

→ estudios

## Manual de evaluación para políticas, planes, programas y actividades de juventud

Este es un Manual bajo el formato de una guía sistemática, cuya finalidad es la de ayudar a evaluar las políticas, los planes, los programas y cualquier actividad que se realice en el ámbito de juventud.

Esta afirmación inicial no implica que los ámbitos de juventud requieran estrategias particulares de evaluación. Porque la evaluación, sus métodos y técnicas son de aplicación general y no sólo para las políticas sociales. Pero en este manual, tanto los procedimientos como los ejemplos seleccionados, se refieren a cuestiones y temas relevantes para los ámbitos de juventud. Asimismo, las políticas de juventud se mueven sobre el terreno de algunos debates y orientaciones concretas. El Manual tendrá en cuenta todos estos hechos.

Por tanto, este Manual posee un alcance general y puede ser utilizado por profesionales de muy diferentes áreas y titulaciones, pero constituye una lectura “obligatoria” para casi todos los trabajadores del ámbito de juventud. Se trata de una lectura “obligatoria” porque todos los profesionales deberían poseer un mínimo nivel de conocimiento en torno a la evaluación y, de forma habitual, en la formación académica recibida, tema al que no se le presta demasiada atención. Hasta hace un tiempo este asunto se resolvía con el recurso de los “expertos externos”, quienes se ocupaban de la evaluación. Pero, como tendremos ocasión de ir viendo, recurrir a los “expertos” no es una buena solución, en una gran medida porque la evaluación se ha convertido en una “tarea de todos”, en la que el evaluador experto sólo se ocupa de ciertas tareas. Como consecuencia, la competencia profesional requiere, de una forma inevitable, la valoración sobre la capacidad para evaluar.

→ estudios

# Manual de evaluación para políticas, planes, programas y actividades de juventud

Manual de evaluación para políticas, planes, programas y actividades de juventud



→ estudios

injuve



**Redacción**

Observatorio de la Juventud en España  
Servicio de Documentación y Estudios  
C/ Marqués de Riscal, 16  
28010 Madrid  
Tel.: 91 363 78 09  
E-mail: estudios-injuve@migualdad.es  
web injuve: www.injuve.migualdad.es  
Catálogo General de Publicaciones Oficiales  
<http://www.060.es>

NIPO: 208-08-026-1

Dep. Legal:

Impresión: Lettergraf S.L.

Las opiniones publicadas en éste estudio  
corresponden a su autor.  
El Instituto de la Juventud no comparte  
necesariamente el contenido de las mismas.

## AVISO

Pocos días antes de morir Nicolás Maquiavelo contó a sus amigos un sueño. Andaba por un camino desconocido cuando se tropezó con un grupo de harapientos mendigos, torpes en el hablar, zafios y descuidados, a los que, tras esforzarse mucho, pudo entender y averiguar que aquello era el reino de los cielos, al que habían accedido por su condición de pobres miserables. Siguió andando y en otro lugar se tropezó con un “círculo de caballeros afables” que hablaban, de forma clara y precisa, de temas interesantes. Entre tales contertulios reconoció a Platón y a Tácito, los cuales, al observar su sorpresa, le contaron que aquello era el infierno, porque la sabiduría humana es enemiga de dios. Si esto es así, les dijo Maquiavelo a sus amigos, cuando muera me gustaría ir al infierno.

Tomado del prólogo de Ana Martines Arancon (traductora) a MAQUIAVELO, N. (1527), ***Discursos sobre la primera década de Tito Livio***, Madrid, Alianza, 1987.



**Manual de evaluación para políticas, planes,  
programas y actividades de juventud**

Domingo Comas Arnau



## **AGRADECIMIENTOS**

Tengo que reconocer las aportaciones y la revisión de los contenidos realizada por Josune Aguinaga de la UNED y Juan Andrés Ligeró de la Universidad Carlos III, mi agradecimiento se extiende al trabajo de corrección de Jacobo Muñoz de la Fundación Atenea. Por parte del INJUVE tengo que agradecer la labor de Julio Camacho y de Blanca Bardo por su dedicación a la edición de este texto.



Prólogo: ¿Para qué un manual de evaluación? .....	13
<b>PARTE I: LOS CONCEPTOS</b>	
1.- ¿Qué es la evaluación? .....	21
1.1.- Una breve historia de la evaluación .....	22
1.2.- La evaluación en España .....	28
1.3.- La función de la evaluación .....	30
1.4.- La estrategia evaluativa .....	32
Bibliografía y documentación complementaria .....	36
2.- Algunos debates en torno a la evaluación .....	37
2.1.- El método científico y la evaluación .....	37
2.1.1.- Investigación versus Evaluación .....	37
2.1.2.- Cuantitativo versus Cualitativo .....	39
2.1.3.- Racional versus Empírico .....	41
2.2.- Terminología .....	42
2.3.- Las relaciones de la evaluación con la planificación y la gestión, la calidad y la satisfacción .....	46
2.4.- Preparados para comenzar a despejar algunos equívocos .....	48
Bibliografía y documentación complementaria .....	50
3.- El diseño general y la planificación de la evaluación .....	51
3.1.- Los tres momentos de la evaluación y los tipos de evaluación .....	51
3.2.- Las fases de la evaluación .....	53
3.3.- Determinar los objetivos (¿qué queremos saber?) .....	54
3.3.1.- El Objetivo general .....	54
3.3.2.- Los objetivos específicos .....	54
3.4.- Las matrices de la evaluación (¿cómo vamos a hacerlo?) .....	55
3.5.- Los requisitos de la evaluación .....	57
3.6.- El papel del evaluador .....	59
3.7.- El ajuste de los tiempos: el desarrollo de la intervención y el cronograma de la evaluación .....	61
Bibliografía y documentación complementaria .....	64
<b>PARTE II. LOS MÉTODOS Y LAS TÉCNICAS DE PRODUCCIÓN Y RECOGIDA DE DATOS.</b>	
4.- Complementarios por deficientes .....	67
4.1.- Estrategias empíricas .....	67
4.2.- La “tabla redonda” de los métodos y las técnicas de recogida y producción de datos .....	68
Bibliografía y documentación complementaria .....	70
5.- La mirada del observador .....	71
5.1.- La observación científica .....	71
5.2.- Procedimientos de observación .....	73
5.3.- La actitud del observador durante la observación y la cuestión de la observación participante .....	74
5.4.- La observación y la evaluación .....	75
6.- Los datos producidos por otras observaciones o investigaciones .....	77

6.1.- Una maraña organizada.....	77
6.2.- Algunos tipos de fuentes secundarias.....	77
6.2.1.- Las estadísticas oficiales.....	79
6.2.2.- Las documentos oficiales.....	79
6.2.3.- Las documentos personales.....	80
6.2.4.- Las contenidos de los Medios de Comunicación.....	80
6.2.5.- Internet.....	80
6.2.6.- La bibliografía especializada.....	81
6.3.- El uso de fuentes secundarias en evaluación.....	81
7.- La entrevista primaria.....	83
7.1.- El dominio intencional.....	83
7.2.- Tipos de entrevistas.....	84
7.3.- La evaluación y las entrevistas.....	86
8.- Registro de casos.....	87
8.1.- Los sistemas de información y registro.....	87
8.1.1.- Una definición o como el todo puede ser la nada.....	87
8.1.2.- Sistemas de información y registro, “estadísticas” e indicadores sociales.....	89
8.2.- Los sistemas de información y registro en la historia de las ciencias sociales.....	89
8.2.1.- Antecedentes. La modernidad y la lógica de la política democrática.....	89
8.2.2.- La influencia de los registros en la constitución de la sociología clásica.....	91
8.2.3.- Las circunstancias de una crisis.....	92
8.2.4.- Los registros en la antropología social.....	93
8.2.5.- la nueva irrupción de los registros.....	94
8.3.- El diseño de un sistema de información y registro.....	95
8.4.- Tipos de sistemas de registro.....	97
8.4.1.- Anónimos y confidenciales.....	97
8.4.2.- Notifican acontecimientos y notifican casos.....	97
8.4.3.- Especializados y no especializados.....	98
8.5.- Dificultades inherentes al marco institucional.....	98
8.5.1.- Hegemonías corporativas.....	98
8.5.2.- Antiguos y nuevos registros.....	99
8.5.3.- Ámbitos de aplicación y perspectivas de futuro.....	99
8.6.- Evaluar a partir de registros: oportunidades y riesgos.....	101
9.- El relato y la percepción de la complejidad.....	103
9.1.- Reconstruyendo con intención.....	103
9.2.- Produciendo historias de vida.....	105
9.3.- La evaluación y los relatos.....	106
10.- Informantes clave e informadores cualificados.....	107
10.1.- La compleja cuestión de la mediación social.....	107
10.2.- Los informantes clave.....	109
10.3.- El método Delphi.....	110
10.4.- Evaluación, mediación e influencia.....	112
11.- La encuesta: la organización industrial y tecnológica.....	115
11.1.- La legitimidad social de la encuesta.....	115

11.2.- Los orígenes de la encuesta representativa .....	116
11.3.- El proceso para la elaboración de encuestas.....	118
11.4.- Estrategias para la entrevista en encuestas .....	119
11.5.- La validez y la utilidad de las encuestas .....	121
11.6.- Las encuestas en la evaluación.....	122
12.- El grupo de discusión: la interacción y la administración de los silencios .....	123
12.1.- El paso final y apertura primordial .....	123
12.2.- La técnica del grupo de discusión.....	125
12.3.- La utilidad de los grupos de discusión.....	127
PARTE III. LOS PROCEDIMIENTOS DE LA EVALUACIÓN.	
13.- El análisis de los datos .....	131
13.1.- Una tarea para los artesanos abiertos a la participación .....	131
13.2.- El análisis de la realidad y el diagnóstico de la situación.....	132
13.3.- El análisis del diseño .....	136
13.4.- El análisis de la implementación y de los procesos.....	139
13.5.- El análisis de los resultados y los impactos .....	141
13.5.1.- La función del análisis de resultados .....	141
13.5.2.- La selección de estándares, dimensiones e indicadores .....	141
13.5.3.- El control de variables y efectos espurios .....	143
13.5.4.- Los grupos control y los diseños cuasi-experimentales .....	144
13.6.- Otros análisis más particulares.....	145
13.6.1.- Análisis de la equidad.....	145
13.6.2.- Análisis de sensibilidad .....	146
13.6.3.- Análisis de sostenibilidad .....	146
Bibliografía y documentación complementaria .....	147
14.- El informe de evaluación y la difusión de los resultados.....	149
14.1.- El informe de evaluación: contenidos .....	149
14.2.- Publicación, distribución, registros y bases de datos.....	150
14.3.- La meta-evaluación.....	151
14.4.- Tipos peculiares de evaluación: la evaluación formativa .....	153
Bibliografía y documentación complementaria .....	154
15.- Los recursos de la imaginación. ....	155
15.1.- La evaluación como acto creativo.....	155
15.2.- Las técnicas participativas .....	157
15.2.1.- Los talleres de contraste de resultados .....	158
15.2.2.- Las técnicas de consulta externa .....	159
15.2.3.- El diagnóstico rápido participativo .....	160
15.3. Las estimaciones basadas en la evolución de los indicadores formales .....	161
15.4.- La noción de triangulación .....	162
Bibliografía y documentación complementaria .....	163

PARTE IV: APORTACIONES PRACTICAS.

