logros terminales", pues por el énfasis
puesto en la posible modificación final de la conducta quedó confinada a analizar los resultados obtenidos. Por eso, Guba pudo decir:

"Una tercera y tal vez la mayor desventaja de este enfoque con-
siste en que el énfasis en la conducta del estudiante como criterio llevó a que la evaluación fuera una técnica post facto o
terminal. Los datos sólo están disponibles al final de un largo período instrumental. Es tal vez irónico que una defini-
ción que sugería tan claramente la retroalimentación y su uti-
lización en el mejoramiento (del proceso de enseñanza-aprendi-
zaje G.B.) haya tenido este efecto. La totalidad de las posibi-
lidades no sólo no fueron realizadas sino que la forma de la
definición congeló la evaluación como un suceso terminal desti-
nado a emitir juicios sobre el producto". (Egon G. Guba, "The
failure of educational evaluation", en Carol H. Weiss, Eva-
luating action program: Readings in social action and educa-
tion, Boston: Allyn and Bacon, Inc. 1972; pág. 255).

Deberíamos decir todavía que las dos concepciones de evaluación expuestas más arriba llevaron a privilegiar el uso de diseños experimentales para medir los efectos de recursos y metodologías en logros educativos de muy distinta naturaleza. La adhesión por largo tiempo a estos diseños -extendida después a programas de acción de otros sectores sociales- ha sido tal vez un factor importante en el lento avance de la teoría y de la práctica de la evaluación, en general. Sin perjuicio de esta afirmación, no hay dudas que tanto la definición de la evaluación como medición de variables, como el uso casi absoluto de definiciones operacionales de los objetivos han sido superados desde el punto de vista conceptual y metodológico en los últimos años.

Una definición de trabajo

En un campo de investigación con tan poca tradición como lo es el de la evaluación de programas, no es sino natural que abunden las definiciones, en un aparente intento por poner claridad en su contenido y en sus funciones. En verdad, un buen número de esas definiciones han sido formuladas en el campo de la educación, específicamente en el área constituida por el análisis del proceso de enseñanza-aprendizaje. De ahí, algunas de ellas, en forma casi mecánica, se han proyectado hacia tareas de evaluación de actividades muy diferentes y, por lo mismo, con necesidades teóricas y técnicas de otra naturaleza. Como también sucede con frecuencia, cada una de las distintas definiciones propuestas acentúan uno o dos aspectos que el autor considera como focos principales o centrales de los estudios evaluativos.

Si bien no pretendemos superar las características anotadas, creemos conveniente ofrecer una definición amplia y flexible que permita orientar la práctica de la evaluación de programas de muy diversos objetivos y coberturas. De acuerdo con esta intención, entenderemos por evaluación de un programa:

Un tipo de investigación que analiza la estructura, el funcionamiento y los resultados de un programa con el fin de proporcionar información de la cual se deriven criterios útiles para la toma de decisiones en relación con su administración y desarrollo.

Respecto de la anterior definición, conviene hacer las siguientes especificaciones:

1. Que la evaluación sea una investigación no debería ser motivo de mayor desacuerdo si aceptamos, en términos generales, que lo que se pretende al estudiar un programa es encontrar nuevos conocimientos sobre el mismo. Pero en esta caracterización queremos destacar diversos aspectos comprendidos en el término análisis utilizado en la definición. Esos aspectos son:
 - a) descripción del estado, situación o nivel de cada uno de los componentes del programa y, en especial, de aquellos que constituyen focos de evaluación. La descripción puede ser de tipo cualitativo o puede implicar mediciones;
 - b) determinación de factores que faciliten o dificulten el funcionamiento

del programa, el logro de sus objetivos y la calidad de los mismos; y

- c) aplicación y desarrollo de esquemas explicativos o interpretativos referidos a las relaciones que se dan entre los factores mencionados más arriba y los estados descritos o medidos en la evaluación. Como se puede apreciar, este análisis -que aquí denominamos análisis evaluativo- es el centro de toda investigación que busca juzgar objetivamente el funcionamiento y los resultados de un programa de acción social.
2. Como investigación que es, conviene distinguir la evaluación de un mero proceso de juzgar subjetivamente la importancia o valor de un programa de lo que constituye su naturaleza esencial, vale decir, un proceso de análisis guiado por un esquema conceptual adecuado al objeto de estudio y en el cual se utiliza un conjunto de técnicas, cualitativas y cuantitativas, con el fin de asegurar la validez del conocimiento obtenido. De aquí que sea más apropiado referirse a este proceso de indagación con el nombre de "investigación evaluativa".
3. La definición propuesta no limita las tareas de evaluación a la determinación de los resultados finales de un programa que, indudablemente, forman parte principal del mismo. Pero, más allá de esa determinación, es de la naturaleza de la administración de un programa preguntarse si éste "trabaja bien" o "trabaja mal" y, luego -muy específicamente, en este último caso- por qué sucede esto o qué factores dan cuenta de los resultados obtenidos. Esto significa que, en mayor o menor grado, la evaluación es una

investigación explicativa. Pero sea esta u otra la dirección del análisis, preguntas de este tipo obligan estudiar la estructura, los recursos y el funcionamiento del programa como focos o áreas potenciales donde se ubican las respuestas que dan cuenta de los resultados del programa, sean positivos, negativos o nulos.

"Recordemos -dice Weiss- que la evaluación tiene como objeto ayudar a tomar decisiones. Los encargados de tomarlas tienen que saber qué fue lo que dio resultados y qué es lo que no sirvió para nada, qué es lo que debería adaptarse en todo el sistema o, de lo contrario, modificarse. A menos que la evaluación pueda presentar pruebas de la naturaleza del programa tal como existió (y no sólo del programa según lo describieron sus ejecutantes) no habrá en qué basar una decisión. En el caso extremo, cuando un programa es un éxito tremendo y cuarenta comunidades desean adoptarlo inmediatamente, ¿qué es lo que les diremos que adopten?" (Carol H. Weiss, Investigación evaluativa, México: Editorial Trillas, 1975; pág. 60).

4. La definición que hemos proporcionado le asigna una finalidad práctica a la evaluación: la de proporcionar informaciones que ayuden a tomar decisiones en relación con la marcha general del programa o en relación con programas similares que podrían desarrollarse con otras poblaciones. Por eso la evaluación es, básicamente, investigación aplicada con algunas características que conviene destacar.
- a) La razón de ser de la evaluación es la utilización de sus hallazgos, la aplicación al programa de los resultados de la investigación evaluativa. Como lo destaca un autor:

"el éxito de un proyecto evaluativo dependerá especialmente de la utilidad que le preste al administrador para mejorar los servicios (del programa) ... A diferencia del investigador básico, el investigador aplicado debe, constantemente, estar consciente de la utilidad potencial de sus hallazgos" (Edward Suchman, Evaluative Research, New York, Russell Sage Foundation, 1967; pág. 21).

- b) Es incorrecto, desde luego, pasar de la anterior característica a la conceptualización de la evaluación como ciencia aplicada. Lo que sí sucede en la práctica es la utilización de conceptos y metodologías proveniente de una o más disciplinas en cuanto ellas permiten descripciones útiles para el desarrollo del programa. Así, entonces, la evaluación es también investigación interdisciplinaria que se apoya, según el caso, en la psicología, la antropología, la sociología, la economía y otras ciencias, pero no es, en sí misma, una ciencia distintiva con su propio objeto teórico de estudio.
- c) En este orden de consideraciones es importante destacar que en el proceso de la investigación evaluativa se establece una especial relación entre teoría y práctica de tal modo que más que la aplicación en abstracto de leyes y principios generales se utilizan conceptos y experiencias específicas y apropiadas al problema o situación que se analiza. En este punto valen, a nuestro entender, los comentarios que hace Foster sobre la antropología aplicada y que luego resume así:

"... ya que la ciencia social teórica no brinda un marco suficiente al estudio aplicado, se deduce que el científico social que trabaja en este campo debe estar preparado para elaborar al menos una parte de su teoría relevante a partir

de sus propias experiencias y observaciones" (George M. Foster, Antropología aplicada, México: Fondo de Cultura Económica, 1974; pp. 217-218).

En todo caso, estas notas acerca del papel de la teoría en la investigación evaluativa tienen mayor sentido cuando ella trata de buscar relaciones entre componentes del programa y, en especial, entre algunos de ellos y los resultados obtenidos; en otras palabras, cuando se tiene un propósito explicativo. La preocupación teórica disminuye, empero, en la evaluación descriptiva, especialmente en aquella en la cual el objetivo principal y a veces único consiste en determinar si la exposición de una cierta población a un programa fue acompañada por un cambio en la dirección deseada. Mientras en el primer caso se tienen bases para generalizaciones de las características de funcionamiento y logros hacia otros programas similares, la evaluación descriptiva casi no tiene capacidad de generalización; sus resultados se refieren sólo al programa que está sometido a estudio.

- d) Digamos finalmente que otra característica que define a la investigación evaluativa como investigación aplicada la constituye el límite que deben imponerse los evaluadores en la selección de informaciones. Por lo ya dicho, es claro que su cantidad y tipo se define en función directa de las áreas del programa sobre los cuales los administradores desean o deberían actuar, de tal modo que el criterio de relevancia práctica tiende a reducir al mínimo de información que será analizada posteriormente. Cuando decimos "deberían actuar" queremos significar ocasiones en las

cuales la evaluación no tiene un carácter administrativo directo, es decir, no ha sido solicitada y convenida con las personas responsables del programa, sino surge desde fuera, pero como investigación evaluativa y con los mismos propósitos de ella. Un ejemplo claro lo presentan los estudios sobre eficiencia de los sistemas educativos o sistemas de salud; sobre la eficiencia externa de programas de capacitación de maestros, etc. en los cuales los resultados obtenidos sirven como normas de contrastación establecidas, como dijimos, desde instancias totalmente independientes del programa. La evaluación toma aquí las características de una investigación crítica.

Propósitos de la evaluación

En la definición que dimos del proceso de evaluación incluimos su finalidad básica o propósito principal: proporcionar criterios para la toma de decisiones en relación con la estructura, el funcionamiento y el desarrollo del programa. Si bien las decisiones que pueden afectarla en mayor profundidad son tomadas por sus responsables directos, existen otras decisiones que tienen que ver con las operaciones diarias, ya sea en relación con los recursos utilizados, la metodología empleada, con el trato o interacción entre las personas o, en fin, con una o más de las múltiples actividades que se desarrollan en los distintos niveles del programa.

Por la diversificación de roles que exige la marcha del programa hacia los objetivos intermedios o finales, por el ambiente cambiante de las situacio

nes en las cuales se actúa, por las expectativas convergentes o divergentes dentro del "grupo técnico" o de la población objeto o de ambos, las necesidades de información y, por lo tanto, los propósitos de la evaluación suelen ser, en la práctica, de muy diferente índole e importancia. A veces, ta les propósitos resultan difíciles de resolver en un solo estudio, pero de to das maneras es necesario tenerlos en cuenta para priorizarlos o tomar indica dores de los problemas que se relacionan con las necesidades que hayan podido ser explicitadas.

A manera de ilustración de los propósitos que pueden perseguirse en una investigación evaluativa, presentamos una lista de ellos, formulados a diferen tes niveles de generalidad:

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, o sea, la relación entre in sumos 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 homogeneizar los efectos, consecuencias e impactos del progra-
ma en diversos subgrupos de la población de referencia del programa.

Como dijimos antes, las necesidades de conocimiento sobre el programa pueden ser muy numerosas y de variada importancia y tener su origen en los adminis-
tradores, en el grupo ejecutor o en la población-objeto. Cualquiera sea el
caso, algunos de estos propósitos surgen de algún indicio o "sospecha" de que
algo "anda mal" en el programa: que sus objetivos no son comprendidos por los
usuarios, que tales o cuales recursos no cumplen la función o la utilidad
para la cual fueron pensados, que la moral del "grupo técnico" es baja, que
no tiene la capacitación adecuada, que existen conflictos con la población,
etc. Tales indicaciones junto a otras fuentes de información permiten defi-
nir los objetivos de la evaluación y, en definitiva, a preparar el diseño me

todoológico que se empleará.

Sin perjuicio de la existencia de propósitos que justifican la evaluación, Weiss (1975: 23-24) ha señalado cuatro circunstancias en las cuales no valdría la pena realizarla. Ellas son, tomadas textualmente:

1. Cuando no hay nada que preguntar acerca del programa. Se está realizando y las decisiones sobre su futuro o no se van a tomar o ya se han tomado.
2. Cuando el programa carece de clara orientación. El personal del programa improvisa sus actividades de un día para otro, basándose en pocas reflexiones y menos principios y el programa cambia y se va transformando, "vaguea" y se orienta como puede. Hay poca razón para llamar a esto "programa".
3. Cuando las personas que deberían saberlo no pueden ponerse de acuerdo en qué es lo que está tratando de realizar el programa. Si existen grandes discrepancias en las metas pensadas, la evaluación carece de fundamento sobre el cual levantarse.
4. Cuando no hay suficiente dinero ni personal lo bastante calificado para realizar la evaluación. Esta es una tarea exigente, que requiere tiempo, dinero, imaginación, tenacidad y destreza.

Podría agregarse que en algunas oportunidades, la existencia de las caracte-

rísticas indicadas en los numerales 1 a 3 forman la base de las racionalizaciones usadas para impedir la realización de una evaluación que tratamos en la sección que sigue. Por esta razón, en contra de la posición de Carol Weiss, nuestra opinión se pronuncia a favor de hacer una investigación evaluativa en las situaciones supuestas, precisamente para, con sus resultados, ayudar a la clarificación de los objetivos de los programas y al mejoramiento de su estrategia.

Resistencias a la evaluación

Al tratar de los propósitos de la evaluación decíamos que, básicamente, ellas se resumían en la necesidad de tomar decisiones en diversos aspectos del programa en vista a mejorar su funcionamiento, su efectividad, etc. Esta situación, sin embargo, no siempre es así y, por el contrario, es común encontrar diversas formas de oponerse a la evaluación. A este respecto, Suchman (1967: 144) señala que el personal encargado de un programa suele apelar a una serie de "razones", ya sea para impedir el estudio evaluativo o para disminuir la importancia que puedan tener sus hallazgos. Entre tales "razones" el autor destaca las siguientes, que traducimos sin mayores modificaciones:

1. Los efectos del programa son a largo plazo, por lo tanto, las consecuencias no pueden ser medidas en el futuro inmediato.
2. Los efectos buscados son de tipo general más bien que específico. De esta manera, no se puede utilizar un criterio único para evaluar el programa;

por otro lado, aún cuando se utilizaran muchas medidas no se lograría llegar a las consecuencias complejas buscadas.

3. Los resultados son pequeños, pero significativos; por lo mismo, no pueden ser medidas efectivamente ya que los instrumentos no son suficientemente sensitivos.
4. Los efectos son sutiles y no hay posibilidades de usar instrumentos capaces de dar cuenta del cambio de las cualidades a las cuales se refiere el programa. Además, la medición perturbaría los procesos involucrados.

Suchman señala también diversas otras racionalizaciones tendientes a disminuir la importancia de los hallazgos de la evaluación y que, por lo tanto, revelan la intención de no introducir cambios o modificaciones en el programa. Debe tenerse en cuenta aquí que, en una forma u otra, algunas personas del programa suelen estar convencidas del valor de sus actividades y de manera natural rechazan una evaluación que podría atentar contra ese supuesto. O como con alguna crudeza lo dice otro autor:

" ... A diferencia de las empresas de negocios cuyo destino está decidido por el balance, las organizaciones de servicio no sufren castigos graves cuando no logran alcanzar sus metas. Les siguen llegando contribuciones privadas o partidas del presupuesto. La clientela descontenta no tiene mayor voz ni voto en la dirección de los asuntos. No sabiendo a veces que se les está engañando, se sienten contentos con solo que alguien les preste atención. Indudablemente los administradores concienzudos no sólo desean que les vaya bien, sino obrar bien, pero pueden encontrar alegrés racionalizaciones para persistir actuando de manera ineficien

te ... Las críticas contra la metodología son un con
tra ataque cada vez más común para defenderse de la eva
luación que no hable bien de uno" (Weiss, 1975: 140-
141).

Existen todavía otras formas de ponerse en guardia contra la evaluación de un cierto programa, como decir que tratan de orientarla hacia ciertos aspec.
tos que el administrador "sabe" que funcionan bien; o alegar que el proyecto va modificando continuamente sus objetivos para adecuarlo al contexto o a los deseos de la población; que basta con la auto-evaluación de su grupo eje
cutor; que lo que en definitiva interesa es lo que se hace en el proyecto y no "resultados cuantificables"; que personas externas al programa (autorida-
des) han expresado juicios muy favorables sobre el mismo, etc.

Niveles de evaluación

El título de este texto y el desarrollo que hasta quí hemos realizado refiere la evaluación a estrategias de acción orientada a fines y que hemos denominado
programas. Si bien deseamos mantener el término como etiqueta general, en la práctica esos cursos de acción se diferencian bastante por su magnitud en términos de población atendida, por su distribución territorial, duración, va
riedad de objetivos, etc. etc. Para dar una idea de nuestra referencia: la idea de programas se extiende desde todo el sistema de educación de un país (nivel macro), pasando por un programa de reforma agraria, hasta un pequeño proyecto orientado hacia niños que han desertado de la escuela y viven en una determinada población (nivel micro). Reconocida esta dimensión, pretendemos que el término programa, a lo menos, en algunos de los aspectos que trata,

pueda ser utilizado en la evaluación de las siguientes estructuras funcionales:

1. Sistemas y subsistemas institucionales (educación, salud, agricultura, educación básica, vocacional, etc.)
2. Organizaciones institucionales (escuelas, hospitales, sindicatos, etc.)
3. Programas generales (alfabetización, educación de adultos, etc.)
4. Programas específicos (capacitación vocacional de la mujer campesina; drogadicción, etc.)

Según sea el nivel del programa serán los diseños, tipos de información y técnicas de análisis que utilice la evaluación. Dentro de estas decisiones, es indudable que al movernos de un nivel a otro, cambia el número y las características a las cuales se referirá la investigación evaluativa. Para dar una idea de esta situación: la evaluación del subsistema de educación básica podrá requerir de solo algunos indicadores globales (tasa de escolarizazación, de deserción, etc.); en cambio la evaluación de un programa destinado a promover la participación de los habitantes de una comunidad podrá requerir de una información más variada como asimismo la utilización de técnicas apropiadas para este nivel, como grupos de discusión, entrevistas en profundidad, determinación de necesidades y de expectativas de solución, etc. Con esto, queremos anticipar la flexibilidad que necesariamente debe existir en la apli

cación tanto de modelos de evaluación como de técnicas apropiadas al nivel del programa a evaluar.

CAPITULO 2: TIPOS DE EVALUACION

Por tipos de evaluación entendemos estudios evaluativos diferenciados básicamente según criterios funcionales, sin que ellos estén necesariamente determinados por algunos de los modelos que presentaremos en el próximo capítulo u obliguen al uso de técnicas específicas de investigación. Los principales tipos son los que describiremos a continuación.

Algunos tipos

Evaluación intermedia - evaluación terminal

De manera general podemos decir que las evaluaciones intermedias buscan información durante el desarrollo del programa con la finalidad de mejorarlos en los aspectos que sea necesario. En este sentido la evaluación intermedia es una actividad de apoyo permanente que evalúa el desarrollo y "comportamiento" de los diversos componentes del programa. Una tarea principal dentro de esta evaluación consiste en establecer los factores que dificultan la marcha del programa o, a la inversa, que facilitan.

Si bien la evaluación intermedia puede considerar los resultados que va alcanzando el programa, su característica principal consiste en ser una evaluación de los procesos de desarrollo y aplicación de la metodología del programa, con la finalidad señalada más arriba.

La evaluación terminal es el estudio que se realiza después que el programa ha terminado y se hace con el fin de establecer cuál fue su funcionamiento y cuáles los resultados finales estimados. La información recogida permite formular juicios evaluativos sobre el programa y derivar conclusiones que puedan ser utilizadas en la puesta en marcha de programas si -

milares. En este sentido, la evaluación terminal es una evaluación proyectiva.

Mientras la evaluación intermedia tiende a ser descriptiva y sin mayores pretensiones de formalización metodológica, la evaluación terminal busca relaciones entre los componentes y variables del programa como también explicaciones de sus modalidades de funcionamiento y de los niveles y calidades de los resultados obtenidos. Por estos objetivos, este tipo de evaluación es realizada preferentemente por evaluadores externos con el propósito de asegurar la objetividad de los hallazgos.

La mayor rigurosidad metodológica de la evaluación terminal está relacionada con la búsqueda de características del programa sometido a estudio que pueden ser aplicables en otras situaciones, es decir, que puedan ser generalizadas.

Evaluación formativa - evaluación sumativa

Tanto este tipo como su pareja, fueron propuestos por Scriven (1976) en el campo de la educación de donde, como otros aspectos conceptuales y técnicos

de la evaluación, han sido transferidos a otros campos, con evidentes ambigüedades su uso (Michael Scriven "The Methodology of Evaluation" en P.A. Taylor D.M. Gowley, Readings in Curriculum Evaluation, Dubuque, Iowa: Brown Company Publishers).

La evaluación formativa se realiza durante el proceso de elaboración del currículo y tiene como propósito proporcionar información de retroalimentación que pueda utilizarse en el mejoramiento del mismo.

En el área específica de la enseñanza la evaluación formativa se relaciona directamente con la llamada "investigación educacional y desarrollo" entendida como las actividades destinadas a preparar y validar productos instruccionales, como métodos, materiales curriculares, textos de estudio, ayudas audiovisuales, etc. (Para una presentación de este campo, ver el capítulo 18 del libro de Walter R. Borg y Meredith D. Gall, Educational Research, New York: Longman 1979).

La evaluación sumativa se realiza cuando el producto instruccional (programa textos, etc.) está elaborado con la finalidad de establecer su utilización en el aula, sus ventajas sobre otros materiales, los costos para su extensión masiva, etc.

Evaluaciones descriptivas y explicativas

En general, todos los modelos presentados en la anterior sección hacen descripciones de los distintos focos de evaluación o del programa como totalidad.

Pero esta función es el objetivo principal de las evaluaciones descriptivas, objetivo que se manifiesta claramente en las técnicas de análisis de los datos que se utilizan: porcentajes globales o referidos a subgrupos de la población, medios aritméticos, etc. Todo esto sin perjuicio de descripciones de naturaleza cualitativa propiamente tales.

Las evaluaciones explicativas, a diferencia de las descriptivas, utilizan esquemas teóricos para tratar de comprender situaciones ^{de} "éxito" y "fracaso", sea con referencia a los procesos o a objetivos instrumentales y finales del programa. Estas explicaciones, a su vez, están constituídas 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 des-confirmar.

Las técnicas de análisis que se utilizan son las propias para los propósitos buscados: correlaciones, análisis de regresión, etc. Por otro lado, en muchas oportunidades las evaluaciones explicativas parten de hipótesis que vinculan distintas características del programa. Por ejemplo, en relación con la alfabetización funcional se podría tener la siguiente hipótesis: "se supone que en el adulto alfabetizado más que en el adulto analfabeto:

- los proyectos para el porvenir de los hijos son más ambiciosos;
- las ideas relativas a la mujer, su posición social, el matrimonio han evo-

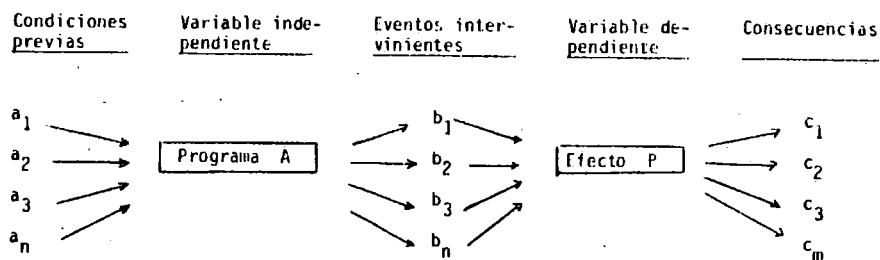
- lucionado en forma favorable;
- el deseo de participar en la evolución de la sociedad es más manifiesto;
- las tendencias a aceptar las soluciones modernas del desarrollo son más fuertes;
- el deseo de instruirse es más profundo;
- la tendencia a aceptar las explicaciones científicas de los fenómenos de la naturaleza es más acentuada".

(Roger Couvert, La evaluación de los programas de alfabetización. París: UNESCO, 1979; pp. 135-136).

Respecto de las evaluaciones explicativas, hemos dicho en otra oportunidad:

"Cualquiera sea la naturaleza de los resultados de un programa, ya se trate de 'éxitos' o de 'fracasos', el evaluador debe buscar los factores explicativos que condujeron a una situación final. En cierto tipo de programas, la no búsqueda de explicaciones dificulta establecer si fue propiamente el programa el que produjo los resultados en evaluación o bien son el producto de otros factores ajenos. De igual manera, la formulación de esquemas explicativos e interpretativos hace posible la generalización de los resultados en la forma de hipótesis y teorías que permitan aprovechar la experiencia en otras situaciones. Al identificar por medio de la explicación los factores que actúan como antecedentes o requisitos de tales o cuales efectos, pueden aplicarse otros programas que buscan objetivos iguales o similares, con un grado anterior ya de confirmación" (Guillermo Briones, Metodología de la Investigación Evaluativa. Bogotá: Universidad Pedagógica Nacional, 1975, p. 53).

Al hacer la distinción entre evaluación descriptiva y evaluación explicativa decía Suchman (1967: 84) que en este segundo tipo es necesario examinar su fundamentación racional que, esencialmente, consiste en concebir una secuencia "causal" en la cual "nuestro programa viene a ser solo uno de muchas acciones o eventos posibles que pueden producir el efecto deseado . Además nuestro programa no solo influirá para que se produzca el efecto deseado, sino que también tendrá otros efectos". De esta manera en la evaluación explicativa se debe trabajar en términos de "multiplicidad de causas" y de "la interdependencia de efectos" que el autor citado representa con el siguiente diagrama:



Finalmente queremos decir que de las evaluaciones que buscan explicaciones basadas en hipótesis o relaciones ya contrastadas deben distinguirse los juicios "interpretativos" que acompañan algunos de los modelos globales. Sin entrar en esta oportunidad a toda la complejidad del problema, diremos que *en* ellos se trata, fundamentalmente, de "comprender" (verstehen) la conducta de los usuarios del programa desde su propio marco de referencia, de acuerdo con ^{una} epistemología "subjetivista" a la cual haremos referencia al presentar este tipo de modelos.

Evaluación de objetivos internos - evaluación de objetivos ~~externos~~.

Esta distinción nace de la diferenciación entre "objetivos internos" y "objetivos externos". Los primeros se refieren a situaciones a alcanzar en la población sobre la cual actúa directamente el programa, como podrían ser nuevos conocimientos, valoraciones, etc. Precisamente, llamamos evaluación interna al proceso destinado a verificar el logro de tales objetivos y la relación con la organización del programa, la calidad de los instructores, los recursos utilizados, etc.

Si esta evaluación se realiza antes de la terminación del programa viene a ser similar a la evaluación intermedia.

Los "objetivos externos" son situaciones o estados deseables de alcanzar fuera del ámbito del programa, cuando éste ya ^{haya} finalizado, sea *en* personas que estuvieron sometidos a él o *en* personas a quienes se transfirieron los

resultados buscados. Por lo general, estos objetivos son de naturaleza económico, social y cultural, se espera que sean cumplidos en la comunidad o en la sociedad global. La evaluación externa se preocupa de este tipo de situaciones y del estudio de los factores que dan cuenta de sus distintas modalidades de cumplimiento y de los factores que explican el "éxito" o "fracaso" del mismo. Un ejemplo de este tipo de evaluación es la determinación de la rentabilidad de la educación, mientras que el cálculo de la eficiencia de un programa educativo lo es de una evaluación interna.

Sucede a veces que un programa logra los objetivos internos y no los externos. Así pasa con alguna frecuencia, por ejemplo, en programas de capacitación de maestros cuyos nuevos conocimientos no se transfieren o sólo lo hacen en mínimo grado a los estudiantes, que es la principal población de referencia de tales programas.

Evaluación externa - evaluación interna

En diversos lugares de la exposición hecha hasta aquí hemos hablado del o de los evaluadores. Pero debemos preguntarnos ahora quién o quiénes, en la práctica de la evaluación, son estas personas. La respuesta va en una de estas tres direcciones: a) es una persona externa al programa; b) es un miembro del equipo ejecutor o de la administración del programa; c) el grupo evaluador está formado por evaluadores externos y por personal del programa. Veamos las ventajas y desventajas de estas soluciones.

Para comenzar, se dice que una evaluación por investigadores externos asegura una mayor objetividad en cuanto a la apreciación del funcionamiento y resultados del programa, comparada con la evaluación hecha por personas involucradas en el mismo, que por su familiaridad con el mismo pueden dejar de ver aspectos que desde otro punto de vista son importantes. Esto sin referirnos a la conciente o inconciente parcialidad en sus juicios. Por otro lado, existe mayor probabilidad que el evaluador externo haya tomado un entrenamiento especial en la evaluación de programas como también tenga experiencias en otras situaciones que pueda aplicar en esta oportunidad.

Frente a las ventajas anteriores, se hace notar que es difícil que una persona extraña al programa pueda comprender bien sus objetivos, su metodología, su población de referencia y, en fin, las múltiples actividades diarias que constituyen el programa en su realidad dinámica. Además, se dice, no es fácil compatibilizar los puntos de vista de un evaluador externo con los que tienen, en los temas de evaluación, el personal del programa. Con lo cual se produce un área de conflicto potencial. Por las razones indicadas, aparece, entonces, como más conveniente una evaluación realizada por personas que forman parte del equipo técnico o de la administración del programa.

No es fácil que la alternativa planteada pueda resolverse en abstracto. Frente a la decisión de evaluar un programa, deberán tomarse en cuenta las ventajas y desventajas de las evaluaciones hechas por investigadores internos y externos en cuanto a calidad y confiabilidad de la información recogida, el

tamaño y naturaleza del programa y el momento en el cual se efectúa, pues bien puede ser que una evaluación intermedia, por ejemplo, a cargo de una o más personas del programa sea suficiente. Sin embargo, podríamos decir que siempre una evaluación terminal debe ser hecha por un evaluador externo, sin perjuicio que recurra, en términos de información, a los miembros del programa.

En ciertos programas, la solución es integrar un equipo mixto. Se tiene así una evaluación participativa con la cual se unen las bondades de ambos tipos extremos y se controlan mutuamente sus desventajas. Un caso especial de esta modalidad se encuentra en la investigación-acción donde la evaluación del programa es una tarea que compromete a todos sus participantes, vale decir, a los usuarios y equipo técnico o asesor del programa. Un breve análisis de este tipo de evaluación se puede ver en el capítulo 6 de este texto.

CAPITULO 3: MODELOS DE EVALUACION.

El análisis de cualquier programa incluido los que trabajan en el nivel micro, revela que su funcionamiento real implica una gran cantidad y variedad de acciones, recursos, personas, situaciones ambientales, etc. frente a las cuales el evaluador debe hacer decisiones en cuanto al diseño general de la investigación y, consecuentemente, en cuanto a la información a recoger y al tipo de análisis al cual serán sometidos los datos.

Frente a esta diversidad de aspectos, el evaluador utiliza con frecuencia un "modelo de evaluación", es decir, un esquema general que señala las características básicas de un programa que deben ser evaluadas y los procedimientos principales para recoger la información pertinente.

En el sentido anteriormente expuesto, se distinguen dos tipos generales de modelos: a) los modelos globales; y b) los modelos analíticos. Los primeros tienen especial aplicación en la evaluación de pequeños programas y los segundos, en programas de mayor envergadura.

En la categoría de modelos de evaluación global de nivel micro, se pueden ubicar los siguientes esquemas de evaluación:

1. Evaluación concertada. .
2. Evaluación "iluminativa".
3. Evaluación focalizada.
4. Evaluación sin objetivos pre-fijados.

En la categoría de modelos de evaluación analítico de nivel macro se encuentran los siguientes modelos tipos:

1. Evaluación del contexto-insumos-procesos-productos (modelo CIPP).
2. Modelos de referentes específicos de evaluación.

En la práctica misma de la evaluación, con frecuencia el investigador combina enfoques y procedimientos de uno o más de los modelos clasificados y con ellos define su diseño específico de evaluación. Por ejemplo, si se va a evaluar el sub-sistema de educación primaria de un país o una región (nivel macro), puede utilizar procedimientos de modelos globales para ilustrar algunos

aspectos específicos del programa, como podría ser la observación del proceso de enseñanza-aprendizaje en unas pocas escuelas o la discusión sobre problemas de funcionamiento con algunos profesores y directores, etc.

Modelos globales

Los modelos globales, como dijimos, tienen especial aplicación en programas pequeños, generalmente de tipo educativo no formal, dentro de cuales los usuarios o participantes a las personas encargadas de su conducción pueden establecer relaciones impersonales directas (relaciones de tipo primario, según la terminología de la psicología social). Dentro de estos modelos se distinguen los modelos subjetivistas, con cuya caracterización comenzaremos esta exposición.

Modelos subjetivistas

Las principales características de estos modelos derivan de una epistemología según la cual la tarea básica de la investigación evaluativa consiste en "comprender" la conducta humana, en captar su "significado", todo esto basado en la experiencia subjetiva y las conexiones que se dan entre ella y la conducta. Por ello, los modelos subjetivistas que

reconocen como uno de sus inspiradores principales a Max Weber con su teoría de la "verstehen", conceden especial énfasis la posibilidad de conocer y comprender a las personas a través de una introspección simpática y

de una reflexión detenida sobre las conductas e interacciones que han sido observadas directamente y descritas detalladamente. Como dice Strike (Kenneth Strike, "Explaining and understanding : The Impact of Science on our Concept of Man" en L.G. Thomas. ed., Philosophical direction of Educational Research, Chicago: University of Chicago Press, 1972;28).

"...los hombres tienen propósitos y emociones, hacen planes construyen culturas y tienen ciertos valores de tal modo que su conducta es influida por esos valores, planes y propósitos. En síntesis, un ser humano vive en un mundo que tiene 'significado' para él y a causa de que esta conducta tiene significado, las acciones humanas son inteligibles en una forma que no lo son la conducta de objetos no humanos".

A niveles más concretos, pero como expresión de esa "epistemología subjetivista", la evaluación en estos modelos utiliza a veces en forma exclusiva métodos cualitativos de investigación, como observación participante, entrevistas en profundidad, historias de vida, etc. Pero estos métodos no están dirigidos a reunir información sobre partes separadas del programa, sino que el investigador trata de comprender la totalidad (gestalt) del mismo, dentro de la cual los acontecimientos individuales adquieren su verdadero sentido.

Por otro lado, esa investigación es inductiva en cuando se trata de captar la situación con un enfoque inductivo, sin previa especificación de variables o de hipótesis. De este modo las interrelaciones que se encuentren surgen naturalmente de la experiencia con el programa. Por lo mismo, el evaluador no debe manipular el ambiente o contexto en el cual se desarrolla el programa,

como lo hace el investigador experimental: la evaluación cualitativa es también, esencialmente investigación naturalística. (Para una exposición más detallada de los modelos globales denominados allí transaccionales, ver Michael Quinn Patton, Qualitative Evaluation Methods, Beverly Hills: Sage Publications, 1980). La crítica a ellos, referida a la teoría de la "verstehen" y a sus expresiones posteriores en el funcionalismo, se encuentran en varias fuentes, entre ellos, Mario Bunge, Materialismo y Ciencia, Barcelona: Ariel, 1981, especialmente pp. 82-96).

Después de esta corta introducción, veamos algunos modelos subjetivistas.

1. La evaluación "sensitiva"

La evaluación "sensitiva" (responsive evaluation), propuesta por Stakes (Robert E. Stakes, Evaluating the arts in education: a responsive approach, Columbus, ON: Merrill, 1975), recibe su nombre del intento que debe hacer el evaluador por descubrir lo que el cliente realmente necesita antes de proceder a preparar el diseño de evaluación. Basado en esas necesidades, su tarea principal como evaluador consistiría en hacer descripciones y juicios acerca de diversos aspectos y componentes del programa, guiado por los principios generales de la epistemología subjetivista que presentamos anteriormente.