16.- Un modelo de indicadores para evaluar políticas, planes, programas y actividades de juventud.....	167
16.1.- El debate sobre los indicadores .....	167
16.2.- Algunos modelos de indicadores .....	169
16.3.- Propuesta de indicadores para la evaluación en los ámbitos de juventud .....	173
16.3.1. El problema del déficit de evaluación y la cuestión de las "Memorias" .....	173
16.3.2.- Indicadores seleccionados.....	174
16.3.3.- El manejo de los indicadores .....	179
17.- Recopilación de ideas clave sobre la evaluación:	
la doble faz de la evaluación.....	181
17.1.- En torno a la noción de evaluación fuerte.....	181
17.2.- En torno a las estrategias de evaluación .....	181
17.3.- En torno a las técnicas y a los métodos de recogida y producción de datos.....	182
17.4.- En torno a los estándares de evaluación .....	182
17.5.- En torno a la acción evaluativa .....	182
17.6.- En torno a los indicadores para evaluar políticas de juventud.....	183
Bibliografía citada .....	184

ANEXOS:

I.- Estándares de evaluación del Joint Comitee on Standard for Educational Evaluation (1981-1994) .....	193
II.- Grupo de Evaluación de las Naciones Unidas (UNEG) Estándares de evaluación en el Sistema de las Naciones Unidas 2005 .....	197

## Prólogo: ¿Para qué un Manual de evaluación?

Este es un Manual bajo el formato de una guía sistemática, cuya finalidad es la de ayudar a evaluar las políticas, los planes, los programas y cualquier actividad que se realice en el ámbito de juventud.

Esta afirmación inicial no implica que los ámbitos de juventud requieran estrategias particulares de evaluación. Porque la evaluación, sus métodos y técnicas son de aplicación general y no sólo para las políticas sociales. Pero en este manual, tanto los procedimientos como los ejemplos seleccionados, se refieren a cuestiones y temas relevantes para los ámbitos de juventud. Asimismo, las políticas de juventud se mueven sobre el terreno de algunos debates y orientaciones concretas. El Manual tendrá en cuenta todos estos hechos.

Por tanto, este Manual posee un alcance general y puede ser utilizado por profesionales de muy diferentes áreas y titulaciones, pero constituye una lectura “obligatoria” para casi todos los trabajadores del ámbito de juventud. Se trata de una lectura “obligatoria” porque todos los profesionales deberían poseer un mínimo nivel de conocimiento en torno a la evaluación y, de forma habitual, en la formación académica recibida, tema al que no se le presta demasiada atención. Hasta hace un tiempo este asunto se resolvía con el recurso de los “expertos externos”, quienes se ocupaban de la evaluación. Pero, como tendremos ocasión de ir viendo, recurrir a los “expertos” no es una buena solución, en una gran medida porque la evaluación se ha convertido en una “tarea de todos”, en la que el evaluador experto sólo se ocupa de ciertas tareas. Como consecuencia, la competencia profesional requiere, de una forma inevitable, la valoración sobre la capacidad para evaluar.

Por este motivo el Manual va dirigido a todos los técnicos y profesionales del ámbito de juventud, y lo hemos calificado de “obligatorio”, no tanto porque haya que leer este manual en concreto, sino porque actualmente resulta imprescindible un cierto nivel de conocimiento sobre la cuestión de la evaluación, para cualquier profesional de cualquier titulación y con independencia del tipo de tarea concreta que realice. Deberíamos reiterar una y otra vez que saber sobre evaluación se ha convertido en un elemento constitutivo básico de cualquier buen desempeño.

Uno de los elementos que distinguen este Manual, frente a otros, es que va dirigido de forma expresa a los técnicos de juventud y que además no es solo un “manual para expertos” sino que intenta ser útil a todos, incluso para las que quieren simplemente conocer las nociones y las destrezas básicas de la evaluación de políticas, programas y proyectos, para poder desempeñar así mejor sus tareas.

En este sentido el manual puede ser utilizado por varias categorías diferentes de lectores, cada una de las cuales puede hacer una lectura muy distinta del mismo:

- a. Una lectura básica para comprender en qué consiste la evaluación y que puede obviar incluso algunos apartados demasiado técnicos sin renunciar a la comprensión global del texto.
- b. Una lectura profesional que permite participar en el equipo de un proyecto de evaluación y que debe incluir la lectura de todo el libro, e incluso de alguno de los textos citados como bibliografía complementaria.
- c. Una lectura para expertos en evaluación, ya que no se ha rehuido ningún debate, se ha tomado posición en casi todos ellos, se ha justificado tal posición y además se ha añadido una bibliografía actualizada.

Ocurre que el mayor problema que afronta la evaluación, y que este Manual trata de evitar para sus lectores, se refiere a la creciente complejidad, cuando no confusión, que envuelve el concepto de evaluación y los límites del territorio del mismo. Para ayudar a afrontar esta complejidad podemos afirmar, a modo de síntesis, que de una parte aparece una concepción autónoma, comprensiva, central y **fuerte** de evaluación, que vamos a desarrollar en las próximas páginas. Pero que, por otra parte, también aparece una numerosa constelación de concepciones **particulares** y corporativas de evaluación a las que nos vamos a referir de forma crítica, porque son las responsables de la confusión.

La primera es una concepción bien delimitada, con autores reconocidos, revistas científicas propias y un notable desarrollo conceptual y operativo. Pero su éxito ha producido una reacción paradójica, porque todas y cada una de las disciplinas académicas y los campos de conocimiento, todos y cada uno de los grupos corporativos, todas y cada una de las orientaciones teóricas y conceptuales, han tratado de constituir sus propios modelos de evaluación.

Esto ha ocurrido, como tendremos ocasión de comprobar, por la creciente importancia atribuida a la evaluación. Casi todo el mundo ha tratado de participar en esta innovación y además se trataba, en parte, de impedir que agentes externos metieran las narices (y valoraran) lo que un determinado colectivo estaba realizando<sup>1</sup>. Como además, durante un tiempo, la llamada “evaluación interna” no tuvo demasiada buena prensa<sup>2</sup>, la respuesta más frecuente ha consistido en adoptar estrategias evaluativas (y lenguajes técnicos) peculiares para que “todo se quedara en casa”<sup>3</sup>. Obviamente cada una de estas orientaciones ha tratado de mostrar, en el ámbito científico global, que trabajaba de una forma adecuada y por tanto debía ser reconocido, porque la mejor forma de alcanzar el reconocimiento es conseguir que esta noción particular de evaluación se utilice en otros ámbitos. Las pugnas políticas<sup>4</sup> y corporativas por conseguir el reconocimiento general han enriquecido el ambiente de la evaluación, pero no han desdibujado la noción fuerte de la misma.

En todo caso nos encontramos con una especie de hipermercado de la evaluación, en el cual diversos productos compiten por llamar la atención del cliente y que éste los compre. Muchos

---

1 Lo cierto es que no se trata de “todo el mundo”. Como han mostrado diversos estudios y por razones que tendremos ocasión de explicar en el capítulo 1, se evalúan las políticas de bienestar (en particular las políticas de servicios sociales, integración social, lucha contra la pobreza y la exclusión así como las educativas), pero a nadie se le ha ocurrido evaluar las políticas económicas, las de justicia, las de seguridad y defensa y por supuesto las infraestructuras (Ligero, 2005) y que ino se le ocurra a nadie! digo yo.

2 Un temor infundado porque, como veremos, la noción fuerte de evaluación requiere el consenso del grupo promotor y apuesta decididamente por las prácticas de auto-evaluación.

3 En una reciente publicación sobre la situación actual de las universidades europeas hemos leído “EVALUAR: Acción que consiste en fijar un conjunto de criterios tales que, aplicados al objeto o persona evaluables, consigue que los resultados obtenidos sean iguales a los resultados deseables” (Bermejo, 2007). Se trata de una opinión que ya comparen muchos agentes sociales y que hay que atajar endureciendo criterios.

4 En otra reciente publicación que se supone evalúa la sostenibilidad del Estado de Bienestar en España, hemos leído que el sistema de pensiones ya no es sostenible a consecuencia de la aprobación en España de las siguientes Leyes, la de Igualdad, la del Matrimonio Homosexual, la de Reforma del Divorcio, la de Identidad Sexual, así como curiosamente la Ley de Investigación en Biomedicina y la Ley sobre Técnicas de Reproducción Asistida (Patxot y Farre, 2007). Para poder hacer esta afirmación los autores consideran que estas leyes “reducirán la natalidad”, lo cual no sólo es incierto según otros estudios empíricos y más centrados en este tema (Aguinaga, 2004), sino que además, como reconocen los propios autores la falta de sostenibilidad del sistema se produciría “en ausencia de migraciones”. Vale.

se presentan además anunciando que son “los únicos” capaces de producir un determinado efecto. Otros, en cambio, se limitan a afirmar que son “buenos y baratos”. Lógicamente el cliente o sabe mucho de evaluación para poder decidir de manera informada o corre el riesgo de comprar algo inadecuado y en ocasiones perfectamente inútil.

Tampoco podemos rechazarlos a todos sin más y quedarnos sólo con el producto genérico, con la mera concepción fuerte de evaluación, porque la propia teoría afirma que “hay que utilizar todas las técnicas y métodos y no se pueden desaprovechar las mejoras vengan de donde vengan”. En realidad muchos de los productos particulares, tanto de procedencia corporativa como de una orientación teórica determinada, han realizado aportaciones de indudable interés.