Stakes describe la esencia de su modelo en los siguientes términos:

"...Para hacer una evaluación concertada el evaluador

prepara un plan de observaciones y de transacciones. Luego se pone de acuerdo con varias personas para observar el programa y con su ayuda prepara algunas narraciones breves, descripciones amplias, representaciones, gráficos, etc. Averigua que es valioso para sus usuarios y reúne otros juicios de diversas personas cuyos puntos de vista difieren... En seguida le pide al personal encargado del programa que se pronuncie acerca de la precisión de sus descripciones y a los usuarios del programa acerca de la importancia que puedan tener sus hallazgos. La mayoría de este trabajo lo hace de manera informal, repetitivo y guarda un diario de las acciones y reacciones realizadas y encontradas" (Stake, 1975:14).

Si los usuarios lo desean, el evaluador puede preparar un informe oficial final del estudio realizado.

2. El modelo "iluminativo"

Otro modelo global lo constituye la "evaluación iluminativa" desarrollado por Parlett y Hamilton (Parlett, M.R. y Hamilton, D.F., "Evaluation as Illumination: a new approach to the study of innovatory programs", en G.V. Glass, ed., Evaluation studies review annual, vol. i, Beverly Hills: Sage Publications, 1976). Esta forma de evaluación tiene como principal objetivo "iluminar" los procesos educativos que se investigan, a través de descripciones obtenidas del funcionamiento del programa, de las actitudes y reacciones de sus participantes y del esclarecimiento de él.

De manera directa ^{la} "evaluación iluminativa" se basa en la "epistemología sub-

jetivista" que, entre otras cosas, rechaza el nexo causal que trata de distinguir entre variables independientes y variables dependientes. Su preocupación principal es la captación del significado del programa para las personas, en especial sus procesos. El significado se busca en símbolos, representaciones y formas lingüísticas usadas por los usuarios del programa. El modelo enfatiza el uso de un paradigma de investigación basado en 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 a los usuarios del programa (con olvido, muchas veces, de los intereses o propósitos de la agencia que patrocina y, desde luego, de sus costos y fuentes de financiamiento). Se supone, además, que la participación de los miembros del programa en la evaluación les permite reflexionar y comprender en mejor forma los problemas y la facilidad o dificultad de las soluciones.

En la "evaluación iluminativa" se distinguen las siguientes etapas: a) identificación de los interlocutores (usuarios, monitores, etc.) b) observación y entrevistas etnográficas (abiertas) por parte de los investigadores; c) profundización de los elementos y aspectos más importantes del programa encontrados en la primera fase (entrevistas en profundidad; estudios de casos, etc). d) validación de los resultados. Esto se hace mediante diversos procedimientos y técnicas cualitativas y cuantitativas (triangulación metodológica); e) redacción del informe final. Este informe se prepara con un lenguaje accesible

a los usuarios del programa, en el cual se colocan dibujos simples, narraciones de hechos concretos, citas textuales de los participantes, con interpretaciones psicológicas y sociológicas, etc.

En Chile se ha utilizado este modelo en la evaluación del Programa Padres e Hijos (PPH) del CIDE y al respecto ha dicho uno de sus más entusiastas presentadores:

"... Puede ser el momento de esbozar la metodología de la evaluación del proyecto, dando cuenta de sus líneas directrices. Las que se pueden denominar: 1) iluminativa, 2) hermeneútica, y 3) dialéctica.

... Iluminativa dice relación con la manera de presentar los hechos concretos... Queremos saber concretamente qué hacen las personas. considerando acciones como la mejor evidencia del compromiso... Este procedimiento se llama iluminativo, porque su afán es esclarecer los hechos, hacerlos visibles, a través de una descripción plástica e incluso ^{con} la fotografía... Hemos preferido en general las descripciones de los propios campesinos, tal como ellos las hacen, con las palabras propias de ellos... Si hay algo de novedoso en la evaluación del PPH ... esto es la valorización de las palabras de los campesinos como datos fundamentales. En vez de capacitar a los campesinos para que ellos hagan tests y encuestas, hemos valorizado sus modelos y sus deliberaciones grupales hallando en ellos la iluminación primer y básica que hace visible su realidad ...

.. Llamamos hermeneútica a nuestra metodología porque se centra en la lectura de cierta especie de texto. El texto es el discurso popular que se ha venido discutiendo...

... Por eso mismo, el método es dialéctico a la vez que hermeneútico. Vale decir, la opción por valorizar la voz de los pobres es una opción dialéctica...

el método merece tal objetivo porque privilegia el conflicto". (Howard Richards. "Presentación de la "Evaluación de la acción cultural: Estudio evaluativo del Programa Padres e Hijos", en Horacio Walker y Howard Richards, Evaluación Iluminativa y acción cultural, Santiago: CIDE, Documentos de trabajo, 1983; citas tomadas de las páginas 55 - 60).

3. Modelo de la evaluación focalizada

Más que un modelo, es una estrategia presentada por Patton en 1978 (Michael Quinn Patton, Utilization focused evaluation, Beverly Hills: Sage Publication, 1978), para hacer decisiones respecto del contenido general de la evaluación pero sin llegar por anticipado a establecer especificaciones de ese contenido. Para lograr este primer objetivo el evaluador comienza por identificar las personas que por su ubicación o situación en el programa serán las que tomarán las decisiones como también ubicará y se pondrá en contacto con los usuarios activos que, efectivamente, utilizarán la información que produzca la investigación evaluativa. Estos dos tipos de personas, en trabajo conjunto con el evaluador, logran focalizar diversos problemas y temas relevantes para la mejor marcha del programa y en base a ellos se hace la selección de las técnicas de recolección de informaciones y procedimientos de análisis más apropiados. Si bien la evaluación focalizada no excluye la posibilidad o necesidad de emplear enfoques y métodos utilizados en modelos analíticos, lo cierto es que el énfasis que pone en la comprensión de la dinámica interna del programa y la importancia central que tienen en él las técnicas cualitativas de investigación (observación, entrevistas en profundidad,

etc.) acercan esta estrategia a los modelos globales o transaccionales.

Las características anteriormente anotadas se expresan con toda claridad en los siguientes términos del autor:

"... Los métodos cualitativos son particularmente apropiados para realizar evaluaciones de los procesos. Para comprender lo único, la dinámica interna de un programa, lo mejor es acercarse a ese programa sin hipótesis predeterminadas acerca de cuáles son sus bondades y debilidades. Tal enfoque abierto permite que estos aspectos emerjan de las observaciones del programa y de las entrevistas, más bien que de las teorías y expectativas del evaluador" (Patton, 1980:61)

La focalización de la evaluación en los procesos pretende comprender la dinámica interna del programa, mediante la búsqueda de respuestas a preguntas como éstas: qué factores son los que, en conjunto, caracterizan a este programa y le dan forma? ¿Cómo se recluta a los usuarios y qué hacen una vez que están en el programa? ¿Cómo perciben al programa sus usuarios? etc. Este énfasis en los procesos significa darle más importancia a cómo se logra un cierto resultado que al producto mismo. De esta manera la evaluación de los procesos exige descripciones y observaciones detalladas y consecuentemente, el uso de técnicas cualitativas.

Respecto de los resultados, el modelo trabaja con el supuesto de que éstos serán diferentes para los diferentes tipos de participantes (resultados individualizados). Por lo mismo, la evaluación necesita recoger información que

caracterice a los diferentes usuarios, y, por lo mismo, no utiliza medidas estadísticas que pretenden caracterizar a la población en su conjunto (medios aritméticos, distribuciones de frecuencia, etc.) En su lugar, se recurre al estudio de casos y se pone especial atención en los "casos desviados" respecto de la norma o característica general. Pero aún así, la base para las recomendaciones para mejorar o ampliar el programa está en los resultados que proporcionen las evaluaciones de los distintos procesos que se dan en él.

Otros modelos globales

Existen diversos otros modelos que podemos clasificar como globales tanto por su aplicación o programas pequeños, formales y no-formales, como por el uso predominante de técnicas cualitativas de investigación en la recolección y análisis de la información. Entre tales otros modelos, se destaca el presentado por Scriven con el nombre de evaluación libre de objetivos que, aparte de sus particularidades, no se basa en la epistemología subjetivista que hemos visto en los tres modelos anteriores.

Según este nuevo modelo la evaluación debe buscar los resultados reales obtenidos por el programa, sin necesidad de determinar por anticipado sus objetivos. Se trata así de evitar una búsqueda limitada de los cambios que pudo producir el programa, entre ellos, los resultados no anticipados o efectos imprevistos. Esta tarea se cumple mejor con el uso de técnicas cualitativas de

observación y de descripción amplia, sin que ella sean guiadas y "moldeadas" por los objetivos pre-establecidos. El evaluador registra todo lo que realmente sucede en el programa: sólo así es posible conocer los reales efectos producidos, o en palabras del formulador del modelo:

"... Mientras menos el evaluador oiga acerca de los objetivos del programa, menos desarrollará una *visión estrecha* (tunnel-vision) y más atención prestará a la búsqueda de los reales efectos (más bien que comprobar los efectos supuestos" Michael Scriven, "Pros and cons about free-goal evaluation", Evaluation Comment, 1972, : 3, 1-7)"

De acuerdo con Scriven, la evaluación sin objetivos predeterminados debería seguir estos pasos:

1. Caracterización de la naturaleza del programa que va a ser evaluado.
2. Finalidades de la evaluación.
3. Indicadores de las relaciones que puedan 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 diversos tipos de costos.
7. Identificación de acciones alternativas a los programas (validez interna).
8. Identificación de los elementos del programa y sus posibles impactos en

el programa.

9. *Conclusión* sobre el valor del programa.

En esta lista, *los* primeros pasos se refieren al programa y sus resultados; los tres últimos se refieren su validez.

Si bien la estrategia propuesta por Scriven tiene el mérito de destacar la necesidad o conveniencia de identificar y describir los efectos no anticipados en el programa, el propósito básico de la mayoría de las evaluaciones es, justamente, conocer si se han logrado o no los objetivos perseguidos, y si se han logrado, saber cuáles son los niveles de logro, la eficiencia con la cual funciona el programa, etc. El demasiado énfasis puesto en la búsqueda de efectos no provistos podría tener las mismas consecuencias que el modelo trata de evitar, pero en la dirección contraria: no ubicar los cambios producidos realmente, apreciados desde los objetivos explícitamente formulados.

Modelos analíticos
=====

En distinción a los modelos que hemos presentado en las páginas anteriores, que pretenden hacer un enfoque totalizador o global del programa durante su evaluación, los modelos analíticos, como lo dice su nombre, distinguen con mayor o menor desgregación los diversos elementos dinámicos o estructurales que componen un programa, con el fin de dirigir hacia ellos las tareas de indagación evaluativa. Aún más: estos modelos admiten evaluaciones parciales,

es decir, de solo alguno o algunas de las partes o procesos que se desarrollan en el programa, cuando así lo requieren las necesidades de su administración.

En esta oportunidad presentaremos dos modelos analíticos, el denominado esquema o modelo de contexto-insumo-proceso-producto y el modelo de referentes específico de evaluación. Ambos se aplican preferentemente a programas formalizados, de nivel medio a macro, aun cuando algunos o todas sus unidades funcionales puedan ser pequeños (por ejemplo, escuelas dentro de un sistema educativo).

1. El modelo contexto-insumo-proceso-producto (modelo CIPP)

Este modelo, como otros aportes a la teoría y a la práctica de la evaluación, se generó dentro del campo educacional, especialmente referido a la evaluación curricular. Fue propuesto por Stufflebeam y otros investigadores en 1971 (Stufflebeam et al., Educational Evaluation and Decision Making, Itasca, Ill.: F.E. Peacock, 1971). Desde entonces ha tenido una amplia aplicación en la evaluación de programas educativos de muy diversa naturaleza. Convendría advertir, sin embargo, que el modelo CIPP, como algunos de los modelos globales que fueron reseñados en la anterior sección, conservan en sus respectivas estructuras y orientaciones metodológicas características del sistema educativo norteamericano, muy diferentes, desde luego, de aquellas que definen los sistemas latinoamericanos.

El modelo CIPP empieza por reconocer como función principal de la evaluación la de proporcionar informaciones útiles para la toma de decisiones. En lo que se refiere a la información misma, el evaluador debe resolver los siguientes problemas: a) tipo de información específica requerida por los decisores; b) formas para obtener la información, organizarla y analizarla y c) elaboración de síntesis en la perspectiva de que ellas sean útiles para los propósitos establecidos en la evaluación.

En relación con el programa mismo, las decisiones se clasifican en cuatro categorías:

- a) decisiones de planeamientos
- b) decisiones de estructuración
- c) decisiones de implementación y
- d) decisiones de reciclaje.

Las decisiones de planeamiento se refieren a las modificaciones que es necesario introducir en el programa cuando se constata la discrepancia entre los cambios definidos en los objetivos y los que se están produciendo en los hechos. Las decisiones de estructuración recaen sobre los medios elegidos por alcanzar los objetivos, sean éstos métodos contenidos, organización, personal, presupuesto, calendario, etc. Las decisiones de implementación se relacionan con la operacionalización y ejecución del programa y comparan la metodología propuesta con los procesos reales que se dan en el programa.

Finalmente, las decisiones de reciclaje apuntan a los resultados del programa y a sus relaciones con los objetivos propuestos. Los resultados pueden ser resultados finales o intermedios.

La información para las cuatro decisiones nombradas está dada por cada una de cuatro tipos de evaluación, que forman el centro del modelo CIPP: a) evaluación contextual; b) evaluación de insumos; c) evaluación del proceso; y d) evaluación del producto.

a) Evaluación contextual. Consiste en un estudio exploratorio destinado a ubicar los problemas o necesidades educativas no satisfechas en un cierto contexto educativo, con el fin de diseñar un conjunto de objetivos específicos sobre la base de los cuales pueda elaborarse un programa instruccional. Por otro lado, si se está evaluando un programa en marcha, la evaluación contextual podría encontrar necesidades y aspiraciones, sea de estudiantes, profesores o padres de familia, que no están incluidos en el currículum original.

b) La evaluación de insumos trata de especificar los recursos que se necesitarían para alcanzar los objetivos definidos en los programas, ya sea que estos provengan de fuentes internas o externas. También la evaluación establece diversas formas en las cuales esos recursos deberían emplearse y señala la estrategia que mejor serviría para alcanzar los objetivos instruccionales.

c) Evaluación del proceso. Se dirige a establecer si la estrategia definida

en la fase anterior se está aplicando o no y a ubicar los factores que pueden dificultar el desarrollo de los procesos instruccionales. De esta manera, podrá proporcionar información a los administradores del programa que les permitan anticipar y resolver esos problemas.

d) Evaluación del producto. Tiene como objetivo principal medir e interpretar el logro de los objetivos instruccionales, ya sea durante el desarrollo del programa educativo o a su terminación. Los logros alcanzados se comparan con las expectativas que se tenían respecto de los niveles a alcanzar.

La preocupación del modelo CIPP por la toma de decisiones y por la correcta identificación de las personas que requerirán la información que proporciona se nota con claridad en esta cita de los autores del modelo:

"... Esta interacción del evaluador y el decisor se puede compendiar en términos de un conjunto de preguntas que deben responderse:

- Quién tiene la autoridad legal para tomar la(s) decisión(es)?
- A quién se ha delegado la responsabilidad de tomar la(s) decisión(es)?
- Quiénes son las personas o grupos extralegales que influyen en las decisiones o las ratifican?
- Quiénes son los clientes o usuarios probables de la información?
- Qué nivel de información se suministrará a esos clientes o usuarios?
- Cuándo se necesita la información?
- Puede o no la evaluación satisfacer de manera realista el itinerario-cronograma de las necesidades de información?
- Técnicamente es factible suministrar la información?
- Qué clase de decisiones se van a tomar: (planeamiento, estructuración, implementación o reciclaje)?

-Cuál es la situación de cambio (homeostático, incremental o neomobilístico)?

2. El modelo de referentes específicos de evaluación

Es un esquema de referencia para la orientación de investigaciones evaluativas desarrolladas por el autor de este texto aplicado a programas educacionales y sociales de diversos países latinoamericanos.

El modelo parte por enfatizar el carácter de investigación aplicada de la evaluación y, por tanto, la relevancia directa que debe tener su diseño y sus resultados para la toma de decisiones en diversas instancias y momentos de la marcha de un programa.

Las decisiones pueden corresponder a los administradores del programa y referirse a su totalidad a algunos de sus componentes (recursos, estrategia general, usuarios, etc.) o bien las decisiones pueden corresponder a los ejecutores directos de la metodología del programa, sea en los aspectos técnicos de ejecución, *en sus* relaciones con los usuarios, etc. Debido a estas diversas necesidades de *información*

es conveniente, según este modelo, tener claridad respecto de cuáles han de ser los referentes específicos de la evaluación .

Dentro de la intención principal ya señalada, las principales características del modelo de referentes específicos son los siguientes:

1. La evaluación de cualquier programa, cuando las necesidades de información así lo requieren, se dirige ya sea a sus objetivos, a su estructura, a su funcionamiento y a sus resultados o a sus usuarios. Tales componentes reciben el nombre de "objetos de evaluación". Tales aspectos pueden ser observados a nivel general (por ejemplo, el funcionamiento del programa) o a niveles específicos (ejemplo, el funcionamiento administrativo del programa). Según esto, se distinguen "objetos generales" y "objetos específicos" de evaluación.
2. La evaluación es, por principio, descripción, medición y comparación de propiedades que definen el estado de los objetos de evaluación seleccionados, sean éstos elementos estructurales o procesos. Estas propiedades o atributos se denominan "variables de evaluación". Así, si el referente específico de evaluación son los resultados que pudo producir el programa, entonces la evaluación misma se dirige hacia la descripción o medición de características de esos resultados, como podrían ser los niveles de logro o efectividad del programa, su calidad, etc. Si el objeto de evaluación son los usuarios del programa, el atributo de evaluación podría estar constituido por sus actitudes frente a los ejecutores del programa, o a sus expectativas, etc.

En algunos textos suele denominarse, **en** forma equívoca, criterios a las llamadas variables de evaluación (término que también utilizó este autor en su trabajo introductorio de 1975, pág. 47).

3. Cuando las características de los objetos de evaluación no pueden describirse o medirse directamente, es necesario recurrir a indicadores, sean éstos de naturaleza objetiva o subjetiva, ^{o cualitativos.} cuantitativos. Según lo dicho, la evaluación no siempre utiliza variables entendidas en la forma que se presentan en estadística ya que en muchas oportunidades es necesario o conveniente hacer descripciones cualitativas de situaciones o procesos componentes del programa o de su contexto.
4. El modelo de referentes específicos distingue entre variables o propiedades de evaluación correspondientes al programa como totalidad y variables o propiedades correspondientes a ese programa. De este modo, por ejemplo, la evaluación puede considerar tanto:

- la eficiencia del programa como la eficiencia administrativa.
- la efectividad del programa como la efectividad del equipo técnico; etc.

No siempre, como en los ejemplos dados, a cada propiedad del programa corresponde una propiedad similar en algún componente.

5. Como el proceso de evaluación es, por esencia, una actividad de compara-

ción y juzgamiento, aparece como evidente que es necesario especificar cuáles son los referentes de comparación, es decir, las normas o criterios que permitirán formular juicios sobre el programa o sus componentes en términos de "buen funcionamiento", "regular"; etc. En otras palabras esas normas permiten hablar del "éxito", "~~fracaso~~" o situaciones intermedias del programa.

Estos referentes de comparación pueden tener diversos orígenes, como los que se señalan a continuación:

- a) Cifras o situaciones correspondientes al mismo programa en años anteriores o correspondientes a programas similares.
- b) Promedio o situaciones nacionales.
- c) Criterios internacionales para clasificar situaciones en diferentes ámbitos institucionales. Por ejemplo, promedio de las tasas de deserción en América Latina; tasas de mortalidad infantil; etc.
- d) Cifras o situaciones propuestas por los administradores del programa, por los evaluadores o por ambos.

Cualquiera sean los referentes a utilizar, conviene explicitarlos con anterioridad a la evaluación misma, con el fin de evitar discrepancias posteriores.

Por ejemplo, supongamos que en un programa de capacitación laboral se inscriben 1.000 personas, de las cuales terminan 300, o sea, un 30%. Indica esta cifra el "éxito" o el "fracaso" del programa? Si ese porcentaje se aprecia por personal del programa, tal vez pueda decir: "El 30% de los alumnos que iniciaron el curso lo terminaron satisfactoriamente". En cambio, una persona externa al mismo podría decir: "Apenas un 30%, es decir, menos de un tercio de los alumnos que lo iniciaron, lograron terminar el curso". El ejemplo pretende mostrar la necesidad de tener una norma establecida para apreciar los resultados obtenidos por un programa, o la forma y niveles de funcionamiento, etc.

6. El análisis de los objetivos del programa constituye el primer objeto de evaluación. En este sentido, dicho análisis constituye una "pre - evaluación básica" que ayuda a definir el análisis real o posible de evaluación, frente al modelo ideal que pudieron diseñar los autores del programa. (ver capítulo sobre evaluación de objetivos)
7. En la evaluación de objetivos, es conveniente diferenciar entre "objetivos finales" que son los objetivos del programa en su totalidad y "objetivos instrumentales" que son situaciones, niveles, etc., deseables de alcanzar en estructuras o procesos componentes (objetivos de funcionamiento, de recursos, etc.)
8. En relación con los objetivos finales, el modelo distingue varios tipos de resultados: previstos - no previstos; logros - consecuencias, etc. (ver evaluación de resultados).
9. Respecto de los procesos, se hace la distinción entre procesos de mantenimiento y desarrollo del programa y procesos constituidos por la aplicación de la metodología para conseguir los objetivos (ver evaluación de procesos).
10. El modelo de referentes específicos utiliza, tanto en la recolección de informaciones como en el análisis de los datos, técnicas y procedimientos cuantitativos y cualitativos. En estas tareas, la evaluación se guía

por los siguientes criterios: 1) propósitos de la evaluación; 2) objetivos de evaluación; 3) objetos y variables de evaluación y 4) diseño metodológico básico (experimental, no-experimental, mixto).

Representación esquemática del modelo

Las características del modelo de referentes específicos desarrolladas en la acción precedente se incluyen en el esquema N° 1, con algunas especificaciones de los principales objetos de evaluación que pueden ser referentes de una investigación evaluativa según sean las necesidades de información para la toma de decisiones.

La especificación, ya sea en términos de objetos, características o indicadores pueden referirse a cualquiera de los objetos generales de evaluación de la primera columna.

Debe agregarse que la utilización de los componentes del modelo depende de las características del programa (grande, pequeño, formal, no-formal; etc.) y de los propósitos de la evaluación. En algunas situaciones, la evaluación podrá dirigirse a unos pocos objetos, en otros a todos los que comprende el modelo. De igual manera, la información recogida (señalada aquí por los indicadores) podrá ser cualitativa y cuantitativa. También en algunos casos la evaluación será de carácter descriptivo o bien podrá dar explicaciones tanto de la forma y niveles de funcionamiento, de las reacciones de los usuarios, de los niveles de logro de los objetivos, de su calidad, etc.

A. Objetos generales de evaluación

1. Objetivos del programa
(pre-evaluación de base)
2. Estructura
3. Administración
4. Recursos logísticos
5. Funcionamiento
6. Población o usuarios
7. Grupo técnico
8. Contexto de interacción
9. Metodología (estrategia y recursos instrumentales)
10. Resultados

B. Especificación: objetos específicos de evaluación

C. Características o variables

Información
o
Indicadores

Quantitativas -objetivos
-subjetivos

Cualitativos -objetivos
-subjetivos

D. Análisis evaluativo:

Descriptivo
(Cuantitativo-Cualitativo)

Comparación --> Explicativo
(normas)

E. Informes

F. Decisiones

G. Retroalimentación

CAPITULO 4: VALIDACION DE UN PROGRAMA DE ACCION

En el Capítulo 2 de este texto hicimos una breve presentación y diferenciación de dos tipos de evaluación: la explicativa y la descriptiva. Ahora que remos decir que ya sea que se elija uno u otro de esos tipos básicos, no puede dejar de reconocerse que todo programa es una estrategia destinada a producir cambios en un determinado contexto. Esta estrategia se basa en el supuesto de que "las actividades A, B, C,... (el programa o "variable independiente) ... producirán los resultados ... X, Y, Z, ("variable dependiente)". Desde luego, esta conexión entre esas actividades y esos resultados es una relación a fines, de naturaleza causal, sin cuya aceptación no tendría sentido montar un programa que, por definición, se dirige a producir modificaciones de diversos tipos en un cierto ambiente o en una cierta población. Pero por otro lado esa conexión es una hipótesis que necesita ser contrastada

empíricamente, actividad o proceso que aquí denominamos con los términos de validación del programa.

Si bien la validación del supuesto básico tiene diferente importancia en la evaluación descriptiva que en la explicativa, su vigencia teórica como soporte del programa no puede ser ignorada cualquiera sea el tipo metodológico básico que se utilice. Esto dicho sin perjuicio de reconocer que el análisis de "causa-efecto" no constituye un foco principal de evaluación en la mayoría de los modelos "subjetivistas" o "globales", según lo hicimos presente en la sección dedicada a su exposición.

Que el modelo de relación medios o fines es la base estructural y teórica de todo programa de acción, cualquiera sea su naturaleza y la estrategia diseñada para producir el cambio, puede constatarse a primera vista en la presentación que se haga de ellos. Así sucede, desde luego, en los ejemplos que damos a continuación.

1. Programa de desarrollo rural (México)

Medios

El objetivo del programa es proporcionar inversiones y servicios en áreas rurales seleccionadas

(La 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 N° 332, 1973, pág. 9)

Fines (objetivos)

Con el fin de:

- a) elevar los ingresos y niveles de vida del sector rural mediante la introducción de actividades directamente productivas;
- b) aumentar los niveles de empleo permanente y temporario;

llo Rural PIDER - México, Banco Mundial. Documento de Trabajo N° 332, 1973; pág. 9)

- c) contribuir al logro de una distribución más equitativa del producto social mediante un mejor equilibrio entre el desarrollo de las ciudades y las comunidades locales.

2. Programa de capacitación campesina (Chile)

Mediante la utilización de la radio el programa pretende

(Ana María de Andraca y Marcela Gajardo. Un proceso de aprendizaje colectivo, PIIE-ICECOOP; pp. 52-53), 1979.

- a) posibilitar una mayor participación e integración del campesino a sus organizaciones (cooperativas, sindicatos, etc.) y una formación oportuna y adecuada tanto de la realidad de las mismas como de la vida nacional; y
- b) aprovechar las condiciones masivas del medio de comunicación radial para establecer un circuito permanente de relaciones entre el campesino, sus organizaciones de integración y los demás sectores del país.

3. Programa padres e hijos (Chile)

Mediante el entrenamiento de los padres, el aumento de las horas de contacto con sus hijos, la lectura y análisis de folletos; la sensibilización de la comunidad para apoyar nuevas actitudes hacia los niños.

(L. Brahm; E. Schiefelbein y T. Scully. Un caso de evaluación formativa; proyecto padres e hijos. CIDE, 1980; cuadro 3).

Se pretende lograr el desarrollo de la potencialidad intelectual, afectiva y motriz del niño (de 0 a 4 años de edad) en su situación actual y concreta.

4. Programa de educación abierta y a distancia (Colombia)

La educación superior abierta y a distancia es un enfoque y una estrategia metodológica de organización y administración

que busca ampliar las oportunidades de acceso a la educación superior, formal y no formal, facilitando el ingreso a esta modalidad educativa de un mayor número de estudiantes.

(Humberto Serna. Universidad abierta y a distancia. Bogotá: Instituto Colombiano para el Fomento de la Educación Superior, 1983, Introducción, iii).

5. Proyecto experimental de educación bilingüe-Puno (Perú)

Implementar un programa primario bilingüe en centros educativos seleccionados del medio rural de las áreas quechua y aymará del Departamento de Puno

El proyecto tiene como finalidad última contribuir al mejoramiento cualitativo de la educación en las zonas rurales del Departamento de Puno; coadyudar al proceso de revalorización y desarrollo de las lenguas y culturas nativas ...

(Ministerio de Educación - Instituto Nacional de Investigación y Desarrollo de la Educación, Proyecto Experimental de Educación Bilingüe - Puno, Lima. 1982. pp. 5 y 13.

6. Programa de solidaridad social por cooperación comunitaria (México)

El Instituto Mexicano del Seguro Social se compromete a establecer y operar durante este año 890 nuevas unidades médicas rurales y urbanas

El programa tendrá por objetivos propiciar la ampliación del ámbito de cobertura de los servicios de solidaridad social y la consolidación de un sistema de cooperación comunitaria en beneficio de las colectividades destinatarias de dichos servicios.

(Programa IMSS-COPLAMAR de solidaridad social por cooperación comunitaria, México: COPLAMAR, Presidencia de la República, 1981, pp. 10-11)

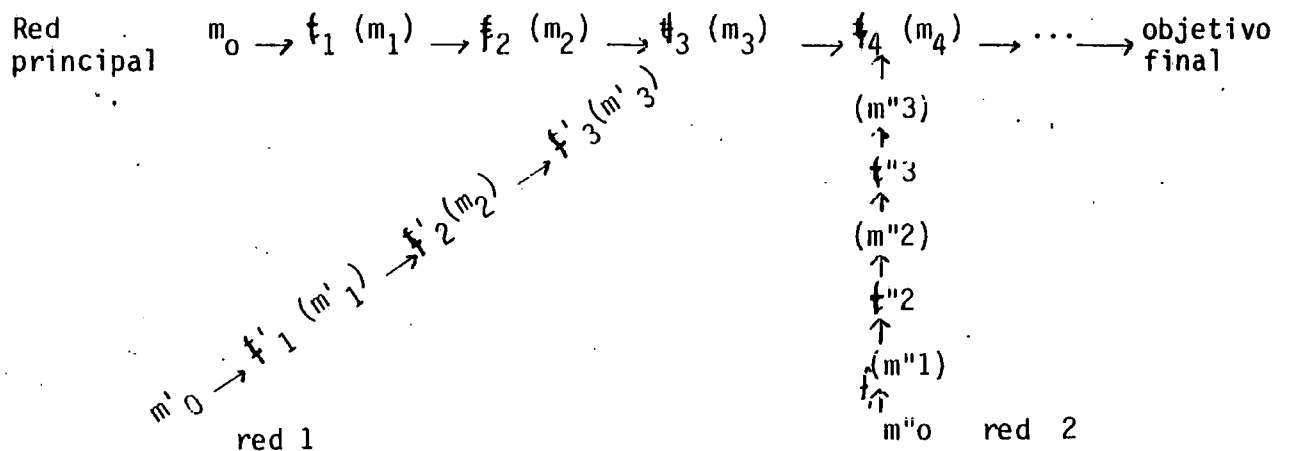
El modelo de medios a fines toma diversas formas de acuerdo con la mayor o menor complejidad del programa, lo que tiene que ver con sus objetivos y las estrategias que se utilizan para su logro. Entre ellas queremos destacar dos tipos principales: a) la "cadena lineal" con fines intermedios y b) el modelo de "redes causales".

- a) El modelo que aquí llamamos "cadena lineal" implica un conjunto de sucesos variables intervinientes a través de las cuales se logra el objetivo final del programa. Podemos representarla así:

$$m_0 \rightarrow f_1(m_1) \rightarrow f_2(m_2) \rightarrow f_3(m_3) \rightarrow \dots \rightarrow \text{objetivo final}$$

En esta cadena las f_i representan fines u objetivos intermedios, los cuales, a su vez, sirven de medios para el logro de otros fines que siguen en la secuencia y así, sucesivamente, hasta llegar al objetivo final.

- b) El modelo de "redes causales" toma formas como esta:



El diagrama quiere decir que, por ejemplo, para lograr el fin ₃ se requiere tanto de las acciones en secuencia de la red principal como de las acciones o medios de la red N° 1. Lo mismo podría decirse del fin ₄. En términos de evaluación es importante poder establecer estas redes y, en particular, la red a la cual pertenecen algunas acciones. En otras palabras, la tarea de evaluar debe poder responder a esta pregunta ¿a qué red causal pertenece la acción, medio o recurso X? Un ejemplo de un programa al cual subyace un esquema de redes es el proyecto bilingüe del Perú que citamos hace poco. Allí se dice que su desarrollo "comprende la realización de acciones interdisciplinarias en cinco áreas estratégicas de trabajo: a) investigaciones; b) adaptación del currículo; c) elaboración y producción de material educativo; d) capacitación y supervisión y c) difusión" (op. cit. pág. 14).

En lo que dice relación con la validación del modelo de medios a fines, la investigación *evaluativa* puede llegar a algunas de las conclusiones que se presentan en el cuadro N°

Cuadro N° 1

Los medios instrumentales



actuaron como
"causas" de
los resultados

no actuaron como
"causas" de los
resultados obtenidos (hubo otras
"causas")

actuaron como
"causas" de re
sultados no pre
vistos.

no produjeron resulta
dos que hayan sido per
cibidos hasta el momen
to.

(1)

(2)

(3)

(4)

En ciertos programas, piénsese en los de vacunación, como ejemplo - la efectividad entre medios y resultados ha sido comprobada, con alta confiabilidad, por anticipado; en otros, en cambio, esa efectividad está lejos de ser obvia, entre otras razones por la complejidad de las situaciones a modificar, o por vacíos teóricos en las disciplinas de base que se refieren a ellas (respecto a actitudes, a prácticas innovadoras, formas de "conciencia", etc., por ejemplo). De igual manera y de igual importancia es la necesidad de verificar la validez de los instrumentos o de los procedimientos empleados para la producción de cambios.

"El mayor obstáculo -dice un autor preocupado de estos temas- al desarrollo de un programa de investigación de signado para comprobar los componentes en el cambio de conductas es que, a pesar de la existencia de una gran cantidad de estudios en temas como cambio de actitudes, dinámica de grupos, influencias sociales en el aprendizaje y relaciones inter-grupales, este material no ha sido sistemática y cuidadosamente examinado con el fin de determinar qué componentes de los procesos de cambio

han sido científicamente validados". (John Mann, "Outcome of evaluative research, en Weiss, Evaluating action program, pág. 280)

Los problemas aquí señalados y que hemos aplicado a la investigación evaluativa deberían tomarse en cuenta en la etapa de formulación de los programas de acción, en especial, en la tarea de elegir la metodología destinada a producir modificaciones en la población de referencia. En ciertas situaciones orientadas a la introducción de innovaciones conviene poner en marcha proyectos demostrativos que permitan una primera validación del esquema de medios o fines que sirve de base al programa mismo, antes de proceder a su aplica-ción masiva o generalizada. (A esta situación y a otros aspectos de la validación nos referimos en la siguiente sección).

Al mismo punto se refieren los autores de un estudio analítico referido a diversos programas destinados a inducir a los usuarios de los servicios de eléctricidad de ciertas comunidades norteamericanas a ahorrar en el consumo. Entre otros aspectos se refieren a la metodología empleada por varios de esos programas basados en las técnicas de marketing y dicen al respecto:

"En verdad, aun en el campo de la investigación de mercados, existen dudas acerca de los efectos de la propaganda sobre la conducta. Tal vez el mejor ejemplo aquí es el experimento de Ackkott y Ems-hott (1975) por el cual el presupuesto de publici-dad de Anheuser-Busch se modificó en gran propor-ción en algunos mercados. Ackott y Emshott no en-contraron ningún efecto en la cantidad de cerveza comprada lo cual llevó a Anheuser-Busch y a otros

a dudar que la publicidad tiene cualquier efecto predecible. En suma, el aumento en la publicidad no resulta necesariamente ~~en~~ aumento de la conducta deseada. Más aún, la utilidad de la publicidad dentro del campo de la conservación de energía no ha sido establecida".

(Lawrence T. White et al., "Energy conservation research of California's utilities: a metaevaluation", Evaluation Review, vol. 8 No 2, April 1984, pp. 167-187).

Diseños de validación

Como lo dijimos anteriormente, la efectividad de un programa, es decir, su capacidad para lograr los objetivos que se ha propuesto, no pasa de ser una hipótesis en tanto no haya sido sometida a contratación empírica y así pueda mostrar su validez. Ahora bien : este proceso de validación debería hacerse antes de comenzar el programa mismo, en un programa demostrativo o en ensayo experimental de magnitud reducida. El ensayo, de mostrar resultados po sitivos permitiría, a su vez, introducir modificaciones en la estructura, en el funcionamiento, en la metodología utilizada, etc. según conviniera y si el programa mostrara no ser efectivo podrían buscarse los factores que impiden que lo sea y al actuar sobre ellos logra los resultados buscados.