Realmente la noción fuerte de la evaluación no posee un corpus de técnicas y métodos propio, sino que utiliza las técnicas, conceptos e instrumentos de diferentes disciplinas y áreas de conocimiento, en particular el área de Ciencia Política y las de Pedagogía, Economía, Psicología y Sociología. Este conjunto de aportaciones no ha concluido, aunque en este momento resulta esencial preservar la propia concepción autónoma y fuerte de la evaluación.

Por este motivo el manual trata, utilizando criterios científicos de evidencia y funcionalidad, así como una alta dosis de sentido común, de ofrecer respuestas razonables para aplicar y elegir aplicaciones a la hora de evaluar los programas.

En resumen, estamos ante un intento de sistematizar, pero el resultado es una sistematización limitada, porque la visión de las diversas alternativas conforma un panorama abierto. A pesar de lo cual, creo que el manual ofrece respuestas prácticas muy útiles para evitar perderse en el bosque de la evaluación.

Se trata, pues, de un intento quizá demasiado ambicioso, justamente en un Manual que en diversos momentos llama la atención sobre las consecuencias negativas de ser demasiado ambicioso en el diseño y planteamiento de las evaluaciones. Un intento que, en todo caso, “debería evaluarse”, cosa que es perfectamente posible, porque creo que he expuesto el objetivo del mismo de una forma clara y transparente.

## UN EJEMPLO Y ALGUNOS COMENTARIOS PARA IR DESBROZANDO UNA APROXIMACIÓN A LA EVALUACIÓN.

En el número 75 del Boletín **Hilero Eguneratuz**, correspondiente al mes de mayo de 2007, y cuya consulta puede ser la mejor manera de estar al día en el complicado territorio de los Programas y las Investigaciones sociales, aparecía la siguiente noticia ([www.siiis.net](http://www.siiis.net))<sup>5</sup>.

### **Uso de un experimento aleatorio a gran escala para evaluar un programa de inserción**

[De uso común en ciencias naturales, los experimentos con muestras seleccionadas al azar son poco frecuentes cuando se trata de evaluar políticas sociales. Sin embargo, tal como demuestra esta investigación, pueden constituir una estrategia válida para determinar la eficacia de una nueva medida]

WALKER, R. Y OTROS **Making random assignment happen: evidence from the UK Employment Retention and Advancement (ERA) demonstration**. Londres, Department for Work and Pensions, 114 págs., 2006.

El presente informe recoge los resultados del mayor experimento aleatorio realizado hasta la fecha en el Reino Unido con el fin de valorar una política social a gran escala. Esta metodología de investigación consiste en tomar una muestra de la población destinataria de un servicio o prestación, dividirla de modo aleatorio entre un grupo que recibe ese beneficio y otro que no, y analizar las diferencias que se derivan de ello.

El programa evaluado se denomina Employment Retention and Advancement

<sup>5</sup> Hilero Eguneratuz es una publicación editada por el SIIS, Centro de Documentación y Estudios de la Fundación Eguia Carega de San Sebastián.

(ERA). Empezó su andadura en 2003 en seis zonas diferentes de Inglaterra, Escocia y Gales con la vocación de diversificar las políticas activas de empleo. Los programas tradicionales de activación del empleo en el Reino Unido se dirigen a los desempleados que perciben algún subsidio, y su misión es ayudarles a conseguir un trabajo remunerado y a que lo conserven a corto plazo. En cambio, el ERA restringe su alcance a tres colectivos, -padres o madres solos y sin trabajo, parados de larga duración, y padres o madres solos, con jornadas parciales-, que suelen encontrar dificultades para mantener un empleo durante cierto tiempo. Precisamente, ése es el cometido de este programa: ofrecer asesoramiento personalizado y bonificaciones económicas para que, una vez que logren un puesto de trabajo, permanezcan empleados durante más tiempo consecutivo y puedan avanzar dentro del mercado laboral.

La evaluación del programa se realizó desde cuatro ángulos complementarios: resultados conseguidos, proceso de implementación, costes y eficiencia (cotejo de costes y beneficios). Los datos que sirvieron de base para la evaluación se recogieron mediante el seguimiento durante 15 meses de 2004 de una muestra de más de 16.000 usuarios de los servicios de empleo, quienes aceptaron participar en el experimento y fueron distribuidos de modo aleatorio en dos grupos: uno al que se le permitió participar en el ERA y otro al que esa posibilidad le fue negada. La comparación de los resultados de ambos contribuyó a determinar los beneficios del programa.

Como ya avanza en su título, *Making random assignment happen...* indaga en un aspecto muy concreto de la citada evaluación, a saber, la manera en que se asignaron los casos a uno de los dos grupos. Mediante el uso de diversas técnicas cualitativas -observaciones, entrevistas semi-estructuradas y grupos de discusión-, estudia la puesta en marcha de esta metodología en las oficinas de empleo, valora la eficacia del experimento y extrae una serie de recomendaciones de cara a la aplicación de este enfoque en otras evaluaciones de políticas sociales.

**Resultados.** Un primer tema de interés gira en torno a la obtención del consentimiento informado por parte de los participantes en el experimento. Según el informe, en la fase de asignación se produjo un conflicto entre la necesidad de incorporar usuarios al estudio y la exigencia ética de que éstos tuvieran suficiente información para decidir al respecto. Vista la complejidad del programa y del experimento, así como el tiempo disponible para la entrevista de presentación, los técnicos de empleo se inclinaron por ofrecer pocos detalles sobre ambos, salvo que se les pidiera expresamente, lo que sucedió en raras ocasiones. Además, si bien el programa les parecía positivo, tendían a ocultar aquellos aspectos que no les convenían. Por todo ello, aunque las negativas a participar fueron escasas, pocos usuarios tenían una idea exacta del programa y del experimento.

Por lo que se refiere a la gestión del proceso, el informe destaca que parte de los técnicos de empleo recibieron una formación escasa sobre el programa, y que ésta se centró en los aspectos burocráticos más que en su contenido y objetivos. También hubo un malentendido sobre el modo en que los centros de empleo debían organizarse para realizar la asignación aleatoria. Finalmente, se comprobó que el reclutamiento de usuarios para el experimento entraba en conflicto con los fines de otros programas de inserción laboral.

A pesar de los fallos detectados, los autores de esta investigación consideran que esos errores no invalidan el experimento y, por tanto, concluyen que la asignación aleatoria puede resultar útil para evaluar políticas sociales a gran escala. Las lecciones aprendidas les sirven, además, para proponer algunas recomendaciones. Desde la perspectiva de la gestión, aconsejan asegurarse de que el experimento no interfiera con el funcionamiento normal de la institución. Asimismo, insisten en la

conveniencia de proporcionar una formación adecuada a todo el personal involucrado en el estudio, de modo que los usuarios decidan participar de manera libre e informada. Por último, y desde el punto de vista científico, sugieren que se realice una prueba piloto y que se revise la implementación del procedimiento en cada uno de los lugares donde se lleve a cabo, lo que permitirá detectar y corregir las eventuales desviaciones que se produzcan.

-----

La lectura de este texto nos permite introducir las primeras ideas clave en torno a la evaluación. Entre algunos componentes que iremos dilucidando en el texto, podemos comenzar por cuatro cuestiones que conviene presentar de entrada. De esta manera, se intentará romper con algunos estereotipos sobre la evaluación, tratando a la vez de evitar que tales estereotipos nos impidan una buena comprensión del texto.

- EL MODELO EXPERIMENTAL NO SE ADECUA A LOS FINES DE LA EVALUACIÓN. El modelo experimental ha sido el modelo científico por excelencia, sin embargo, en el terreno de los estudios sociales aparecen numerosos problemas epistemológicos cuando tratamos de recrear las metodologías experimentales. El enunciado de analogías entre lo físico y lo social no garantiza una correspondencia en las reglas que rigen la lógica y el funcionamiento de ambos mundos. No se trata de renunciar a las reglas del conocimiento científico, sino de tener presentes las dificultades metodológicas que comprometen los resultados empíricos en el área de lo social.
- EL TAMAÑO DE LA RECOGIDA DE DATOS NO GARANTIZA LA CALIDAD DE LA EVALUACIÓN. La calidad de la evaluación no es una cuestión de tamaño en términos de recursos, porque un gran experimento puede resultar fallido si no somos capaces de proporcionar una cierta coherencia en el diseño, en la aplicación, en la selección de casos, en la recogida de datos y en otras normas básicas de la evaluación. A la vez, algo muy simple puede resultar muy exitoso si cumplimos con estos mismos requisitos.
- LOS EXPERTOS EXTERNOS NO BASTAN. Se supone que la evaluación compete a todos los implicados y no sólo a los evaluadores, ya que sin una adecuada colaboración (que implica una buena información, el consentimiento y la participación) de todos los agentes activos, bien en el programa, en la acción o en la política, la evaluación mostrará todas sus insuficiencias. De hecho *“La evaluación en sí sólo puede concebirse como una política resultante de un interés común”* (House, 1994).
- LAS BUENAS EVALUACIONES PROVOCAN LA EMERGENCIA DE COMPONENTES INESPERADOS. La idea clave se refiere aquí al hecho de que en el proceso pueden aparecer resultados y hallazgos inesperados, que en algunos casos incluso pueden llegar a mostrar que nuestro diseño es inadecuado. Cuando esto ocurre debemos estar satisfechos, porque este tipo de “descubrimientos” es muy positivo. Se entiende que en estos casos la evaluación ha servido para evitar los errores futuros, siempre que seamos capaces de interpretar estos hallazgos y de saber integrarlos en los programas y en futuras evaluaciones. Aquella evaluación que se limita a confirmar que todo va bien es menos útil que aquella evaluación que nos sorprende, que nos dice cosas nuevas, que nos señala las equivocaciones.

## TRES TEXTOS CLAVES SOBRE LA EVALUACIÓN EN GENERAL.

Actualmente, para hablar de evaluación el texto canónico de referencia es **Evaluation: methods for studying programs and policies**, de Carol H. WEISS (1998), **Nueva Jersey, Prentice Hall**.

Por desgracia, no es un libro traducido al castellano. Por este motivo se proponen otros dos textos más accesibles:

BAÑON, R. (2003), **La evaluación de la acción y las políticas públicas**, Madrid, Díaz de Santos.

ROSSI, P. H. y FREEMAN, H. (1989), **Evaluación, un enfoque sistemático para los programas sociales**, México, Trillas.