En la realidad, lo cierto que un gran número de programas se inician sin ha ber cumplido esta etapa de validación por lo cual se hace necesario introducir como objetivo principal de la evaluación la determinación de la efecti vidad del programa, como el programa ya está en marcha, los procedimientos de validación suelen presentar diversas dificultades y debilidades metodológicas.

El problema de la validación de un programa consiste en encontrar procedimientos o estrategias que permitan asegurar que los resultados que obtiene el programa (si es que se produce) se deben efectivamente a las acciones desarrolladas en él y no a la intervención o influencia de factores ajenos al mismo. Se trata, como se ve, de establecer la llamada validez interna del programa. Al respecto, Campbell y Stanley destacan ocho factores que pueden perturbar o alterar dicha validez (Donald T. Campbell y Julian C. Stanley, Diseños experimentales y cuasi-experimentales en la investigación social, Buenos Aires: Edic. Amorro.tu, 1973). Ellos son:

1. Historia. Como los programas se desarrollan en períodos más o menos prolongados, se presentan diversas oportunidades y fuentes para que otras acciones distintas a las acciones del programa puedan dar cuenta de los cambios o curridos en la población-objeto.
2. Maduración. En programas que buscan modificaciones de tipo intelectual y/o afectivo debe tenerse en cuenta que el paso del tiempo produce cambios biológicos y psíquicos en los individuos, independientemente del tratamiento del programa.
3. Efecto de la prueba. Se trata aquí del efecto que pueda tener la aplicación de una prueba para constatar la situación de "entrada" de las personas en las variables de evaluación sobre los resultados finales apreciados con una medición "después".

4 Instrumentación. Los posibles cambios detectados podrían deberse a diferencias entre el instrumento (prueba) aplicado antes de iniciar el programa y el instrumento aplicado después o al final.

5. Regresión estadística. Este factor se presenta cuando para detectar cambios en la variable a evaluar se seleccionan personas que tienen características que pueden ser consideradas "altas" y que están correlacionadas con la primera; la segunda medición puede mostrar valores más bajos que los iniciales, por la tendencia de todos los valores de moverse hacia la media aritmética. La situación inversa al ejemplo general también se produce. La explicación de estas situaciones consistiría que en las primeras mediciones se presentaron factores al azar que no se presentan en las mediciones posteriores.

6. Selección diferencial. Consiste en la diferencia de características que pueden tener las personas de los grupos experimentales y de control. La solución a este problema se busca, como se sabe, por la asignación al azar de las personas a ambas condiciones.

7. Mortalidad experimental. Es la pérdida de personas que puede producirse tanto en los grupos experimentales y de control y de tal modo que modifiquen la igualdad que se había logrado con la randomización.

8. Interacción selección - maduración. Es el efecto perturbador que se produce cuando los grupos, por diferencia en su composición de edades, tienen

períodos distintos de maduración física y psicológica.

Frente al problema planteado, la investigación evaluativa encuentra a su disposición dos tipos de diseños metodológicos básicos que, en mayor o menor medida, controlan la validez interna de los resultados de un programa. Son ellos: a) los diseños experimentales; y b) los diseños cuasi-experimentales.

Diseños experimentales

Los diseños experimentales controlan en mayor nivel que los otros, la validez interna de un programa. Existen diversos tipos, de los cuales, a manera de ilustración, sólo nos referiremos a dos: a) el diseño experimental clásico; y b) diseño con mediciones "después". (Para otros diseños ver Campbell y Stanley op. ct.).

a) Diseño experimental clásico

En su versión clásica el diseño experimental se compone de dos grupos, uno experimental y otro de control, al primero de los cuales se le aplica un estímulo o tratamiento y al otro no.

El diseño tiene por objetivo verificar el efecto de una variable independiente-el programa, en nuestro caso - sobre una determinada variable dependiente. En su versión clásica, el diseño experimental comprende los siguientes pasos:

alos
1) asignación al azar de las personas a los grupos experimental y de control; 2) medición de la variable dependiente (conocimientos, actitudes, etc.) tanto en el grupo experimental como en el de control (pre-test); 3) aplicación del estímulo al grupo experimental (conferencias, discusiones, lecturas, acciones del programa, etc.); y 4) medición de la variable dependiente en ambos grupos (post-test).

La asignación al azar de las personas a los grupos es una característica esencial del diseño experimental clásico y tiene como finalidad lograr la igualación de las características de los sujetos de tal modo que se pueda hacer la atribución del efecto a la variable experimental y no a diferencias en la composición de los grupos.

El diseño experimental de dos grupos se representa habitualmente en esta forma:

$$\begin{array}{lcl} E & : & O_1 \quad \times \quad O_2 \\ C & : & O_3 \quad \quad O_4 \end{array}$$

Los subíndices "1" y "3" significan mediciones "antes" y los subíndices "2" y "4", mediciones "después". La "x" indica la introducción de la variable independiente (tratamiento, estímulo o acciones del programa).

Una vez terminado el experimento, si el investigador encuentra diferencias a favor del grupo experimental, como los grupos han sido igualados en sus características, salvo la de recibir o no el programa, entonces puede concluir

que esas diferencias han sido ocasionadas por el programa.

El análisis de los datos comienza por comparar los medios aritméticos O_1 y O_3 para establecer la igualdad de los grupos en el estímulo. Si no existe diferencia significativa, la comparación de los medios de O_2 y O_4 mostrará el efecto del estímulo o de las acciones del programa.

El diseño experimental clásico controla los factores de invalidación denominados "historia", "maduración", "regresión", "selección" y "mortalidad", los tres primeros por usar un grupo de control y los dos últimos, por la asignación al azar de las personas a los dos grupos.

Ejemplo de aplicación

A continuación hacemos un resumen de la validación de un programa de estimulación precoz en Chile en la cual se utilizó un diseño con tres grupos experimentales para tres variables diferentes y un grupo de control.

De acuerdo con los autores, el Programa de Estimulación Precoz está destinado a lactantes de nivel socio-económico bajo, de entre 0 a 2 años de edad, con los siguientes objetivos:

- a) Incrementar significativamente el desarrollo psíquico de los niños.

- b) Aumentar la eficacia de las madres como agentes de estimulación de sus hijos.
- c) Demostrar que un modelo factible y eficaz es utilizar la infraestructura existente en el Servicio Nacional de Salud (Chile), capacitando previamen al personal paramédico disponible.
- d) Proponer un procedimiento que sirva de base para extensión masiva del programa.

En lo que se refiere a los objetivos de evaluación-validación propiamente tales, se formuló una hipótesis general de la investigación y cuatro hipótesis específicas que determinaron el diseño experimental a usar. Esas hipótesis son:

a) Hipótesis general.

Es posible prevenir el retraso del desarrollo psíquico del lactante de nivel socio-económico bajo a través del Programa de Estimulación Precoz propuesto.

b) Hipótesis específicas.

1a. El currículo (manuales de estimulación) aplicado desde el comienzo, acompañado de visitas periódicas al hogar y de una evaluación mensual del desarrollo psicomotor del niño, producirá un incremento en el desarrollo psíquico de los niños de nivel socio-económico bajo de tal magnitud que alcanzarán el mis

mo rendimiento que un grupo de niños de nivel socio-económico medio-alto, no sometidos al programa.

2a. La evaluación periódica del desarrollo psicomotor, per se, producirá un incremento en el desarrollo psíquico de los niños.

3a. El contacto periódico de las madres con el personal de salud, a través de visitas al hogar, producirá, per se, un incremento en el desarrollo de los niños.

4a. El currículo, acompañado de evaluación mensual y visitas domiciliarias periódicas, e iniciado desde el nacimiento, producirá un incremento del desarrollo psíquico de los niños, mayor que si es iniciado más tardíamente.

Para someter^a validación la primera hipótesis del programa (la única que trataremos aquí) se conformó un grupo experimental, con 42 niños de ambos sexos, los que fueron sometidos al curriculum de estimulación a partir de su nacimiento y hasta los 24 meses de edad, es decir, fueron estimulados sistemáticamente durante 24 meses.

Las madres de estos niños recibieron mensualmente un manual con instrucciones y fueron visitadas periódicamente. Los niños de estos grupos fueron evaluados mensualmente en su desarrollo psicomotor, con una escala debidamente estandarizada.

Los niños fueron asignados al azar a este grupo experimental de una muestra de 159 madres tomadas de cuatro consultorios pediátricos ubicados en cuatro sectores de Santiago.

Para efectos de comparación se formaron dos grupos de control. El primero que fué constituido por 40 niños de ambos sexos que recibieron como única inter- vención una evaluación trimestral de su desarrollo psicomotor. Este grupo fue extraído al azar de la muestra indicada más arriba. El segundo grupo lo forma ron niños de otra muestra, utilizado en la estandarización de la Escala de Eva luación del Desarrollo Psicomotor. Eran niños de nivel socio-económico medio-alto no sometidos al curriculum de estimulación.

La comparación de los promedios trimestrales de desarrollo, mediante la esta dística "t", validó la hipótesis en el sentido que el curriculum de estimula ción iniciado desde el nacimiento influyó positivamente sobre el desarrollo psíquico de niños de nivel socio-económico bajo, permitiéndoles alcanzar el mismo rendimiento que un grupo de nivel socio-económico medio alto, ajeno al Programa.

Los autores terminan su trabajo examinando diversos aspectos analizados, co mo la diferenciación de resultados según características de los niños, facto res del programa que tuvieron mayor influencia, etc. Por otro lado, si bién consideraron validado el programa a nivel experiemntal tienen dudas, en tér minos de costos, en relación con su extensión masiva.

(Hernán Montenegro; S. Rodríguez; M.I. Lira y otros, Programa piloto de estimulación precoz para niños de nivel socioeconómico bajo, entre 0 y 2 años. Informe Final". Santiago: Servicio Nacional de Salud, 1977).

b) Diseño con grupo de control y mediciones "después"

En forma similar al diseño experimental clásico, en éste se forman también dos grupos, uno experimental y otro de control, cuyas personas son asignadas, a cada uno de ellos, en forma aleatoria. El grupo experimental recibe el tratamiento o las acciones del programa; el efecto se mide por la comparación de las mediciones "después" que se hacen en cada grupo. En el esquema que se presenta, se supone que ese efecto estará dado por la diferencia $O_1 - O_2$:

E : X O_1

C : O_2

Por usar un grupo de control, el diseño controla la "historia" y la "maduración"; por la asignación al azar de las personas, controla la "selección" y como no se hacen mediciones "antes", se controla el posible efecto de éstas sobre las mediciones "después"

Dentro de su sencillez, este diseño puede asegurar al investigador que lo

usa que las modificaciones que encuentra se deben al estímulo experimental, es decir, según el tema nuestro, a los efectos del programa sobre sus usuarios.

Diseños cuasi-experimentales

los diseños cuasi-experimentales, denominados así por Campbell y Stanley, toman en cuenta la dificultad de aplicación de los diseños experimentales en sentido estricto, en particular, la formación de grupos de control y la asignación al azar de las personas tanto a esos grupos como a los grupos experimentales. El no cumplimiento de estas condiciones afectan diversos aspectos de la validez interna y externa de las cuales el evaluador debe tener plena conciencia al analizar e interpretar los resultados de sus investigaciones.

Existen diversos diseños cuasi-experimentales de los cuales creemos de mayor aplicación en la investigación evaluativa los ~~tres~~ siguientes: a) diseño con un grupo de control no equivalente; y b) diseño "parchado" y c) diseño de series de tiempo.

a) Diseño con un grupo de control no equivalente.

A diferencia del diseño clásico, este diseño utiliza un grupo de control que no proviene del universo del grupo experimental o, en nuestro caso, del grupo que constituye la población usuaria del programa. En otros términos, esto quie

re decir que la asignación de las personas a los grupos no se ha hecho al azar. Por otro lado, como en el diseño experimental, en ambos grupos se hacen mediciones "antes" y "después" para comparar las diferencias finales y evaluar así, si lo hubo, el efecto del programa, Campell y Stanley representan este diseño en la forma que sigue:

$$\begin{array}{rcccl} E & : & O_1 & \times & O_2 \\ & & \text{-----} & & \\ C & : & O_1 & & O_2 \end{array}$$

La línea quebrada indica que los grupos no han sido formados al azar.

La posibilidad de atribuir con fundamentos el efecto del programa en la característica que se quiere lograr o modificar en la población de referencia se basa, en buena medida, en la similitud que tengan los grupos experimental y de control de modo tal que puedan ser considerados como muestras de un mismo universo. De ahí que una de las tareas principales de este diseño consiste en formar un grupo de comparación adecuado. Sin perjuicio de aplicar otras formas, se pueden sugerir algunas soluciones que tienen cierto grado de aplicabilidad general:

- a) Una de estas soluciones consiste en elegir un grupo de personas cuyas características se controlan por igualación con las de aquellas que recibirán el programa. Las características usuales de control son: sexo, nivel socio-

económico, escolaridad, etc.

b) Una segunda solución se basa en tomar como grupo de comparación un conjunto de personas que, inscritas para participar en un programa, se retiraron antes de recibir las acciones del mismo o sólo las experimentaron en baja proporción.

c) Una tercera alternativa la constituye la oportunidad de utilizar como control un grupo de personas semejantes sometidas a un programa con iguales objetivos, pero que utilizan distintas estrategias o procedimientos instrumentales. En tal situación, la evaluación mide la efectividad del programa estudiado y, en comparación con el otro, permite evaluar el mayor o menor éxito de su operación.

Ejemplo de aplicación

La ejemplificación del diseño cuasi-experimental presentado anteriormente la haremos con un resumen de la validación de un programa de no escolarizado de educación inicial realizado en Lima y Cuzco, Perú, entre los años de 1977 a 1979. El programa propiamente tal, basado en el llamado modelo Portage (S. Blume y A. Shearer, Portage guide to early education, Portage, Wisconsin: Cooperative Educational Service Agency, N° 12, 1976) se propuso desarrollar las destrezas que los niños emplean en aprender durante la educación inicial. Para lograr ese objetivo central se elaboró un currículum y una metodología

de trabajo con base en los padres de familia.

El curriculum atendió a ~~tres~~ requerimientos: en primer lugar, responder a las orientaciones socio-políticas de la Reforma Educativa Peruana y a los objetivos definidos en el país ~~para~~ la educación inicial (desarrollo integral del niño, con atención afectiva, biopsicomotor, intelectual y socio-emocional; etc.) En segundo lugar, el curriculum incorporó objetivos del Modelo Portage y, finalmente, consideró diversas características de la población a la cual se dirigía el programa.

La metodología empleada es similar, en general, a la propuesta en la Guía Portage, y comprendió las siguientes acciones: a) Evaluación por parte de los animadores de los niños de la experimentación de acuerdo con los objetivos del curriculum; b) Programación curricular: Con posterioridad a la evaluación la animadora selecciona ~~ba~~ de la lista de objetivos las actividades a ser desarrolladas por el niño. Se empezaba con un objetivo por semana y gradualmente se aumentaba la cantidad hasta llegar a la meta de tres objetivos semanales; c) Proceso de enseñanza en el hogar: iniciado el programa, la primera acción de la animadora durante su visita al hogar consistía en repasar con la madre y el niño las actividades que dejó la semana anterior y en el registro del logro o no logro de los objetivos. Luego presentaba las nuevas actividades enseñando a la madre (o al padre) la forma de realizarlas, el refuerzo correspondiente y la manera de ayudarlo si experimenta ^{ba} dificultades, d) Reuniones semanales: semanalmente, las animadoras se reunían con una coordinadora con

la finalidad de revisar las tablas de actividades, discutir las dificultades, recibir nuevas orientaciones, etc. La investigación evaluativa se propuso tres objetivos:

1. Determinar el grado en que el programa experimental logró sus metas de mejoramiento de los niveles de desarrollo de los niños de diferentes contextos socio-ecológicos y culturales con la directa y activa participación de los padres.
2. Determinar la factibilidad de implementar el programa en áreas marginales -rurales y urbanas- para conocer si era posible un buen funcionamiento con el personal disponible.
3. Estudiar los costos del programa e identificar sus beneficios con el fin de establecer la factibilidad económica de su aplicación en gran escala.

La validación se realiza en dos barrios marginales de Lima-que es la que seguiremos en este resumen - y en seis áreas rurales del Cuzco.

En Lima se tomó una muestra de 283 niños de 3 y 4 años de edad con los cuales se conformaron dos grupos experimentales y dos grupos de control, en cada barrio, equiparados en cuanto a la edad, nivel socio-económico y sexo. Como no se pudo hacer una distribución al azar entre los dos tipos de grupos, se utilizó un diseño cuasi-experimental (grupo ~~experimental~~ y grupo de control no equivalente).

En las mediciones "antes" y "después" se utilizó la escala Mc Carthy de Habilidades para niños, que explora las habilidades del niño en seis áreas: verbal, perceptivo-motriz, cuantitativa, cognitivos general, memoria y motora.

También se utilizó un instrumental basado directamente en los objetivos curriculares del Programa experimental.

Los resultados de las mediciones de logro (para el primer objetivo, único que tratamos aquí) se compararen así: a) los datos de la etapa pre-test entre grupos experimental b) los datos del post-test entre ambos grupos y c) el cambio entre grupos experimental y de control. En estas comparaciones se utilizó el análisis de la varianza. (estadística F).

Las principales conclusiones fueron:

- En el rendimiento intelectual global, los resultados obtenidos en la escala Mc Carthy (que funciona como criterio externo), indican que al término del experimento los niños del grupo experimental muestran un nivel de desarrollo significativamente más alto que los niños del grupo de control.
- En relación a las áreas exploradas en el Mc Carthy, el área perceptivo-motriz es la que muestra significación estadística favorable para el grupo experimental.

En la validación de la hipótesis con el otro instrumento (criterio interno)

se utilizaron sub-muestras de los grupos experimentales y de control en los cuales se hicieron solo mediciones después. Aquí se encontraron diferencias significativas a favor del grupo experimental a nivel de desarrollo global y en cada una de las tres áreas: biopsicomotora, intelectual y socio-emocional.

(María Alcántara, Programa no escolarizado de educación inicial con base en el hogar para la atención de niños de 3 a 5 años, Lima:, Instituto Nacional de Investigación y Desarrollo de la Educación, 1982.)

b) Diseño "parchado" (patch up)

Es un diseño que trabaja con dos o más grupos que se pueden colocar en una secuencia temporal (niños que entraron antes con niños que entraron después; niños que cumplen una cierta edad con niños que tienen la edad que acaban de cumplir los primeros; etc). La característica del diseño es la utilización de las mediciones "después" de un grupo con las mediciones "antes" del siguiente; las mediciones "después" de éste con las mediciones "antes" del siguiente y así sucesivamente. Este esquema se representa así:

Grupo	A	:	X	O_1	-----		
Grupo	B	:		O_2	X	O_3	-----
Grupo	C	:				O_4	X O_5

(Puede haber más grupos).