**PARTE I**  
**LOS CONCEPTOS**

*“Las teorías son redes que lanzamos para apresar aquello que llamamos **el mundo**: para relacionarlo, explicarlo y dominarlo. Y tratamos de que la malla sea cada vez más fina”.*

Karl R. POPPER (1958), **La lógica de la investigación científica**, Madrid, Tecnos, 1972.



## ¿Qué es la evaluación?

**La evaluación es un canon sistemático de prácticas sobre políticas, programas y otras actividades sociales e institucionales, que nos permite recoger información, valorarla y devolverla a los interesados y afectados, tratando de mejorar estas intervenciones. Se trata de prácticas fundadas en el método científico y en la razón, que de forma global pretenden provocar un cambio positivo en la vida de las personas<sup>6</sup> y en la organización de las sociedades.**

La versión correspondiente para los ámbitos y los organismos de juventud se formularía como sigue: *“La evaluación es un canon sistemático de prácticas sobre las políticas, los planes, programas y actividades de juventud, que nos permite recoger información para devolverla a los jóvenes, los técnicos de juventud, los responsables políticos y administrativos y a la sociedad en general, tratando de mejorar (en términos de eficiencia, transparencia, eficacia, cobertura,...) todas estas intervenciones. Se trata de prácticas fundadas en el método científico y en la razón, que de forma global pretenden provocar cambios positivos en la vida de los jóvenes, tanto en su condición individual como colectiva, así como en su participación e integración en la sociedad”*.

El conjunto de autores que citaremos en las próximas páginas han propuesto completar esta definición con otras frases y términos que, sin embargo, no la modifican en lo sustancial. Pero podemos citar estas expresiones (que se repiten y se interfieren) para redondear la noción fuerte de evaluación. Así algunos autores hablan de *“la aplicación sistemática de los procedimientos de investigación social para realizar una reflexión sistemática sobre una política, un programa o una actividad”*; una parte de los mismos le añaden la idea de *“para tomar decisiones”*, que algunos matizan afirmando que *“toda evaluación exige constatar su utilidad”*. Una definición muy utilizada nos dice que la evaluación es *“la colección sistemática de informaciones acerca de actividades y resultados de una política, un plan o un programa”*, claro que en este caso conviene siempre añadir que se trata de *“informaciones útiles para contribuir a la mejora del programa”*.

También se ha descrito con frecuencia la evaluación como *“un juicio de valor cualificado”*, pero aun siendo esta escueta definición muy precisa y muy exacta, tanto que la he utilizado en todas mis anteriores publicaciones, tiene algunos inconvenientes porque nos obliga a determinar cuáles son los criterios de *“valor cualificado”*, lo cual nos conducía, en la propia definición, al com-

<sup>6</sup> En ocasiones de una manera directa a través de la mejora de las acciones concretas y en otras de una forma más indirecta proporcionando criterios de eficacia de las intervenciones sociales. En este sentido y a modo de ejemplo “evaluar la eficiencia de una intervención” no tiene demasiado sentido si esta acción no repercute sobre la vida de la personas, bien sea evitando el despilfarro de los recursos públicos o bien sea resolviendo algún problema sectorial muy concreto.

plicado debate sobre “las técnicas y los métodos de recogida de datos”, tema al que tendremos que dedicar una reflexión en el siguiente capítulo y de forma íntegra una de las cuatro partes en las que se divide este Manual.

Asimismo conviene apuntar que a muchos autores les gusta añadir en el contenido de la propia definición, que la evaluación no se refiere sólo a los resultados, sino también “*al diseño y la implementación*”.

Por último también parece conveniente incluir una aclaración terminológica relacionada con la abundante literatura en inglés, de donde proceden estas definiciones, y que concierne a los términos “*value*” y “*merit*”. En los diccionarios más comunes la traducción literal respectiva es “valor, mérito, valía, sentido, aprecio y estimación” para el primero de ellos, mientras que “merecimiento, premio, virtud, excelencia y mérito” para el segundo. Una parte de los autores ingleses trata de sustituir “*value*”, término considerado excesivamente cuantitativo, por “*merit*”, de carácter más cualitativo y contundente. En las traducciones se utiliza “valor” por “*value*” y “mérito” por “*merit*”. Una parte de los autores españoles afirma que en castellano, “valor” y “mérito” son lo mismo y por tanto la distinción no funciona en nuestro idioma, pero otra parte asume para ambos términos la diferencia que se supone que mantiene el inglés. ¿Qué nos dice el Diccionario de la Real Academia (DRAE) en su última edición?

**VALOR:** (14 acepciones, relevantes para el tema 4).

- Grado de utilidad o aptitud de las cosas, para satisfacer las necesidades o proporcionar bienestar o deleite.
- Alcance de la significación o importancia de una cosa, acción, palabra o frase.
- Fuerza, actividad, eficacia o virtud de las cosas para producir sus efectos.
- Cualidad que poseen algunas realidades, consideradas bienes, por lo cual son estimables. Los valores tienen polaridad en cuanto son positivos o negativos, y jerarquía en cuanto son superiores o inferiores.

**MÉRITO:** (3 acepciones, relevante para el tema 1).

- Aquello que hace que tengan valor las cosas.

Parece claro que el término “mérito”, empleado en inglés para hablar de evaluación, funciona en castellano como un simple sinónimo, más débil incluso, de “valor”, el cual, al contrario que en la lengua anglosajona, tiene una gran potencia en nuestro idioma.

Aparte de estos matices, y como puede verse en cada definición de evaluación, incluida la propia, todas ellas coinciden en destacar tres componentes: **su condición de investigación científica, su orientación aplicada y su finalidad valorativa**. Siempre que incluyamos estos tres componentes una definición de evaluación es perfectamente válida.

### 1.1.- Una breve historia de la evaluación.

La trayectoria histórica de la evaluación, que a la vez nos sirve para comprender la definición que encabeza este capítulo, se suele siempre explicar de una forma similar, aunque existen diferentes versiones en cuanto a los periodos o fases que la caracterizan. En una publicación anterior del INJUVE, establecí un esquema propio que trataba de sintetizar todas estas versiones y que ha tenido una cierta proyección (Cuadro 1.1).

La evaluación como una tarea consustancial de las intervenciones sociales aparece en Estados Unidos en la década de los años 30 del siglo XX. La idea de evaluar los programas sociales apareció como consecuencia de la implantación de la New Deal, que fue la respuesta de la Administración de Franklin Delano Roosevelt (1933-1945) a la profunda crisis económica del año 1929. La New Deal adoptó el principio de John Maynard Keynes de utilizar el gasto público, así

como algunas medidas fiscales y monetarias, con la finalidad de mitigar los efectos adversos de los ciclos económicos.

Pero aquella constelación económica, social y cultural tan particular, producida en Estados Unidos por la crisis de 1929, concluyó con la Segunda Guerra Mundial y con la etapa de las administraciones de Truman y Eisenhower en especial con el momento del Macartismo (1950-1956), que supuso la desaparición de muchos programas públicos y el desmantelamiento de todas aquellas iniciativas y prácticas evaluativas que implicaban actitudes de innovación. Pero la idea de la evaluación se mantuvo gracias al trabajo de Ralph W. Tyler, un psico-pedagogo de orientación conductista que, aparte de crear la propia noción de “evaluación”, sistematizó un modelo de evaluación para los ámbitos educativos. Ralph Tyler estableció la idea de que el sistema educativo se debía fijar objetivos y “evaluar” si se habían alcanzado, insistió además en que lo importante es “medir el cambio obtenido” y finalmente consiguió que le encargaran la evaluación de los resultados obtenidos por diversos proyectos de renovación pedagógica (Alvaro, 1988).

Cuadro 1.1. **Las fases históricas de la evaluación.**

FASE	CONTENIDO	RETOS
New Deal Años 30	Los programas sociales mejoran la vida de los ciudadanos, hay que optimizar recursos para ofrecer más servicios	Lo más barato no es lo mejor
La visión de la ciencia positiva Años 50-60	Hay que aplicar las técnicas de la ciencia experimental para seleccionar las mejores intervenciones	No se pueden extrapolar sin más los resultados a distintos contextos
La visión de la administración Años 70-80	Los recursos públicos son escasos y no se pueden despilfarrar	Son los ciudadanos los que pueden determinar las prioridades
La visión ética, política y social Años 90	Hay que democratizar la toma de decisiones y la participación de los ciudadanos más allá de los simples procesos electorales	Las estructuras evaluativas y administrativas previas quieren mantener su poder corporativo

Adaptado de Rossi y Freeman (1989) por Comas (2001)

Fueron también estos primeros evaluadores, la mayoría situados en el ámbito educativo, los que identificaron el problema que aparecía en aquellas primeras evaluaciones de los programas sociales: el mero cálculo del coste/eficacia en términos monetarios no servía como indicador para garantizar ni la calidad, ni la máxima eficiencia del programa. Expresado en otros términos, el coste no podía ser considerado como variable independiente y era conveniente abandonar el criterio de “gasto social mínimo” de la economía clásica si lo que se pretendía era optimizar los resultados y producir efectos verdaderos sobre el sistema social.

Pero de forma inesperada este argumento alarmó a las instituciones y contribuyó a la crisis de los programas sociales en Estados Unidos y al desarrollo de la “cultura económica liberal de los años 50”, de Milton Friedman y la Escuela de Chicago, lo cual, a la postre, desdibujó el papel asignado a la evaluación en los años 30. Al mismo tiempo, en Europa, en aquellos países en los que se estaba implantando el Estado de Bienestar, el mismo argumento servía para situarse en una posición inversa: el bienestar era un derecho de ciudadanía y los costes relacionados con este derecho era algo que no necesitaba ser evaluado. Es decir, en ambos casos se llegó a considerar que la evaluación era “una opción política inadecuada”. Sin embargo, mientras en Norteamérica el

peligro se resumía en que la evaluación podría justificar “un exceso” de gasto social, en los países europeos que apostaban por el bienestar social, la “amenaza” se refería a que la “evaluación” podía cercenar derechos de ciudadanía.

Quizá por este motivo, especialmente en los Estados Unidos, la evaluación abandonó el territorio de las prácticas y se instaló en el ámbito académico, en el que adoptó el perfil del “diseño experimental” de los programas construidos ad-hoc para ser evaluados y llegar a conclusiones universales en torno a su validez. Este es el periodo caracterizado por el psicólogo Donald Campbell, quien, con una propuesta muy cercana al diseño experimental, es una figura emblemática de la tradición de los estudios basados en la teoría del aprendizaje, la cual dominará el territorio académico durante las décadas de los años 50 y 60 (Campbell y Stanley, 1966).

Como consecuencia, en dicho periodo la mayor parte de los trabajos evaluativos en el campo social tuvieron un marcado carácter académico, se mantuvieron en el área de influencia de la psicología conductual y en la estela del diseño experimental (Herman, 1987), lo cual provocó numerosas críticas.

Se trata básicamente de críticas relacionadas con el procedimiento metodológico, ya que, siguiendo los criterios de Campbell, se trata de validar con las pretensiones de universalidad atribuidas al diseño experimental, un programa en un contexto y en un momento determinados. A la vez se supone que el éxito de la validación permite sucesivas aplicaciones clónicas que, aunque en muchos casos fuerzan situaciones particulares, lo cual conduce a evidentes malos resultados, eso no importa ni se puede discutir porque el trabajo ya “fue validado” en una universidad anglosajona. Un segundo tipo de críticas, más jocoso, hace hincapié en el hecho de que la totalidad de los programas que han sido a la vez diseñados y evaluados con una metodología experimental, desde instancias académicas norteamericanas, en los últimos treinta años y para ser distribuidos y vendidos en el ámbito internacional, han sido, sin excepción y a pesar de representar en ocasiones opciones antagónicas, perfectamente validados.

Lo cierto es que en la práctica existe un fuerte prejuicio favorable a cualquier programa diseñado y evaluado desde una perspectiva experimental, aunque para conseguir dicha evaluación haya sido preciso construir una ficción social. A la vez, los programas sociales reales, los que no se sostienen sobre un diseño cerrado y validado en una universidad anglosajona, permanecen, al menos, en el limbo de la sospecha. En otras palabras: la ficción construida en el laboratorio pretende representar la realidad social, mientras que la realidad no nos vale porque carece de una adecuada simetría metodológica. Quizá por este motivo el propio Donald Campbell, consciente de que no se podía sostener la idea de que la realidad social debe adaptarse al ambiente impoluto del laboratorio, sino a la inversa, propuso “*contrastar los diseños experimentales con otras metodologías más acordes con el desarrollo real de los programas en un contexto social variado*” (Campbell, 1979).

Las críticas contra el modelo de la evaluación-ficción comenzaron a expresarse a principios de la década de los años 60 y pudieron generalizarse gracias a una serie de cambios en el contexto social y político norteamericano, los cuales facilitaron la expansión de alternativas metodológicas innovadoras<sup>7</sup>. Tales cambios tuvieron que ver con las nuevas políticas impulsadas por la administración de John Kennedy y de forma más intensa (pero menos reconocida) por Lyndon Jhonson, en particular su “guerra contra la pobreza”, así como por la explosión del movimiento de los Derechos

<sup>7</sup> Entre estas innovaciones se sitúan los trabajos de Charles Wrigth y Herbert Hyman emprendieron en los años 30 evaluando programas de prevención de la delincuencia, campamentos para jóvenes infractores y otras intervenciones sociales relacionadas con prácticas de riesgo. Se trata de la experiencia más compleja de evaluación nunca realizada (con estudios de seguimiento que se prolongaron durante más de treinta años y utilización de grupos control y experimental en los que se controlaron muy bien las pérdidas), pero a la vez es la más invisible. Tanto Wrigth (sociólogo) como Hyman (psicólogo social), ambos de la Universidad de Columbia, son bien conocidos por sus estudios conjuntos en el terreno la Comunicación y en el tema de la “participación social y las asociaciones de voluntarios”, en los que se les reconoce un cierto protagonismo teórico. Pero en paralelo desarrollaron un intenso trabajo en el campo de la evaluación que muy pocos conocen porque la mayor parte de sus aportaciones están en Informes Institucionales restringidos y no publicados, aunque han sido descritos por otros autores como Donald Campbell. Un resumen de los mismos fue publicado en 1968 (Wrigth, 1968). De haber sido accesibles podríamos haber entendido, hace ya décadas, que una parte sustancial de las intervenciones sociales en el área de riesgos ni alcanzan sus objetivos, ni mejoran sus resultados en relación al grupo control, pero también habríamos aprendido que nos aportan otras ventajas, que la evaluación pone en evidencia y que sirven para fundamentar este tipo de programas.

Civiles, que también impulso múltiples programas sociales, o algunas innovaciones en la intervención social como el modelo CIPP (Stufflebeam, 1971). Todo este movimiento reforzó la noción de evaluación.

En Europa se produjo una situación muy similar aunque más tardía, ya en los años 70, cuando a partir de la crisis económica de 1973-74 (la famosa “crisis del petróleo” tras la guerra del Yon Kippur) comenzó a plantearse la necesidad de “racionalizar” los derechos adquiridos del Estado de Bienestar y la evaluación dejó de verse como una amenaza para los derechos ciudadanos para percibirse como una estrategia para prestarlos.

Es decir, mientras en Estados Unidos la evaluación se desarrollaba como consecuencia del refuerzo de las organizaciones cívicas y los programas sociales, en Europa era precisamente la supuesta “crisis del Estado de Bienestar” la que abría la puerta a la evaluación. Se trata de un tema importante, porque la mayor parte de la literatura sobre evaluación, en particular la evaluación fuerte, procede de Estados Unidos, de organizaciones cívicas y posee una lógica de “reivindicación social”; mientras que en Europa, las instituciones, nuestras instituciones, son las que desarrollan la evaluación siguiendo una lógica de “control del gasto”. También es cierto que la mayor parte de las agencias gubernamentales norteamericanas han adoptado en los últimos años una lógica a “la europea” (y enloquecen con sus demandas de evaluación a las organizaciones civiles), mientras que en Europa los profesionales (y las ONGs) que apuestan por la evaluación fuerte están alcanzando una cierta masa crítica.

En todo caso y desde puntos de partida muy divergentes (un primer desarrollo mínimo de los derechos sociales en una sociedad liberal y la crisis de algunos derechos consolidados en una sociedad socialdemócrata) se produjo una confluencia en torno a “la necesidad de la evaluación”. En ambos casos el supuesto problema político era “el gasto”, lo que facilitó la entrada de los economistas en el tema de la evaluación. Al mismo tiempo, al tratarse de programas reales e implantados, se recurrió a los sociólogos para que desarrollaran metodologías que resolvieran los problemas del trabajo de campo que los diseños cuasi-experimentales, controlados y de laboratorio que habían caracterizado la etapa precedente, no se habían planteado. Tanto para uno como para otro la noción de “validez universal” de un programa se sustituyó por la noción de “evaluación concreta” de un programa. Conviene señalar que también en ambos casos, es decir, tanto en Europa como en EEUU, tal y como hemos explicado en el prólogo, la evaluación se limitó (y se limita) al campo del gasto público dedicado a programas de bienestar, mientras que el gasto público destinado a seguridad, defensa o infraestructuras, debe ser “eficaz por sí mismo” porque a nadie se le ocurre evaluarlo (Ligero, 2005). En realidad esto ocurre porque muchos siguen pensando que desarrollar un programa de prevención de la violencia escolar, o promocionar el contacto inter-cultural, pueden producir despilfarro de recursos públicos, mientras que comprar armas no exige ninguna evaluación. Quizá tengan razón.

La figura más representativa de la nueva etapa es la de Michael Scriven, ya que revolvió de una forma radical las tranquilas aguas de la evaluación experimental con conceptos tan conocidos como son “evaluación formativa”, “evaluación sin metas” y la “meta-evaluación”, que supo aplicar a la propia evaluación de programas y políticas. El argumento central de Scriven es muy claro: los evaluadores se han dedicado de forma tradicional a preguntar cuáles eran los objetivos del programa para, a continuación, tratar de averiguar si se habían alcanzado o no. Pero a partir de ahora, decía Scriven, hay que tratar de averiguar, primero y de forma directa, cuáles son las características esenciales y reales del programa o política, incluidos sus fundamentos teóricos, su misión, así como la decisión que lo ha puesto en marcha, para determinar a continuación sus efectos. Es decir, se pasaba de un modelo de evaluación cerrado y predefinido a otro abierto tanto en el diseño como en los posibles resultados, porque primero teníamos que preguntarnos en qué consistía verdaderamente el programa (Scriven, 1991). El trabajo de Scriven ha sido ampliamente difundido, aunque en la práctica no se hayan asumido todas sus consecuencias (Shaw, 1999; House, 1994), en una gran medida porque a sus conceptos teóricos, como puede ser la

idea del meta-análisis, no le correspondieron avances metodológicos equivalentes hasta bien entrados los años 90.

El camino abierto por Scriven fue seguido por Lee Cronbach, quien plantea dos cosas muy sencillas por evidentes, pero que hasta aquel momento nadie había considerado. La primera señala que si hablamos de evaluación lo que importa es la validez externa de la misma en un contexto local, es decir, el valor concreto de un programa estable en un contexto de gobierno determinado; la segunda indica que para determinar el grado de validez, podemos utilizar cualquier combinación metodológica, siempre que la elijamos entre lo que Mario Bunge llama "*los métodos reconocidos y utilizables en el campo de conocimiento*". En cuanto al evaluador, su aportación es lo que Cronbach llama "*una mentalidad crítica, erudita e ilustrada*" que irrumpe sobre una práctica rutinizada en un contexto determinado (Cronbach, 1982).

Debemos tener en cuenta que la noción de "evaluador ilustrado" supone que el evaluador es un agente externo, lo que se justifica porque tiene que ser "*un profesional con destrezas específicas, un experto sin compromisos con las partes y capaz de ejercer de árbitro independiente y reconocido*" (Simon, 1999). Claro, que la función de este experto es parecida a la de un "animador", porque no es él quien realiza físicamente la evaluación, sino que ésta es una tarea que corresponde al propio equipo que realiza la intervención. Se trata de la noción del "evaluador (o evaluación) que responde"<sup>8</sup> de Robert Stake y que significa que el evaluador tiene como principal tarea la responsabilidad de proporcionar información accesible a aquellos que han realizado la intervención y recogido los datos, así como a otras personas como clientes y autoridades (Stake, 1975).