La medición O_1 , hecha después de la aplicación del programa, da el efecto de ést (si se produjo); la diferencia entre O_1 y O_2 indica ese efecto menos las modificaciones sufridas por los sujetos debido a los factores de "maduración" e "historia"..De nuevo, la diferencia entre O_2 y O_3 de el efecto del programa y la diferencia entre O_3 y O_4 , los efectos de "maduración e historia"; etc. Como se ve, los grupos (salvo el primero, en este caso), actúan a la vez como grupos experimentales y de control. (Sobre este diseño, ver Bruce W. Tuckman, Conducting educational research, New York: Harcourt Brace Jovanovich. 1978, pp. 146 - 147).

Cabe hacer notar que este diseño puede ser aplicado en la validación de programas en marcha, cuando es posible establecer grupos en secuencia. Esta y las otras características que hemos presentado pueden verse en el siguiente ejemplo.

Ejemplo de aplicación

El ejemplo está constituido por el Programa Experimental de Estimulación Temprana en Base a la Familia, de Lima, Perú. del cual damos aquí un resumen muy escueto con el fin de ejemplizar el diseño de validación dado más arriba. El programa fué elaborado en 1981 y puesto en marcha como experiencia piloto, sometido a control experimental, por un período de seis meses calendario (julio a diciembre de 1982) en dos unidades muestrales ubicadas en el Pueblo Joven Cooperativa Las Chincas de Andahuaylas en el Distrito de El Agustino de la pro

vincia del mismo nombre (ambos barrios marginales). Los usuarios estuvieron formados por niños menores de tres años y por sus madres o tutoras. Conformaron el personal ejecutor de las acciones directas dos grupos de animadoras, jóvenes mujeres no profesionales, que fueron capacitadas para este fin y que contaron con la coordinación y participación de personal docente, especialistas en educación inicial de la respectiva Zona Educativa de las que dependían administrativamente las dos unidades de estudio.

La validación se refirió especialmente a la siguiente hipótesis, basada en el objetivo central del Programa:

" Los niños sujetos al Programa de Estimulación Temprana (grupo experimental) mostrarán ganancias significativas en su desarrollo integral en sus áreas bio-psicomotoras, intelectuales y socioemocionales, en comparación a los logros obtenidos por grupos control. Esas ganancias se medirán por el Test de Denver - Examen de Desarrollo del Niño.

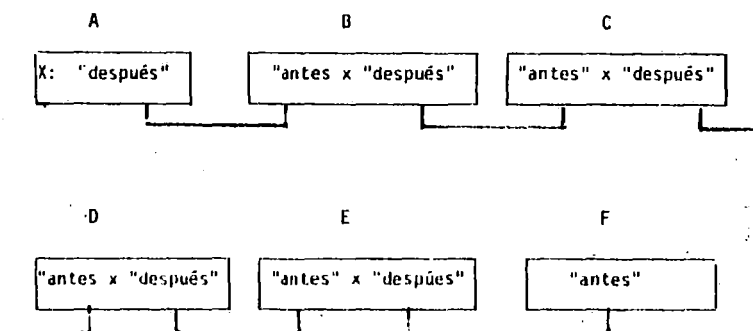
Para efectos de identificar el trabajo de estimulación se definieron objetivos en cada una de las áreas biopsicomotora, intelectual y socioemocional, para cada uno de los grupos de edad que se indican más abajo, cada uno con sus respectivos indicadores. La estrategia general consistió en la interacción de madre-hijo o padre-hijo en el propio hogar, interacción que fue enseñada por animadores especialmente entrenados. Las demostraciones en el hogar duraron una hora aproximadamente. Como materiales de instrucción se utilizaron fichas de

actividades (con alternativas de acción para cada uno de los objetivos de estimulación) y un manual de nutrición, higiene y cuidados permanentes. A nivel de los monitores se utilizó también una tabla de control de avance de los niños.

En el caso de Lima (que seguiremos en esta exposición) se conformaron 12 grupos de niños con edades en secuencia, tomados de un censo realizado en la comunidad en 1981. Los grupos según las edades y clasificados en dos niveles fueron: (se presenta sólo el primer nivel):

<u>Grupo</u>	<u>Edad cronológica promedio relativa al período transcurrido entre el pre-test y post-test</u>
A (6 casos)	15 días a 7 meses
B (6 casos)	7 meses a 13 meses
C (5 casos)	13 meses a 19 meses
D (8 casos)	19 meses a 25 meses
E (4 casos)	25 meses a 31 meses
F (6 casos)	31 meses a 37 meses

El diseño de validación elegido fue el de "diseño parchado", con mediciones "antes" y "después" realizadas en los siete grupos de la forma siguiente:



El grupo A se considera como estrictamente experimental y el F, como estrictamente de control. Los otros son experimentales y de control a la vez, ya que el B, por ejemplo, con su medición "antes" (cuya edad de inicio es la misma que la edad terminal de A) controla la maduración de A; la medición "antes" de C, controla el efecto maduración de B; etc. Realizado este control, el efecto del programa lo dan las composiciones "antes" y "después" dentro de cada grupo. Esas composiciones fueron tratadas con la estadística "t", con un nivel de significación del 0,05.

Los autores terminan la validación del Programa de Lima en esta forma:

" Resumiendo, de los resultados de los doce grupos muestrales conformados según edades, sólo en una de las contrastaciones, entre el grupo experimental de niños de 30 meses de edad y el grupo de control respectivo, arroja una diferencia significativa al 0,01 en el performance general a favor del grupo que ha recibido el Programa de Estimulación Temprana con Base en la Familia. Sin embargo, es necesario señalar que se producen diferencias con valor negativo - aunque no significativas - en tres de las composiciones que se realizaron con niños de las edades de 19, 24 y 31 meses. Lo cual estaría señalando que los grupos controles, que no habían recibido el Programa, rindieron en ese momento ligeramente mejor que los grupos experimentales".

(Flor de María Espinoza y Luis Alfredo Inga, Evaluación Sumativa del Programa Experimental de Estimulación Temprana con Base en la Familia, Lima: Insti

tuto Nacional de Investigación y Desarrollo de la Educación - Ministerio de Educación - 1983)

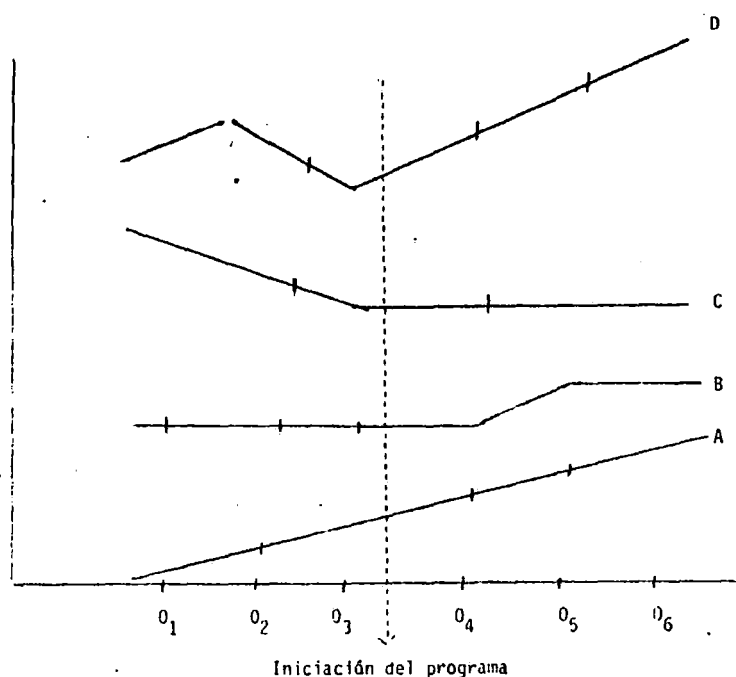
Los resultados del ejemplo muestran con claridad la necesidad de validar un programa antes de su extensión masiva; ¿que falló en este caso? ¿La teoría de la estimulación utilizada? ¿la metodología? ¿El trabajo de los monitores? Estas y otras preguntas podrían formularse y deberían contestarse antes de continuar con un programa así. Todo esto, además de revisar el diseño que se empleó en su validación.

c) Diseño de series de tiempo

Este diseño cuasi-experimental consiste en una serie de descripciones y mediciones de la población objeto realizadas unas antes y otras después que el programa haya comenzado a funcionar. El efecto del programa, si se da, se revelará en la diferencia entre las mediciones "antes" y las mediciones "después". El gráfico N° 2 presenta algunas situaciones posibles de darse antes y después de la aplicación del programa.

Gráfico N° 2

Resultados posibles en un diseño de series de tiempo



La situación A indicaría una tendencia presente en la población en la característica de evaluación y, por lo tanto, la diferencia entre las mediciones anteriores y posteriores no se debería a la acción del programa; la B señala una línea de elevación que luego se estabiliza; la situación C significaría una disminución que luego se estabiliza en un cierto nivel; finalmente la si

tuación D, en estos ejemplos, indicaría un efecto positivo del programa.

El diseño de series de tiempo puede ser especialmente útil cuando, por diversas circunstancias, no es posible aplicar otro modelo en situaciones en que la variable de evaluación es un proceso que puede ser medida periódicamente. En ciertos casos, estadísticas secundarias confiables (sobre niveles de ingresos; productividad, etc.) proporcionan datos que corresponden a las mediciones "antes" y ayudarían a la validación del programa.

Validez ecológica del programa

Así como la validación experimental o cuasi experimental de un programa demostrativo tiene como objetivo verificar su efectividad, los investigadores ocupados de esta tarea deben preocuparse por la validez ecológica de los resultados obtenidos en la validación interna. Esa validez-que forma parte de la validez externa de un experimento - se refiere a la capacidad del programa de producir en la población de usuarios definitivos, los mismos o similares formas de resultados y de funcionamiento que se dieron en la validación experimental.

Esa validez ecológica se asegura ... cuanto mayor sea el control de los factores que inciden en la validez interna, pero además, en cuanto la población de usuarios sea lo más semejante posible en sus características individuales y de contexto en el cual viven a las que se dieron en la validación del programa.

ma. Campbell y Stanley (op. cit. pp. 17 - 18) señalan, a su vez, cuatro factores que amenazan la validez externa o representatividad de un diseño; de los cuales destacamos éstos:

1. El efecto reactivo o de interacción de las pruebas. Se daría en cuanto en la experimentación del programa, la aplicación "antes" de las pruebas podría sensibilizar a las personas respecto a las aplicaciones "después". Este efecto no se daría en la aplicación masiva del programa en cuanto no se aplicaran esas pruebas "antes".
2. Efectos reactivos de los dispositivos experimentales. Hace referencia a la diferencia que existe entre el ambiente experimental de la validación que por lo novedoso ~~puede~~ de llevar a algunas personas a rendir positivamente, mientras a otras ~~puede~~ ocurrirles lo contrario. Lo cierto que existen diferencias de varios tipos entre una situación experimental y una situación más natural en la cual se aplicará el programa masivamente.

A estos y otros factores podríamos agregar el efecto Hawthorne que resulta cuando un grupo de personas se da cuenta que está sirviendo en una experimentación y de lo que desean los ejecutores del programa y por ello proporciona la información que cree adecuada a las expectativas de estos últimos.

CAPITULO 5: DISEÑOS DE EVALUACION I.

Al presentar los modelos de evaluación, dijimos que ellos 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 combinaciones de ellos 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 de tal programa. Justamente llamamos diseño de evaluación al plan elaborado para adecuarse a esas exigencias, es decir, al plan que define las diversas estra-

tegias y procedimientos para llevar a cabo la evaluación del programa.

De manera típica, el diseño de evaluación comprende los siguientes componentes:

1. Antecedentes y descripción del programa.
2. Objetivos generales y específicos del programa.
3. Finalidades de la evaluación.
4. Objetivos de evaluación.
5. Diseño metodológico.
6. Informes.

En cuanto los numerales 1 y 2 de este diseño se limitan a reproducir las características principales del programa y el tema de finalidades de la evaluación ya ~~foe~~ fue suficientemente tratado en el primer capítulo de este libro, 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 necesita 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 averigue "si el programa funciona bien o mal", objetivo de conocimiento que puede significar varias cosas, 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 les presta el programa, etc.

En el problema reseñado más arriba, puede suceder lo que señalaba Hyman 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). Decía así el metodólogo citado, al referirse al trabajo de otros dos autores: "Notan, por ejemplo, que el cliente a menudo no formula su problema en términos que puedan ser reducidos 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, Glencoe, Ill.: The Free Press, 1955, pp. 44-45).

Una dificultad adicional en la formulación de los objetivos de evaluación que, insisto, son objetivos de conocimientos y, por lo tanto, en relación con los cuales se recogerá y analizará la información, consiste que tanto éstos como los propósitos que están detrás de ellos varían en un mismo programa según el

nivel o 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 tiene que ver con la ampliación, la reducción o la modificación del programa, que para los ejecutores directos del programa que podrán estar interesados en saber qué opinan los beneficiarios de la metodología empleada, o del uso de los recursos instrumentales, etc.

Objetivos descriptivos

Cuando la evaluación es de tipo descriptivo los objetivos de evaluación se derivan directamente ~~de~~ los objetivos del programa o/y por los otros componentes que se han tomado como focos u objetos de evaluación. En este caso, los objetivos suelen redactarse así:

- a) Los objetivos de la evaluación son los siguientes (aquí se refieren directamente a los objetivos del programa):
 1. Evaluar el efecto sobre los beneficiarios de las inversiones terminadas 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 a 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 microregión, comparando las previsiones con las realizaciones. (Michael M. Cernea, La cuantificación de los efectos de los proyectos: la vigilancia y la evaluación en el Proyecto de Desarrollo Rural PIDER-México, Washington: Banco Mundial, Documento de Trabajo N° 332, 1979; pp. 64-65.

b) La evaluación se propone evaluar los siguientes componentes y características del programa (se nombran los objetos y variables de evaluación):

- Funcionamiento administrativo: cumplimiento de funciones; calificación del personal, etc.
- Resultados: niveles de logro de los objetivos; calidad; distribución entre los usuarios; etc.

Los objetivos del programa, en este caso, 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.

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), como, en algunos casos, descripción de la calidad y distribución diferencial de esos logros; y 3) comparación de esos logros cuantitativos y cualitativos con alguna norma o meta, criterio deseado, con el fin de pronunciarse sobre el "éxito" o "fracaso" del programa.

- c) En evaluaciones que utilizan diseños experimentales o cuasi-experimentales, los objetivos de evaluación se redactan, con frecuencia, 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 de hemos presentado, lo conveniente sería - por las razones que dimos en otro capítulo del texto - 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 objetivos curriculares de las tres áreas de desarrollo"; respecto del trabajo de los ejecutores directos del programa:

"Las animadoras tendrán un desempeño del 70% de logro en el trabajo metodológico que realizan con las familias".

En fin, para seguir con el ejemplo, en la misma evaluación se propuso una escala valorativa porcentual para la interpretación de porcentajes referidos a las características del personal ejecutor, manejo de instrumentos, etc.) con la siguiente estructura:

Valor	Intervalo
- Alto	70 a 100 %
- Mediano	50 a 69 %
- Bajo	30 a 49 %
- Muy bajo	0 a 29 %

(María Alcántara y Cipriano Olivera, Evaluación formativa del Programa de Estimulación Temprana con 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" o den cuenta de 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 en el concepto de eficiencia interna, deberíamos luego buscar factores o variables que podrían estar conectadas en la eficiencia para formular algunas hipótesis pertinentes. Demás está decir que en tal búsqueda se manifiesta la formación, experiencia e imaginación del evaluador.

Algunos ejemplos podrían ayudar a clarificar la situación que se presenta en la determinación de objetivos de evaluación y la búsqueda de factores explicativos:

- 1) "Cabe preguntarse si un mal desarrollo del programa tiene por causa ciertas deficiencias en la organización de los cursos o 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 socio-profesional), 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 Cou-

vert, La evaluación de programas de alfabetización, París: UNESCO, 1979; pág. 118).

- 2) "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 objetivos particulares (cada uno implica hipótesis sobre factores que inciden sobre las características a 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 asiduaad escolar.
 - c) 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".

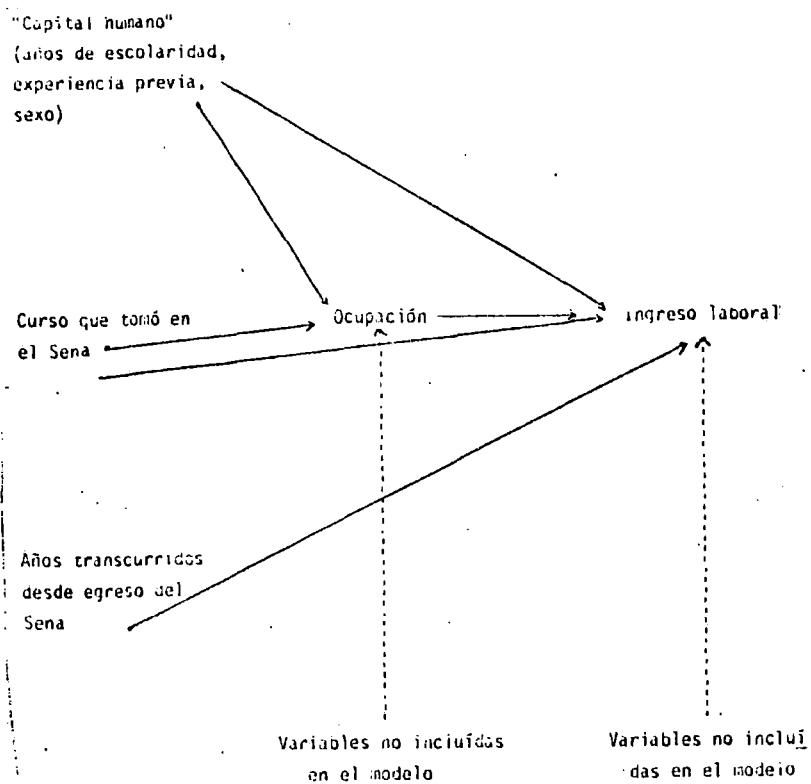
(Sylvia Schmelkes, "Investigaciones sobre las escuelas radiofónicas de la Tarahumara", Revista del Centro de Estudios Educativos, Mexico, vol. 1, N° 3, pág. 128).