Todos los autores citados en los últimos párrafos protagonizaron, entre mitad de los años 60 y mitad de los 80, la ruptura con la evaluación "clásica" representada por la teoría del aprendizaje<sup>9</sup>. De esta ruptura nace la idea moderna de la evaluación fuerte, representada en la cuarta y última etapa del cuadro 1.1, en cuya aparición debe reconocerse, como señalan casi todos los autores, la labor de Carol Weiss, la cual partiendo del ámbito de la pedagogía ha construido una noción muy autónoma de evaluación que es la que presentamos en estas páginas. Weiss parte de una aseveración de sentido común, especialmente evidente en los ámbitos de juventud, aunque por ser tan obvia nadie parecía tener en cuenta hasta que ella la formula: los programas a evaluar son el fruto de decisiones político-administrativas, tanto en lo que se refiere a su diseño como a su continuidad; asimismo, la evaluación es una decisión político-administrativa que suele contener una "elección metodológica". Por tanto, no se pueden evaluar programas si no se tiene en cuenta la realidad de la subsidiariedad político-administrativa de los mismos (en Palumbo, 1987).

Pero Carol Weiss va mucho más lejos (Wagner y Weiss, 1991), cuando demuestra, de forma empírica, que los resultados de una evaluación sólo son aceptados, o mejor dicho, aceptables, según supuestos criterios científicos, cuando concuerden con la línea de las decisiones político-administrativas previas. Es decir, aunque el ejemplo podría ser el contrario, podremos probar mediante repetidas evaluaciones que la concreción de los programas y su gestión desde un organismo de juventud es más favorable para los jóvenes que las medidas de carácter transversal. Pero si el discurso político-administrativo, condensado en los presupuestos, considera que las

8 El término en inglés es "responsive evaluación" y una traducción muy adecuada no existe. Juan Andrés Ligerio propone "evaluación (o evaluador) sensible".

9 En este resumen no le he atribuido ningún protagonismo a la irrupción del modelo del constructivismo, que se presentó así mismo, en los años 80, como la "cuarta generación" de evaluadores. Sin duda la obra de Egon Guba e Yvona Lincoln (1987) tiene una gran trascendencia y la vamos a citar en diferentes momentos del texto. También cabe reconocer que sin el constructivismo no habría sido fácil conceptualizar una noción fuerte y autónoma de evaluación. El constructivismo puso en evidencia los elementos constitutivos reales de la evaluación y realizó una revisión muy consistente a la supuesta neutralidad de la ciencia positiva y de las administraciones públicas. Pero, una vez reconocida su aportación y sus, aún importantes, posibilidades críticas, el constructivismo nos ayuda poco a desarrollar esta noción fuerte y autónoma de evaluación. Es más, adopta incluso posiciones dogmáticas en torno a los métodos cualitativos y las llamadas "intervenciones focales" que desentonan con su supuesta actitud crítica. En España no estaría de más realizar una reflexión rigurosa en torno a los límites y las relaciones entre constructivismo y evaluación, pero este Manual (que incluye numerosas aportaciones del constructivismo) no parece el lugar idóneo para hacerlo.

“medidas transversales son más adecuadas”, la única evaluación aceptable será la que demuestre este aserto. Es decir, la única evaluación aceptable es aquella que confirma como “correcta” la política de juventud emprendida, lo cual no deja de ser lógico, ya que los agentes que promueven dicha política creen en la misma y se juegan su prestigio, y posiblemente su futuro electoral, sino resulta un “éxito”.

A modo de caso extremo (y ejemplo global) de esta situación, podemos observar como los diversos autores que evalúan políticas sociales han puesto en evidencia que las inversiones del Estado de Bienestar son, incluso en términos puramente económicos, muy eficientes. Sin embargo esto no ha servido de nada frente a la hegemonía política de los “razonamientos” del individualismo metodológico de la Escuela de Chicago, los cuales han sido utilizados como “argumentos de orden superior”, para, desde principios de los años 90, ir reduciendo, en todo el mundo desarrollado, las inversiones sociales aún a costa de disminuir la eficacia del capital humano y de la propia eficiencia económica. Es decir, la idea política del neoliberalismo es más importante que los resultados de las evaluaciones porque es una “ideología que orienta la acción política” y a la que se supone deben ajustarse los resultados de una evaluación, porque si no se ajustan seguro que la evaluación esta mal.

La consecuencia de todo ello es, en la versión de Carol Weiss, que en el ámbito de la evaluación el mero conocimiento instrumental, es decir, aquello relacionado con las buenas prácticas metodológicas, resulta insuficiente porque no tiene en cuenta el nivel político-administrativo en el que se toman las decisiones y en el que se refrenda la “validez formal” de unas evaluaciones y la condena al silencio de otras. La proyección de la evaluación se convierte entonces en una cuestión ética que compete especialmente a los profesionales que realizan los programas. En esta línea, Ernest House avanza en la idea del “evaluador ilustrado” de Crombach y afirma que *“la propia evaluación es ilustración”* y representa por tanto un movimiento de reforma y justicia social, entendidos no tanto como una práctica política, sino una orientación de los profesionales en su labor cotidiana (House, 1994).

Estamos por tanto hablando de una “evaluación justa” para profesionales y usuarios, que se superpone a la vieja idea de la mera “evaluación científica” (y la evaluación institucional) que respondía a un proyecto metodológico, cuando no ideológico, concreto. Es decir, la evaluación está fundada de forma imprescindible e irreversible en el método científico y la razón, pero esta fundamentación resulta insuficiente si no se tienen en cuenta aspectos éticos relativos a la misma. Además, para que una evaluación sea justa se supone que tiene que ser una “evaluación realista”. Un concepto que vamos a desarrollar en las próximas páginas, pero que esencialmente se refiere a la necesidad de tener en cuenta el contexto político-administrativo en el que se han desarrollado tanto el programa como la evaluación.

### **UN EJEMPLO PARA CREAR UN IMPACTO EMOCIONAL**

En cursos de formación suelo hacer el siguiente juego con diapositivas. En la primera proporciono una definición “clásica” de evaluación: valorar para mejorar sin más. Después explico lo que es eficiencia, eficacia y efectividad. A continuación indico que voy a mostrar un programa bien planificado y evaluado y que, como consecuencia, consiguió alcanzar un óptimo de eficiencia, de eficacia y de efectividad. La primera diapositiva muestra un edificio que casi nadie reconoce y a continuación informo que voy a mostrar los resultados del programa que aparece como una nueva diapositiva con una montaña bastante espeluznante de cadáveres desnutridos en un campo de concentración. Finalmente indico que la diapositiva anterior era la entrada de Auschwitz.

En general esto provoca enfado y un cierto malestar entre los alumnos, pero en el debate posterior resulta evidente que la definición clásica de evaluación nos permite concluir que aquel genocidio fue eficiente (se consiguió el máximo de re-

sultados con el mínimo de recursos), eficaz (la administración de Eichman alcanzó las metas y los resultados propuestos) y en parte efectivo, porque tuvo capacidad para dar una respuesta al antisemitismo de aquella sociedad sin producir apenas rechazos explícitos, al menos mientras aquella “política” se aplicó. Llegados a este punto los alumnos asumen la necesidad de recurrir a argumentaciones éticas y morales para justificar la evaluación, lo que me permite exponer que sin este tipo de criterios no vale cualquier evaluación, por muy bien diseñada que esté.

## 1.2.- La evaluación en España.

En España la implantación de la evaluación ha seguido un proceso histórico un tanto singular. De entrada el término “evaluación” se utilizaba, en las décadas de los años 60 y 70, de forma casi exclusiva en el ámbito de la psicológica para referirse al diagnóstico y en particular a la evaluación de actitudes mediante test (Pallares, 1981). Durante un largo periodo muchos profesionales españoles se han visto arrastrados a la confusión por la identificación entre psico-diagnóstico y evaluación, lo cual contribuyó al retraso en la implantación de las prácticas evaluativas.

Conviene también tener en cuenta que en España las instituciones tradicionales de control de la acción pública (en particular el Tribunal de Cuentas y los diferentes cuerpos de inspección y control) seguían, y siguen, un modelo administrativo muy bonapartista (y muy republicano lo que no deja de ser paradójico), centrado en la “*revisión jurídica y económica de las acciones*” (Ruano, 2003). La Constitución de 1978, en particular el artículo 136 (en concordancia con el 66 y siguientes), confirman esta visión, sin que ningún otro artículo permite entrever la noción moderna de evaluación (Iturriaga, 2003) y aunque permite que el propio Tribunal de Cuentas y el Parlamento adopte estrategias novedosas en su función de control y fiscalización, lo cierto es que la evaluación se ha tenido que plantear como un desarrollo paralelo a este procedimiento hegemónico y obligatorio.

Pero en todo caso, el retraso español en materia de evaluación ha confirmado la idea de Carol Weiss cuando afirma que “*democracia y evaluación son términos similares y el uno no puede existir sin el otro*” (Weiss, 1998), ya que sólo después de la transición democrática se comenzó a hablar de evaluación en términos formales. Pero además, a partir de la propia transición, tuvimos que afrontar otras dificultades que no han facilitado la recuperación del tiempo perdido. De una parte, la mirada social, política y académica se focalizó hacia otras urgencias y sólo a finales de los años 80 comenzó a aparecer una cierta literatura en torno al tema de la evaluación (Subirats, 1989; Mora, 1991; Alvira, 1991; Comas, 1992; Ballart, 1992; Aguilar y Ander-Egg, 1992; Cohen y Franco, 1993; Grupo GID, 1994; Fernández-Ballesteros, 1995), que con diferentes orientaciones trataba de recuperar el tiempo perdido e informar de lo que se estaba cocinando en otros lugares.

Pero, de otra parte, esta misma mirada se enfrentaba a la propia complejidad histórica de la trayectoria de la evaluación. No era fácil visualizar, de pronto, de entrada y sin ningún aviso previo, todo el complejo mundo de la evaluación, por lo cual lo más frecuente era limitarse a adoptar el punto de vista de un autor o una corriente concreta y proponer una estrategia de evaluación limitada a dicha propuesta.

Pero esta falta de madurez se complicó con la irrupción de las “obligaciones de evaluación” de los programas financiados con fondos europeos, unas obligaciones que, además, algunas administraciones españolas adaptaban de una forma mimética. Las causas que explican esta situación son sencillas: de una parte ya hemos visto nuestras carencias en forma de formación e información sobre el tema; de otra, en Europa también hemos visto cómo la evaluación se desarrolló a partir de la noción de “crisis del Estado de Bienestar”, lo que supuso la creación de un modelo de evaluación muy ligado a la cuestión del control de los gastos (del gasto público en particular). A la vez, España era el país de la UE que percibía más fondos europeos y, por tanto, “el más vigilado”. Como consecuencia, el modelo de evaluación resultante en España era, y es, una opción

muy burocrática, compleja y en ocasiones inteligible, que supone mucho trabajo y obligaciones y que “no devuelve nada” al equipo que realizó la intervención. De hecho son evaluaciones que nadie sabe exactamente qué es lo que hace con ellas la Comisión Europea.

Con tales antecedentes no es extraño que en España la noción de evaluación produzca reacciones paradójicas. De una parte, los profesionales entienden, en términos teóricos, que la “evaluación es una necesidad práctica” para mejorar su trabajo, pero de la otra tienen la experiencia de que las evaluaciones sólo les exigen más trabajo, no les aportan ni les devuelven nada y además suponen una amenaza, que nunca se concreta, en torno a la continuidad y el futuro del programa. Por este motivo, se suelen además encargar “evaluaciones” que se utilizan básicamente para afirmar que “el programa ha sido evaluado y se ha demostrado que sus resultados han sido positivos”.

También es cierto que, en los últimos diez años, sobre este variado panorama comienzan a aparecer tendencias que indican una cierta madurez. La evaluación comienza a ser comprendida y aceptada por las administraciones y se ha alcanzado una discreta, pero suficiente “masa crítica” de profesionales de las evaluaciones, que conforman **una cultura** capaz de ofertar en España un conjunto consolidado de conocimientos en torno a la evaluación fuerte.

Asimismo, el desarrollo institucional de la evaluación comienza a ejercer una influencia positiva. De una parte tenemos el actual Instituto de Evaluación del Ministerio de Educación y Ciencia, heredero de la más antigua y relevante de nuestras tradiciones evaluadoras, aunque representa una respuesta institucional y cerrada a la “evaluación espontánea en los ámbitos educativos”, que ya alcanzó su propia “masa crítica de evaluadores en Educación” en los años 80 (García, 2003), lo que explica que el tema de la evaluación fuera uno de los componentes esenciales de la implantación de la LOGSE (alumnado, profesorado, centros, modelos educativos, administración educativa,...) generándose un intenso “estilo evaluativo” que se ha mantenido en la actual LOE (Zufiaurre y Albertín, 2006). Por otro lado el INCIE difunde numerosos estudios y ha creado un buen Sistema Estatal de Indicadores en Educación.

Por otra parte, los programas europeos también se han modernizado y han abierto las posibilidades de evaluación. Así en Juventud el programa “Juventud con Europa (1995-1999)”, ya se pudo evaluar siguiendo una estrategia abierta según el tipo de programa realizado (INJUVE, 2000). Un proceso que parece estar cambiando a fondo los planteamientos tradicionales de la Comisión Europea (Viñas, 2001).

Finalmente la creación de la Agencia Estatal de Evaluación de Políticas Públicas y Calidad de los Servicios del Ministerio de Administraciones Públicas, que inició como tal sus actividades el 1 de enero de 2007 (a partir de la Ley 28/2006), representando, por fin, una apuesta decidida a favor de la promoción de una cultura evaluativa en España. El Informe de la Comisión de Expertos, que entre 2004 y 2006 revisó la situación en España y realizó una serie de valientes propuestas en torno a la Agencia, refleja un antes y un después de la evaluación en nuestro país. Aparece en el Informe una descripción sin tapujos de los problemas de la implantación de la evaluación en España, reflejada mediante una toma de posición cautelosa por el temor que suscita la mera palabra evaluación en nuestro país (por razones más que justificadas); al mismo tiempo que apuesta sin prejuicios por una noción de evaluación fuerte y realista en términos de gobernanza, innovación y modernización (Garde, 2007).

Pero para lograr un adecuado grado de credibilidad la Agencia Estatal debe superar algunos retos. El primero de ellos es el mismo que aparece en el cuadro 1.1, que, aunque se refiere al ámbito internacional, es de clara aplicación en España y que he enunciado como *“las estructuras evaluativas y administrativas previas quieren mantener su poder corporativo”*, las cuales, además en España, han desprestigiado la evaluación convirtiéndola en un sinónimo de control y hasta de despotismo institucional. El segundo reto se refiere a una necesaria distinción operativa entre evaluación y calidad. Un tema al que voy a referirme en extenso más adelante.

Conviene también señalar que en España la pluralidad territorial y la distribución de competencias en los niveles autonómicos y locales inciden sobre la cuestión de la evaluación. El modelo

de la evaluación fuerte es el más potente en el nivel internacional (pero está, por ejemplo, poco representado en las instituciones de la Unión Europea y en algunos países concretos) y también en el conjunto del Estado. Pero a la vez distintos modelos particulares y corporativos pueden mantener una importante presencia en ámbitos locales, por ausencia de otras alternativas, incluso a pesar de que la evaluación fuerte esté cada vez más diseminada. Un cierto desconocimiento de la pluralidad de opciones en la evaluación ha provocado una cierta tendencia a recurrir “al equipo académico (o a la empresa de de servicios) más próximo”, que en ocasiones puede resultar un tanto peculiar. Esta situación será cada vez menos frecuente por la creciente diseminación institucional y académica de la evaluación fuerte, pero esto no evita que en algunos territorios los ámbitos de juventud puedan tener dificultades a la hora de apostar por las propuestas de este modelo de evaluación.

En cualquier caso el grado de ajuste al conjunto de propuestas (diversas entre si) que aparecen en el libro coordinado por Rafael Bañón (citado como lectura recomendada en el prólogo), permite valorar si una determinada propuesta de evaluación resulta “muy excéntrica” en relación a los modelos comúnmente aceptados de evaluación. Este Manual también debería contribuir a este fin.

Este resumen histórico nos permite entender que la **evaluación fuerte** es el resultado del feliz encuentro entre diversas tradiciones intelectuales y campos de conocimiento. Se sitúa en una gran caja institucional repleta del aire fresco que aportan las nuevas ciencias políticas y en particular la noción de gobernanza y buen gobierno. Pero en la caja encontramos otros objetos. El orden general y conceptual lo establece la pedagogía a partir de la experiencia que han aportado las sucesivas reformas de la educación del siglo XX. Pero los objetos clave (diseño, métodos de investigación y análisis de datos) los aportan economía, psicología y sociología. En este lugar de confluencia se gesta en las décadas de los años 70 y 80 (y nace a principios de los años 90), un **nuevo campo de conocimiento** científico, autónomo, dinámico y necesario, que se llama evaluación.

### 1.3.- La función de la evaluación.

¿Cuál es la función de este nuevo campo de conocimiento? En uno de los más importantes textos para la historia de la evaluación y que hemos recomendado en el prólogo, Carol Weiss explicaba que los usos de la evaluación se pueden resumir en cuatro (Weiss, 1998):

1. EL USO INSTRUMENTAL: en el que se utilizan los resultados obtenidos para tomar decisiones concretas en torno al programa o la política. Se trata de decisiones del tipo “esto va mal y tengo que cambiarlo” o “esto va bien y voy a mantenerlo”. Asimismo, en algunos casos se plantean varias opciones y se puede elegir la más correcta. Este uso instrumental aparece, en la mayor parte de las ocasiones, subsumido y formando parte de los otros usos citados a continuación.
2. EL USO CONCEPTUAL: en el que los resultados sirven para dar explicaciones sobre el porqué o el cómo se hacen las cosas. Por ejemplo, “*nos dirigimos a este colectivo porque es el que tiene más necesidades o del que obtenemos una mejor audiencia,*” o a la inversa, “*no realizamos tal actividad o no incluimos tal logotipo porque provoca el rechazo de los jóvenes que constituyen la población diana*”. Este uso conceptual tiene que ver en ocasiones con la propia misión y los conceptos que sostienen de la intervención, por ejemplo, “*el programa requiere un mecanismo de participación para ser efectivo*”.
3. EL USO PERSUASIVO: en el que la evaluación sirve para movilizar apoyos para conseguir recursos o cambios en el programa. Por ejemplo, “*la evaluación nos demuestra que el programa está siendo muy bien valorado por la opinión pública y la Concejalía o la Caja de Ahorros harían bien en apoyarlo mejor*”. En otro caso, “*la evaluación nos muestra que ampliando los horarios del programa de “hipoteca joven” la audiencia se duplicaría*”.

4. EL USO POLÍTICO O TRASCENDENTE: en la que se pretende influir sobre la toma de decisiones políticas o en el cambio de las creencias sociales. Por ejemplo, *“la evaluación demuestra que la inversión en programas de empleo resulta rentable en términos económicos e incluso reduce los costes de las administraciones públicas porque evita otros problemas”*; o a la inversa, *“la inversión pública en vivienda no conviene porque las ayudas contribuyen a su encarecimiento”*.

En otro texto propio, que también se inspiraba en Carol Weiss, señalaba cuatro razones, muy similares a las anteriores, que explicaban las funciones de la evaluación, aunque en este caso el texto iba dirigido muy directamente a los Agentes de Igualdad que trabajan en el ámbito institucional de los programas de igualdad de género (Comas, 2005):

“Se quiere saber algo por algún motivo, casi todos los motivos son legítimos, pero en el caso de la evaluación aparecen cuatro tipos de motivación, entre los cuales, en general, debemos elegir uno sólo o una parte del mismo, aunque también es posible una motivación más particular que combine diferentes tipos de motivos.