Como es fácil de ver en los ejemplos anteriores, los objetivos de evaluación contienen relaciones de determinación entre la variable de evaluación (el desarrollo del currículum, 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, etc).

En algunas evaluaciones explicativas, los objetivos de evaluación contiene 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 por 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): Gráfico N° 3.

Gráfico N° 3

Modelo causal para explicar los ingresos laborales



En resumen: la formulación de objetivos en evaluaciones explicativas toma formas como las que se indican:

1. La evaluación se propone determinar los factores que facilitan (o dificultan):
 - 1.1. El logro de los objetivos, en cuanto a su nivel y calidad;
 - 1.2. el funcionamiento del programa;
 - 1.3. la aplicación de su metodología, etc.
2. La evaluación se propone determinar (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 daría 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 de pendientes e independientes como el dado en el gráfico recién presentado).

Cuando en las formulaciones anteriores se mencionan "factores que facilitan o dificultan" tales o cuales resultados a la marcha del programa, debe entenderse que tales factores pueden corresponder a indicadores objetivos o subjetivos (por ejemplo, en este último caso, opiniones de los usuarios; actitudes del grupo técnico; etc).

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 que se necesita para satisfacer los objetivos de evaluación. De esta manera, a nivel general, el diseño consiste en los procedimientos que se utilizarán para resolver los 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 princi-

pal de la investigación evaluativa consiste, paralelamente, en establecer si en los hechos el programa, con la metodología empleada, alcanza o no el o los objetivos propuestos. Se trata, como se ve, de verificar la efectividad del programa.

Ahora bien. Puede suceder que tal efectividad haya sido comprobada en la validación del programa. La tarea del evaluador, en este caso, 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 diseños de validación que presentamos antes (experimentales y cuasi experimentales) o en otros procedimientos distintos pero que cumplen con la misma finalidad.

a) 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 por inferencia el evaluador pueda pronunciarse acerca de la efectividad de éste. En el ejemplo que utilizamos aquí, 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, N° 2, dic. 1982, pp. 775 - 788).

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 la sala de clases. En una sesión, que dura unos 15 a 30 minutos, el profesor guía a un grupo de 5 - 12 estudiantes en la discusión de un tema particular. El profesor emplea métodos que estimulan la expresión por parte de los estudiantes de sus ideas y sentimientos. La participación es siempre voluntaria y no se permite la confrontación y el debate. El elemento clave del Círculo es que cada alumno tenga una oportunidad de contribuir - aún cuando él o ella prefiera permanecer en silencio; por otro lado, cada participante es escuchado cuidadosamente por los otros miembros del Círculo.

Para la evaluación fueron procesadas 13 escuelas elementales públicas, predo-

minantemente de clase media, del Norte de Carolina, sobre la base de las características de sus estudiantes, profesores, etc. Una por una de las escuelas fueron asignadas al azar a la condición experimental y a la condición de control. En definitivas, el grupo experimental estuvo formado por 102 niños y 115 niñas y el control, por 131 niños y 119 niñas, todos del tercer grado.

Para evaluar la implementación 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 sobre 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, otra vez, al final del año escolar para vigilar la implementación y valorar la adecuación del entrenamiento; 6) observaciones de clase por un profesor especialmente entrenado de los profesores mientras 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 cualquiera difusión del entrenamiento en la estrategia del programa.

La situación de los estudiantes en los grupos experimental y de control, en las variables a evaluar, fueron medidas, antes y después de iniciarse el Programa, mediante escalas de auto-observación referidas a relaciones con sus compañeros, sus profesores y sus escuelas. En la prueba "después" se midieron, también, la creencia en el auto-control, la responsabilidad en los procesos escolares y la auto-estimación.

El análisis de los datos reveló que el Círculo Mágico fue una experiencia positiva para los profesores experimentales. Les agradó el entrenamiento y el realizar las sesiones de discusión con sus estudiantes. Comparado con los profesores de control, los profesores experimentales mostraron un mayor incremento en su satisfacción personal con la enseñanza.

A nivel de 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 auto-estimación 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 antes, los investigadores sugieren dos posibles conclusiones: o el efecto del

Círculo Mágico en los estudiantes es idiosincrático o bien es espúreo.

Problemas para la aplicación de diseños experimentales

Al tratar de la validación de programas demostrativos pudimos apreciar las bondades de algunos diseños experimentales y cuasi-experimentales en lo que se refiere al control de diversos factores que pueden reducir o anular su validez interna. La posibilidad de su uso queda bastante limitada, sin embargo, 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 iniciado. No hay posibilidad, por lo tanto, de constituir grupos experimentales y de control por asignación al azar como tampoco hacer mediciones "antes".
2. En aquellos casos en que la formación de tales grupos es posible (recuérdese el ejemplo del Círculo Mágico que dimos hace poco), no siempre es 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, etc).
3. También, en el caso de poder aplicar algún diseño experimental, 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 acciones del programa significan algún beneficio para la población: ¿por qué proporcionar servicio a esas personas y no a los otros?

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 bién es posible, en circunstancias especiales utilizar algunos tipos cuasi-experimentales como los que vimos anteriormente).

b) Diseño con medición sólo "después" con grupo de comparación

Este diseño trabaja con dos grupos: uno formado por la población objeto y otro formado por un grupo externo de comparación. Esto, con el fin de poder 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 las personas del otro grupo. La tarea principal en este diseño es lograr un grupo de comparación lo más parecido 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 con 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 le permite al investigador controlar en alguna 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 son equivalentes, en cuanto no han sido formados al azar, puede estar presente el factor de "selección" con lo cual tampoco se puede estar seguro de si un grupo era diferente del otro en la variable dependiente antes de aplicar el programa con 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 lo dice su nombre, de perfeccionar en forma masiva a profesores de ciencias naturales sin que abandonen sus lugares de trabajo. En 1971 el PPS tenía unos 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 encargados del mismo:

1. Material de auto - instrucción.

Contiene los aspectos teóricos de cada uno de los procesos científicos a desarrollar en el aula, de acuerdo con el programa de su curso o nivel (observar, medir, etc).

2. Comentarios sobre los niveles del proceso.

Desarrolla y comenta 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.

Comprende una serie de actividades específicas que debe realizar el profesor 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.

Plantea 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 de evaluación que permiten ponderar aspectos cuantitativos del trabajo del profesor, en particular, con el material de auto-instrucción. 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 pudieron haberse presentado.

La evaluación tuvo como objetivo central medir los resultados logrados en los procesos de observar, medir, comunicar y clasificar tanto a nivel de los profesores como de los estudiantes que en 1971 cursaron 5° año Básico. El estudio se realizó en marzo de 1972.

Como 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 solo 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 comparación por 28. Ambos grupos presentaron características demográficas y 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 consistió en realizarlos efectivamente para lo cual se proporcionó a cada profesor los elementos necesarios (un globo, una balanza, una re-

gla, 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. Hubo también la aplicación de un cuestionario a los profesores sobre satisfacción en el trabajo, contenidos de los programas de ciencias naturales, opiniones sobre el PPS, etc. Los alumnos, por su parte, contestaron 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 procesos 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 evalúa con la norma deseada.
3. La evaluación del PPS respecto de una norma muestra que los logros totales en los cuatro procesos científicos quedan bajo esa norma: promedio de 49 puntos contra un valor esperados de 58 - 60 puntos (la norma fue propuesta por el equipo PPS).
4. Los alumnos con profesores PPS logran, en promedio, mayores puntajes 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, pp. 53 - 102)

c) 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 sub-grupos 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 debe tenerse presente que cuando un programa actúa en un contexto o población en la cual no actúan 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 dice:

"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 estar seguro 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 "historia", 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 a los que propocionan los diseños experimentales y cuasi-experimentales.

d) Diseño con medición solo "después"

Como lo indica su nombre, es un diseño que, como 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, solo puede disponer de una sola medición, la medición "después", realizada en algún momento de la marcha del programa o justamente a su término. Por estas características, el diseño está abierto a la mayoría o 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 como consecuencia de la selección sesgada de tales usuarios o por la maduración biopsicológica de los mismos.

Por las razones dadas, 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 por esta técnica pueden ser cuantificados y empleados en la búsqueda de conexiones causales entre los resultados del programa y características de sus componen

tes, mediante el análisis de senderos (path analysis), por ejemplo. Por otro lado, la información a recoger puede cualificarse y ampliarse mediante entrevistas en profundidad, preguntas abiertas, etc.

Cuando el problema de la validez interna del programa tiene especial relevancia, es posible pronunciarse sobre ella con un manejo analítico bastante simple de los datos, como lo es el cruce de los resultados con el grado de incidencia del programa en distintas personas de la población objeto. Por ejemplo, la asistencia a reuniones del programa puede tomarse como uno de los indicadores de la acción del mismo y de esta manera conviene elaborar cuadros del siguiente tipo:

		<u>Asistencia</u>	
		<u>Alta</u>	<u>Baja</u>
<u>Logros</u>	Altos	p_1	p_2
	Bajos	q_1	q_2
		100 %	100 %

Si el programa fuera efectivo, es decir, produjera las modificaciones buscadas en los objetivos, entonces debería tenerse que " p_1 " es mayor que " p_2 ".

Ejemplos de aplicación

Ejemplificaremos el uso de diseños no experimentales 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, por que 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 similar en varios aspectos al que estamos exponiendo en este capítulo. Así se justifica, además, la extensión que le hemos dado a su resumen.

a) 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 el impacto de los Centros de Salud Mental Comunitaria en Nueva York. El acta federal de 1963, creadora del programa, contenía varios objetivos de la profesión psiquiátrica: proporcionar alternativas para atender en hospitales grandes y distantes; proporcionar atención en las comunidades de residencia con algunos recursos; dar un amplio conjunto de alternativas 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 a las comunidades en lugar que a los individuos, con la participación de ellas en las decisiones.

Estudios previos habían explorado extensamente el impacto de las caracterís-

ticas individuales tanto en el número de personas que recibían los servicios como los resultados obtenidos. Según los autores de la evaluación, ~~era~~ conveniente un enfoque diferente para medir los efectos al nivel de 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 status socio-económico, 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 ser empíricamente contrastados. 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 fue combinado en un modelo causal (path diagram).

De acuerdo con el enfoque distintivo del diseño, la unidad de análisis estuvo constituido por "un área de salud". De las fichas del censo de población de 1970 se tomó una muestra de 333 de tales unidades, en cada una de las cuales se tomó 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:

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 al 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.) Handbook of Evaluation Research, Beverly Hills, CA: Sage Publication, 1975, 341 - 429).

b) 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, a nivel nacional, denominado Acción Cultural Popular -ACPO- que actúa en Colombia desde 1947, 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:

- Radio Sutatenza. Es una radio del programa que transmite cursos de educación básica 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 concentran en los 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 6 horas diarias de transmisión con 4 cursos que, en conjunto, desarrollan los objetivos de ACPO.

- 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.

- 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 de las cartillas. Incluyen temas como cultivo de árboles frutales, piscicultura y pesca, cría de cerdos, apicultura, ganadería, primeros auxilios, salud y nutrición, enfermedades contagiosas, sexo y matrimonio, cuentos de niños, cómo tocar la guitarra, derechos cívicos, etc.

En la práctica es común que la colección de libros de la biblioteca se encuentren situados en una o más casas de la comunidad para su mejor utilización por parte de los campesinos.

- 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 campesino.
- La correspondencia. Son cartas enviadas por los campesinos a las oficinas centrales de ACPQ en la cual expresan sus problemas y necesidades y solicitan ayuda. En 1978 se recibieron y fueron contestadas más de 28.000 cartas de este tipo.
- Los cursos de extensión. Son cursos de corta duración, a veces de un día, desarrollado por un agente interpersonal de ACPQ. Tratan problemas y necesidades locales específicos (cooperativas, utilización correcta del crédito, etc).
- 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 a nivel local por agentes interpersonales de ACPQ. Entre esas campañas están las de nutrición, vivienda, suelo y recreación.
- 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 acciones del programa; juegan un papel muy importante dentro del sistema ACPQ.

- Los profesores como agentes educativos. Son profesores que trabajan en el centro de ACPO 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.
- Agentes educativos de campo. Son las personas (más de 20.000 en 1979) que trabajan en diversas formas de la enseñanza, casa a casa, con los campesinos. Están divididos en 1) auxiliares inmediatos o monitores de las escuelas radiofónicas; 2) dirigentes campesinos; 3) líderes (campesinos adultos que han terminado dos ciclos de cursos en uno de los Institutos Campesinos de ACPO); 4) supervisores (personas que tienen responsabilidades de administración y supervisión en relación con los demás trabajadores de campo a nivel regional).
- 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, ACPD 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 prác-tica de técnicas modernas de agricultura, cuidado de los animales y conservación del suelo.
4. La adquisición de conocimientos ya 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 contribuyen a una mejor rela-ción 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 relaciones humanas dentro de la familia.
11. El reconocimiento de las responsabilidades de la paternidad y la práctica de 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 sobre 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 enfoque en protección de

los derechos humanos y en la importancia de que todas las personas asu
men sus responsabilidades en la sociedad.

Frente a estos objetivos del programa, la evaluación se dirigió a obtener res
puesta a las siguientes preguntas u objetivos de evaluación:

1. ¿Qué tan efectivos han sido los programas de ACPO en mejorar el nivel de conocimientos en salud, agropecuario 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 un cambio de actitudes del campesino hacia la salud, la 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 bajo qué condiciones han estado las familias campesinas utilizando los medios de ACPO?
6. ¿Qué elemento o combinación de elementos ha tenido mayor impacto 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 explicar las diferencias en conocimientos, prácticas y actitudes entre las familias campesinas?
8. ¿En qué medida podría 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 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 in-

cluyó, además una descripción de las regiones geo-culturales en las cuales se realizó el estudio. En segundo término se hizo un análisis de tipo correlacional, con 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."

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 el área de salud.

En el área de agropecuarios se encontró 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 a 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 tener un desempeño rela-

tivamente alto, y que ésta relación persistía cuando los efectos de otras variables educativas del contexto eran controlados. Un resultado paradójico ~~que~~ 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 encontró que aunque buena parte de quienes respondieron la encuesta habían recibido su educación básica en la escuela formal, los más usuarios de ACPO tendían a obtener puntajes altos en la prueba utilizada para medir el desempeño de destrezas. Esto sugiere que aún en los casos en que ACPO no sea la fuente primaria para el aprendizaje de las destrezas en educación básica, sí cumple un rol importante en el mejoramiento y mantenimiento de esas destrezas.

Se encontró que los cinco componentes de ACPO estaban altamente interrelacionados de tal forma que fue imposible concluir que los componentes contribuyeran diferencialmente en los efectos observados. Un alto uso de los componentes de ACPO en conjunto, como se refleja en el índice de uso 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 ésta afirmación no es confirmada por el análisis estadístico, es consistente con el juicio del equipo del proyecto y las entrevistadores de campo" (pp. 10 - 11)

(Robert M. Morgan, Liliana Muhlman y Paul H. Masoner, Evaluación de un sistema de educación no formal a nivel nacional: Acción Cultural Popular, Bogotá: 1980)

CAPITULO 6: DISEÑOS DE EVALUACION

En las páginas anteriores nos referimos a distintos procedimientos que se utilizan en la verificación de la efectividad de un programa y dimos algunos ejemplos al respecto. Desde luego que al preparar el diseño metodológico solo hacemos una descripción general del procedimiento elegido el cual se desarrolla en la marcha misma de la investigación.

Algo similar ocurre con los otros componentes del mismo en el sentido que en la etapa preparatoria basta con referirse a ellos en 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 sub-grupos 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 a qué niveles se tratará de encon-

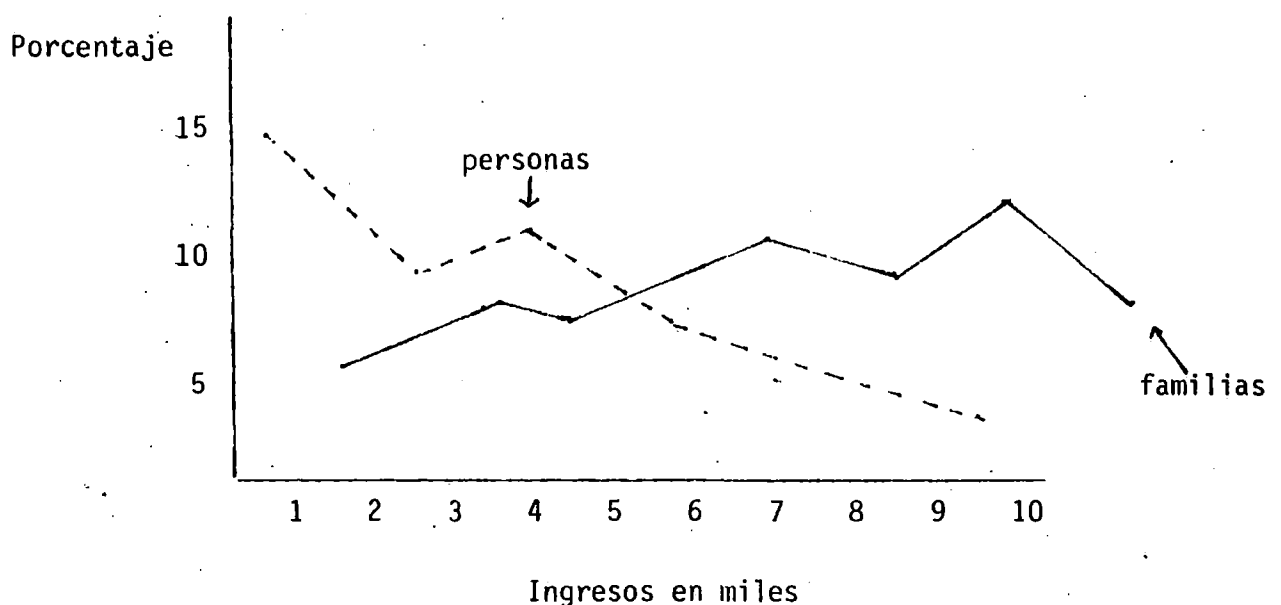
trar las confirmaciones del caso, pues como lo muestran objetivos de diversos o de un mismo programa, la orientación de las acciones instrumentales se dirige unas veces a personas individualmente consideradas, otras veces a grupos naturales (familias, cursos de un colegio, etc.), a categorías de personas (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 tomadas individualmente o a grupos de personas tomadas justamente como grupos? Es fácil de ver 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, creemos, podrá ayudarnos 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 to me a la empresa agrícola como la unidad fundamental de la agricultura campesina, sino la unidad familiar:

"sólo podemos abrigar 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 u nidad pequeña, pero compleja y dotada de un de

licado 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 dar 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:



Como se ve, los resultados son muy diferentes pero, por lo mismo, usados con

juntamente, pueden proporcionar informaciones muy útiles en la evaluación.

En resumen: 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 siguiente: "la mayoría de las personas de más edad tuvieron una alta participación en actividades comunitarias".

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 hacerse nuevas decisiones si existen varias de tales unidades: tomar una, tomarlas todas o seleccionar algunas del total. Esta última decisión envuelve dos alternativas principales: a) utilizar un criterio probabilístico; y b) seleccionar las unidades con 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 probabilísticos. Diremos algo de estos diseños.

Como sabemos, los diseños probabilísticos dan una probabilidad conocida, dis

dista de cero, a cada una de las personas del universo de estudio de entrar en la muestra. En algunas 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 darse el error de muestreo; 3) posibilidad de generalizar al universo los resultados obtenidos en el análisis de la muestra. (Para un tratamiento detallado de los diversos diseños probabilísticos, se puede consultar a G. Briones, op. cit.; capt. 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 en su conformación no es probabilístico, 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 intencional. Pero su principal característica consiste en posibilitar el hacer 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 cuasi-experimental y no experimentales.

1. Selección intencional de casos extremos

Puede que 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 diversas otras características 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 sub-grupos 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 eserción, altos costos, baja eficiencia administrativa, etc. El análisis de situaciones como éstas pueden 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 sub-muestra 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 sub-muestra probabilística de estudiantes o profesores en cada una de las escuelas seleccionadas intencionalmente. El diseño probabilístico puede ser del tipo aleatorio simple, sistemático, estratificado, etc.

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, etc).

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 aprendizaje, elevación del ingreso familiar; etc). También es frecuente, sin embargo, que otras variables necesiten definiciones para comprender su significado como también requieran de indicadores que hagan posible su descripción y, cuando sea posible o necesario, su medición. En este último caso, es necesario utilizar técnicas ya preparadas o que el evaluador debe desarrollar por su cuenta (pruebas, escalas, etc).

A veces una propiedad incluida en los objetivos de evaluación queda bien representada por uno o dos indicadores (por ejemplo, status 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 buena representación válida y confiable.

La tarea es aún complicada cuando la situación a evaluar 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 bien se utilizan características, que implican por ejemplo, variedad de recursos instrumentales empleados, contenidos de aprendizaje, distribución homogénea de los logros obtenidos, etc.

Los indicadores pueden referirse a variables individuales (actitudes, nivel socio-económico, 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 valoraciones 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.) existen listas de indicadores aceptados a nivel internacional 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 puede requerirse el uso de indicadores especiales propuestos por los evaluadores. Un ejemplo que cae en esta solución la encontramos 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 en la forma siguiente:

"Al nivel analítico de la sociedad en su conjunto, se tomará como criterio único 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; pp. 140-141).

También esta evaluación utiliza diversos indicadores subjetivos para evaluar otras características del sistema de aprendizaje, como el siguiente que ilustra la clasificación que hicimos hace poco:

"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 sobre el funcionamiento de la educación superior en Chile en 1981. En este caso, el indicador o índice se definió así:

Porcentaje de alumnos matriculados en la educación superior cuyos padres tienen las ocupaciones X,

Porcentajes 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 neo-liberal: 1974 - 1982, Santiago: PIIIE).

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 damos 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, etc).

4.1.4. Número de magnetófonos y número de cassettes poseídos y 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, consejos, etc).

4.2.2. Número de consejos pedidos a organismos de ayuda social (seguridad social, mutualidades, compañías de seguro, etc).

4.2.6. Número de viajes emprendidos por duración y naturaleza: razones familiares, emigración estacional; búsqueda de trabajo, razones religiosas, etc.

Criterio 3: Utilización social e informativo 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--161. En el texto se dan otros indicado

res 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 verlo en la evaluación del programa de Acción Cultural Popular - ACPD que presentamos hace poco. Por dar una idea de los elementos presentaremos algunos de los indicadores seleccionados en el área de salud:

I 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. pp. 197-201)

Una vez que se definen o eligen los indicadores éstos se traducen en preguntas o en indicaciones para la observación o su constatación en entrevistas o por su uso en otros procedimientos para la recolección de informaciones.

Recolección de informaciones e instrumentos

En el diseño metodológico de una evaluación debe indicarse los procedimientos que se utilizarán para la recolección de informaciones y el o los instrumentos *para el mismo fin.* En cuanto a procedimientos, podrían utilizarse

observaciones, entrevistas de todo tipo, discusiones grupales, etc. Los instrumentos, según el uso, podrán ser protocolos de observación (estructurados o no), guías de entrevistas, cuestionarios, registros de discusiones, etc.

En algunos casos existen instrumentos contruídos y validados por otros investigadores. Un caso así vimos en la evaluación del programa de estimulación temprana con base en la familia, en lo cual los investigadores utilizaron la prueba de Denver, de tipo cualitativo con 105 items, 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, como serán preparados esas personas, cuál será la distribución del trabajo, cómo será la supervisión, etc.

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 utilizaran, sean de tipo cualita

tivo o cuantitativo. Por ejemplo, en el primer caso podrán mencionarse tipologías cualitativas, conexiones entre procesos y entre procesos y resultados, etc. En el segundo, el evaluador podrá referirse a análisis relacional y nombrar técnicas como las de regresión, análisis de senderos, de redundancia, etc. O en diseños experimentales, técnicas como análisis de la variancia, covariancia, etc.

El informe de evaluación

El informe de evaluación da cuenta de la investigación realizada y en él se describen con mayores detalles los componentes del diseño, como también las dificultades que pudieron encontrarse en su aplicación real. 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 termina con sugerencias o recomendaciones basadas en los resultados obtenidos. Sobre este último punto -que no es compartido por todos los evaluadores- dice Weiss:

"La orientación académica lleva a veces a los evaluadores a no sacar conclusiones cuando informan de sus resultados. Según ellos, su trabajo consiste en hacer el estudio y analizar los datos y no en recomendar acciones. Pero... en la medida en la cual raramente los datos

son obvios, la abdicación de esta tarea por parte del evaluador a menudo significa que nadie la hace".

(C.H. Weiss, Investigación evaluativa, 1975, p.136).

Por otro lado, 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 ésta situación:

- Diferencias entre los roles y estilos del personal del programa y los evaluadores.
- Estudios que son insuficientemente rigurosos para ser convincentes.
- Demasiado 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 "pueden ser usados".

(Ray C. Oman and Stephen R. Chitwood, Management evaluation studies: Factors affecting the recommendations" Evaluation Review, vol. 8, number 3, June 1984, p. 286)

De manera típica, el informe de evaluación contiene secciones como éstas:

- Descripción del programa. (Estructura, objeti-

- vos, ubicación, etc).
- Contexto socioeconómico
- La evaluación:
 - Origen
 - propósitos
 - objetivos de evaluación
 - metodología
 - problemas en la marcha del programa
- Resultados
- 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 característica que se dan en muchos de los diseños, está 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 (auto-evaluación, evaluación grupal, etc.), todo esto con el uso de técnicas cualitativas, preferentemente, para la recolección y análisis de la información. Como se recordará, 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 diseño 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:
 - 4.1. Observación participante del personal encargado de la evaluación.
 - 4.2. Reuniones grupales para evaluar el funcionamiento del programa y resultados mediante discusiones y juicios valorativos.
 - 4.3. Entrevistas en profundidad a informantes seleccionados para profundizar temas importantes detectados en 4.1 y 4.2.
 - 4.4. Registro de experiencias personales de algunos usuarios e historias de vida.
5. Análisis de la información recogida:
 - 5.1. Imágenes verbales sobre el programa y sus componentes: Objetivos, necesidades, relaciones interpersonales, etc.
 - 5.2. Determinación de la significación cultural de las expresiones verbales de los usuarios.
 - 5.3. Conexiones entre procesos

5.4. Tipología de comportamientos

6. Determinación de la efectividad del programa. Esquema de "triangulación".

Respecto a este último punto, queremos ejemplificarlo con la validación de los resultados que se hace en la "evaluación iluminativa" mediante el procedimiento denominado "triangulación". En el hecho tal procedimiento consiste en los siguientes tres fuentes de validación:

- a) Confrontación de información cualitativa y cuantitativa (triangulación metodológica).
- b) Comparación de diversas fuentes de información cualitativa (triangulación de fuentes de información) y
- c) 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 ahora, 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 evaluador, de cómo funciona, en este caso, el mecanismo del Programa Padres e Hijos. Se procedió, entonces, a triangular la Imagen Verbal (constataciones 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 estas 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 pre escolar 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	- 9	%
- Bueno	15 niños	52%	17 niños	35%	32	- 42	%
- Aceptable	9 niños	31%	24 niños	49%	33	- 41	%
- Deficiente	--	0	6 niños	12%	12	- 6	%
Total	29 niños	100%	49 niños	100%	78	- 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 en el siguiente cuadro:

Clasificación de dibujos

Mejor dibujo: niño del grupo de control

	Grupo control		Grupo PPH		Total
- Superior	3	25 %	2	42 %	5
- Bueno	3		9		12
- Promedio	6		10		16
- Pobre	9		3		12
- Inferior	3		2		5
	24 niños		26 niños		50 niños

Peor dibujo: niño del grupo control

La evidencia que muestra el cuadro complementó los resultados de los otros estudios que 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; otros 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: Cide, 1982,
pp. 52 - 57).

De manera característica, el informe de la evaluación no formal se redacta en término sencillos, con uso extensivo 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 (diapo-film) de diversas instancias de desarrollo del programa, etc.

(Para algunos otros detalles de la evaluación de programas globales se puede consultar a Thomas D. Cook and Charles S. Reichardt, eds., Qualitative and quantitative methods in evaluation research, Beverly Hills: Sage Publications).

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 en forma principal por el personal encargado del programa, especialmente por el equipo técnico de campo. Así sucede en proyectos de desarrollo rural área en la cual queremos destacar el enfoque de Oakley con las características que este investigador señala. En resumen - y en forma muy apegada a su propia descripción - el esquema utilizado comienza por identificar los criterios e indicadores apropiados de acuerdo con los cuales puede ser apreciado el progreso del programa. Para esto se elabora con los técnicos o monitores una lista de carac

terísticas (criterios) que poseían los usuarios 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:

a) Las características (criterios) de los usuarios antes del desarrollo del programa:

- Individualismo, reflejado en el predominio de la propiedad privada, poco interés en proyectos comunitarios, poca participación en las decisiones.
- Falta de análisis crítico de su situación: incapacidad para identificar las causas de problemas estructurales y buscarle solución.
- Dependencia económica, social y política del patrón y explotación por el 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 a 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 auto-confianza.
- Auto-gestión, auto-suficiencia y autonomía del grupo.

- Recursos colectivos. Reservas de capital.
- Capacidad para la administración del programa.
- Demostración del poder. Responsabilidad colectiva.
- Articulación con otras instituciones controlados 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 arriba; 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 piensa pueden ser utilizados para pronunciarse acerca de los cambios que se han producido o se están produciendo a nivel de los usuarios:

- Actividades realizadas por el grupo: obras de construcción, organizaciones, etc.
- Acciones específicas del grupo con o sin la ayuda del monitor. Ejemplos: acciones por obtener beneficios de seguridad social, o para demandar el cumplimiento de la legislación existente, etc.
- Cambios en la conducta del grupo: tipo de reuniones, uso del lenguaje, naturaleza de la participación, etc.
- Naturaleza de la intervención: cambios en las relaciones del monitor con el grupo como resultado de que el grupo ha asumido la responsabilidad para su propio desarrollo.
- Relaciones con otros grupos: naturaleza de las relaciones con otros grupos y acciones conjuntas, empresas y organizaciones conjuntas, etc.

En términos operativos se sugiere que el personal del proyecto encargado de la evaluación o de la recolección de informaciones para el trabajo a realizar por otro grupo registre continuamente sus observaciones o reuna los datos pertinentes. Para esto pueden emplearse los siguientes métodos:

- Encuesta para el perfil del grupo, es decir, para establecer la situación cuando comienza el programa.
- Recolección de datos cuantificables. Tales datos deben incluir el número de reuniones, de participantes, el número involucrado en actividades particulares y cifras básicas en cualquier actividad económica que se realice.
- Descripciones continuas y observaciones del grupo. Estas 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.
- Minutas en otras formas de registros de reuniones de los monitores.

Una vez recolectada toda la información se procede a ordenarla y a interpretarla en términos de los cambios que hayan podido producirse en relación con

las condiciones iniciales, los 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". Idea RRDC Bulletin, March 1982, 14, pp. 12 - 17)

La evaluación participativa.

Varias de las características de la evaluación que señalamos para los modelos globales se encuentran en la evaluación participativa. Como lo indica su nombre, se trata en ella de lograr la participación de los usuarios del programa o algunos de ellos en la evaluación del mismo. En sentido estricto, la evaluación participativa debería considerar el trabajo de los usuarios en todas sus etapas, vale decir, desde la constitución del grupo evaluador, pasando por la definición de objetivos, la evaluación de procesos y resultados, la recolección y análisis de los datos hasta llegar a la formulación de recomendaciones. En la práctica, sin embargo, la mayoría de las evaluaciones de este tipo solo han incorporado usuarios en la determinación de áreas problemáticas del programa como también en la evaluación de diversos aspectos del mismo mediante auto-evaluaciones y evaluaciones grupales. En todas estas tareas los evaluadores externos colaboran y, en muchos casos, orientan el trabajo del grupo.

Habría que señalar que la evaluación participativa - a cualquier nivel de in

corporación de los usuarios - tiene mejor utilización en poblaciones agrupadas de tal modo que es posible recurrir a evaluaciones grupales todas las veces que sea necesario (barrios, zonas marginales, etc). En poblaciones dispersas, la evaluación participativa debe limitarse a recoger opiniones y juicios individuales como también a auto-evaluaciones (campesinos, etc).

Si bien la evaluación participativa puede utilizar diseños generales como el que presentamos al comienzo de esta sección, lo cierto es que hasta el momento no se pueden derivar normas de las pocas experiencias existentes o de las experiencias que han incorporado usuarios tan solo a las primeras etapas de la evaluación. A este respecto, un trabajo reciente sobre el tema señala que:

"A pesar de esta creciente literatura, debemos enfatizar que las pautas de cómo debe llevarse a cabo la evaluación participativa están poco claras. Aunque parece esencial enfatizar los métodos cualitativos como contrarios a los cuantitativos, la gran variedad de literatura sobre los métodos cualitativos ofrece una gran gama de alternativas que hacen que la elección de procedimientos para evaluar cualquier proyecto debe ser resuelto de acuerdo a la experiencia, conocimiento y juicio".

(Steven Klees, Paulo Esmanhoto y Jorge Wertheim
"Evaluación educacional. Tendencias hacia el desarrollo de enfoques participativos". Marcela Gajardo (ed.) Teoría y práctica de la educación popular, Ottawa: International Development Research Centre, 1983, pág. 483).

A lo dicho convendría agregar que, en término teóricos y metodologías, la evaluación participativa se guía en gran medida por los esquemas conceptuales

de la investigación-acción, la cual no quiere decir, necesariamente, que de ellos puedan obtenerse normas directas para determinar la forma según la cual debería realizarse este tipo de evaluación. Dada esta conexión creemos útil hacer una breve descripción de la investigación-acción, tarea en la cual utilizamos extractos textuales de un trabajo de Stromquist (1983) sobre el tema.

La investigación acción

"La investigación acción pertenece al grupo de los estudios micro-sociales referidos a la familia, la sala de clases, el barrio, la comunidad. Su objetivo principal consiste en describir los procesos que se dan en estos grupos frente a ciertas situaciones problemáticas. Metodológicamente en estos estudios hay un distanciamiento de la preocupación por el muestreo, el uso de técnicas estadísticas y, en general, de la cuantificación de los fenómenos observados. Frente a estos componentes de la investigación tradicional se da prioridad al uso de técnicas cualitativas de muy distinta naturaleza.

En la búsqueda de solución a problemas identificados por el grupo, este grupo y los investigadores sociales comparten el rol investigativo. La aplicación de los resultados de esta tarea no constituye un paso diferente que debe ser ejecutado por otras personas, sino como un elemento integral de la actividad colectiva. Por lo mismo, es necesario que el investigador desarrolle un compromiso ideológico con el grupo.

Relacionada con la idea anterior, está la posición de que un estudio que busque el mejoramiento de un cierto grupo, no puede imponer cambios desde arriba o desde afuera. El grupo en sí tiene que devenir consciente de su situación, sus intereses, sus potencialidad de cambio. En la investigación-acción no se trata de verificar hipótesis ligadas a una tarea. En lugar de eso se busca comprender los factores que contribuyen a crear una situación social que es percibida por los involucrados en ella como problema ... Dentro de este tipo de investigación, las interpretaciones son correctas si éstas son aceptadas como tales por el grupo de participantes.

Una de las debilidades de la investigación-acción es que los datos cualitativos tienden a ser recogidos y presentados con mucha variabilidad en cuanto a profundidad y aspecto. Hay también una tendencia a presentar los datos por caso sin tratar una agrupación analítica de la rica información recogida. Otra debilidad de la investigación-acción es que la ruptura objeto-sujeto trae al proceso de recolección e interpretación de datos ^{a personas} con niveles muy bajos de educación. Las personas con más necesidad de beneficiarse de la investigación social son al mismo tiempo aquellas con menos posibilidad de adquirir conocimiento.

Sin duda una de las mayores debilidades de la investigación-acción es la limitación de sus actividades a un grupo o comunidad y la consecuente duda que este esfuerzo micro se pueda usar para lograr cambios macro y particularmente cambios estructurales".

(Nelly P. Stromquist, Algunas consideraciones metodológicas sobre la investi-

gación-acción. Documento presentado al Seminario Técnico Regional sobre Muje
res y Familias de los Estratos Populares Urbanos en América Latina, Cepal,
1983).