1. **MOTIVOS SOCIOPOLÍTICOS E INSTITUCIONALES:** *se trata de dar respuesta a preguntas que se relacionan tanto con nuestra condición subsidiaria como con la lógica jerárquica de los planes de Igualdad. La mayor parte de las veces las preguntas están ya formuladas e incluso existen cuestionarios a los que debemos responder. Pero en ocasiones no son explícitas y hay que suponerlas en el contexto de una estrategia para garantizar la continuidad de la intervención. En general las preguntas responden en este caso a cuestiones de eficacia, costes y justificación de los recursos empleados.*
2. **MOTIVOS CONCEPTUALES Y CULTURALES:** *son los más personales, ya que en este caso lo que se trata es de valorar, para sí mismo, para el equipo técnico, para la organización que ha impulsado el proyecto o para los participantes, algunos elementos que permitan visualizar los efectos de la intervención. Por ejemplo en un proyecto relacionado con la violencia de género nos puede interesar saber si algunos agentes sociales se han sentido aludidos y que actitud van a adoptar.*
3. **MOTIVOS METODOLÓGICOS:** *tienen que ver con la mejora de los programas, para lograr que sean más eficaces a partir de la experiencia que nos ofrecen los anteriores programas.*
4. **MOTIVOS CORPORATIVOS:** *se supone que los diferentes grupos profesionales, en este caso los propios Agentes de Igualdad, necesitan para poder actuar con eficacia saber lo que ha ocurrido en otros programas. En este caso evaluación se relaciona muy estrechamente con formación.*

En ambos casos resulta evidente que la noción ingenua de evaluación ha sido sustituida por una noción mucho **más realista** y pragmática. La evaluación no es una práctica que aporta resultados neutrales, sino una parte de los propios procesos sociales. Es decir, se trata de una práctica que se basa, de forma ineludible, en el método científico, pero que a la vez pertenece, también de forma ineludible, al campo de una realidad social compleja que la determina y limita. Los “experimentos sociales”, como el ejemplo mencionado en el prólogo, pueden ser científicos y ofrecer datos empíricos perfectamente fundados, pero no cumplen, ni pueden cumplir las condiciones de aislamiento del laboratorio y además, por muy fundamentados que estén, siempre están sometidos a la lógica del debate y del conflicto social. El carácter de “ciencia” lo proporciona, en este caso, aceptar la realidad de este contexto social, en cambio la “idealización científica”(suponer que la sociedad puede ser un laboratorio) sólo es, utilizando de nuevo un término de Mario Bunge, mera pseudo-ciencia.

La propia Carol Weiss tenía claro que esta opción iba a provocar algunas resistencias y afirmó (Weiss, 1973) que *“la evaluación es una tarea racional pero que se lleva a cabo en un contexto político”*, un contexto inevitable en el que hay que tener en cuenta que: 1) los programas sociales son siempre el resultado de decisiones políticas y estas decisiones se someten a consulta electoral en los países democráticos; 2) la evaluación (y sus resultados) forma parte de la confrontación

política; 3) la evaluación, las buenas evaluaciones, son actos políticos porque tienden a mostrar quién tiene razón y quién se equivoca, lo que implica que pueden legitimar (o deslegitimar) las decisiones políticas.

Una parte de la resistencia a esta visión realista de la evaluación procede de los propios profesionales de los programas que prefieren apostar, desde los diversos campos disciplinares, por una imagen aparentemente neutral e ingenua de la evaluación. Se supone que así “no se meten en política” y “obedecen el código deontológico” de su profesión, que les garantiza la neutralidad y les deja al margen de los conflictos políticos. Nada menos cierto. Porque esta es una afirmación profundamente política, e incluso una cierta apuesta por un “sistema político corporativo” que, al menos en España, no podemos asociar con la democracia.

Además, el hecho de que cada una de estas imágenes corporativas contenga elementos antagónicos y se proclame a la vez “ciencia neutral” debería preocuparnos. Como explicó en su día Bertrand Russell (en 1954 en “**Ciencia, Filosofía y Política**”), “*la proclama de la neutralidad ideológica suele ser un síntoma de la presencia de componentes ideológicos*”. Más próximo a nuestros días podemos recordar a Gonzalo Fernández de la Mora, uno de los más reputados “intelectuales” del régimen franquista, que llegó a afirmar en 1965 en “**El crepúsculo de las ideologías**” que “*España es un país donde impera la razón y la verdad en un mundo en el que todos los demás países se organizan en función de diferentes ideologías*”, añadiendo que “*Nuestro sistema político es el más avanzado del mundo porque ha sido el primero que ha podido prescindir de las ideologías en su organización*”. Pues parece que esto mismo pasa en algunos modelos de evaluación que parecen olvidarse de que la política existe, de que la política democrática es necesaria, de que los programas y las intervenciones sociales son “políticas”, y de que los que trabajan en los mismos forman parte de un proyecto político<sup>10</sup>.

#### 1.4. La estrategia evaluativa.

El concepto fuerte (y la definición) de evaluación tiene un punto de partida muy claro en la estrategia que planteó Immanuel Kant para realizar una **evaluación de la ciencia**, en especial de las ciencias físicas de Newton, en el nivel en el que se encontraban a fines del siglo XVIII en el momento álgido de la ilustración (Kant, 1783). Se trataba de una evaluación que partía del llamado **problema de Hume**, referente a si la inducción resulta probatoria<sup>11</sup> y que Popper ha planteado en extenso y le ha dado una nueva respuesta (el falsacionismo)<sup>12</sup>. La cuestión es muy compleja, pero aquí lo importante es resaltar que Kant trataba de evaluar las obras de Hume y de Newton en aquel particular momento histórico y que para hacerlo se planteó resolver previamente tres cosas: A) establecer lo que quería hacer, B) determinar lo que quería saber, C) identificar los medios (las maneras) de hacerlo para obtener un producto concreto D), que en este caso fue la “**crítica de la razón pura**”, cuya difusión E) garantizó un óptimo de certidumbre para la Ciencia en los ambientes institucionales.

Conviene rescatar el planteamiento original de Kant por diversas razones. Primero para situar la evaluación en el contexto científico, e incluso reivindicar a Kant como evaluador. La segunda

<sup>10</sup> Aunque es cierto que en un determinado nivel hay que apostar por la neutralidad técnica de la administración, pero en esta apuesta, afirma Carol Weiss, no se puede ser ingenuo.

<sup>11</sup> La “inducción” es el tipo de razonamiento mediante el cual, a partir de varias proposiciones particulares obtenidas de forma experimental y que resultan coincidentes, se deduce una proposición universal. Hume, el promotor del empirismo, planteó sin embargo que no existe “conexión lógica” entre las repetidas medidas particulares y la proposición universal, lo que invalidaba la categoría “ciencia” tal y como la comenzaban a utilizar los físicos. Fue Kant quien dio una respuesta, que aún sigue siendo la que se utiliza en las Ciencias de la Naturaleza, mostrando mediante argumentos racionales la validez de la inducción.

<sup>12</sup> Para Karl Popper la inducción no posee ninguna base lógica, pero como procedimiento científico es el más operativo, el más útil. Por este motivo las proposiciones universales, lo que ahora se llaman evidencias científicas, tienen que ser “falsables”, es decir, deben poder modificarse a partir de nuevas pruebas inductivas, de nuevos resultados, que permiten “superar” anteriores proposiciones universales. Una proposición (una evidencia) que no permita ser falsable no es, por tanto, científica.

porque en la actualidad son muchos los supuestos evaluadores que inician su tarea a partir de la tercera cuestión (es decir el “como”), porque dan por sabidas las dos primeras preguntas, pero en realidad esta es una manera de imponer visiones implícitas del “qué” y del “para qué”. En territorios como las políticas de juventud, este “qué” y este “para qué” representan además preguntas especialmente pertinentes y muy sensibles, porque determinan la orientación global de las políticas de juventud.

Por tanto, como hizo Kant, vamos a plantearnos un tipo de evaluación que tenga en cuenta qué queremos hacer (A), para qué vamos a hacerlo (B), para pasar a continuación a plantear cómo vamos a hacerlo (C), antes de hacerlo efectivamente (D) y decidir finalmente lo que vamos a hacer con nuestros hallazgos (E). Pero dicho así, este esquema corre el riesgo de confundirse con una típica lógica lineal, cuyo rechazo constituye un componente esencial de la estrategia de la evaluación fuerte.

En realidad el proyecto evaluativo de Kant se corresponde literalmente con las propuestas que ya he citado de Michael Scriven (aunque no lo cite), y esta correspondencia nos permite explicar cómo el rechazo a la lógica lineal (se hace lo que sea y después se evalúa) constituye uno de los ejes que articulan la noción de evaluación fuerte y de las estrategias evaluativas realistas en las que la evaluación es parte de la intervención. Se trata de una cuestión compleja que trataré de sintetizar, aunque esta síntesis puede resultar un reto duro para algunos lectores.

El gráfico 1.2 muestra una serie de correspondencias conceptuales entre la estrategia kantiana de evaluación, la concepción de la evaluación fuerte, la estrategia de planificación de las políticas sociales y el modelo del diseño experimental científico. Dicho gráfico introduce además una serie de términos que iremos explicando en las páginas siguientes.

En la primera columna se resume la actuación de Kant, cuando evaluó un programa ya concluido y cerrado, el programa que había llevado a cabo Newton hacía ya casi un siglo y que Hume había justificado poco antes. Por tanto, Kant se planteó lo que se llaman “preguntas post-facto”, es decir, las que se hacen después de que un acontecimiento ocurra o un programa se haya realizado. Por tanto, si nos situamos en el terreno de las estrategias de evaluación clásicas no hay correspondencia entre las fases de “planificación” e “intervención” y las correspondientes preguntas del gráfico 1.2, porque Kant se hizo todas las preguntas después de que Newton y Hume hubieran pensado y escrito sus textos, es decir, fue en el “momento posterior” cuando realizó la evaluación.

Pero a la vez Kant en los “*Prolegómenos a toda metafísica futura que querrá presentarse como ciencia*” (Kant, 1783), insiste en que planteó post-facto la “*Crítica de la razón pura*” (1781),

Cuadro 1.2. **La estrategia de la evaluación.**

ESQUEMA DE KANT	POLITICA SOCIAL	CIENCIA EXPERIMENTAL	EVALUACIÓN CLÁSICA	EVALUACIÓN ACTUAL
¿Qué queremos hacer? ¿Para qué vamos a hacerlo? ¿Cómo vamos a hacerlo	Necesidades objetivos y recursos	Material y método	Planificación de la acción o la política	Análisis de la realidad, diagnóstico y evaluación de diseño
Hacerlo efectivamente	Actividades	Experimento	Intervención	Evaluación de implementación y procesos
	Producto	Hallazgos y comparación	Evaluación	Evaluación de resultados e impacto
¿Qué haremos con los hallazgos?	Memoria	Publicación científica	Informe de evaluación	Registro y difusión de resultados

Fuente: elaboración propia.

porque no le quedaba más remedio que hacerse tales preguntas **después** de que el proyecto hubiera sido desarrollado, es decir, después de que Hume y Newton hubieran concluido y cerrado sus trabajos, pero que lo correcto es que la ciencia, la futura ciencia que va a tratar de fundamentar, debería hacerse tales preguntas no después de concluir el programa, sino en el momento oportuno de la investigación o la intervención.

Afirma Kant que al plantear la evaluación de los programas como un concepto “ex-novo” (es decir, inédito hasta entonces y novedoso) no le queda más remedio que hacerlo “después” de que Hume y Newton hubieran “realizado sus programas”, pero que una vez establecida la necesidad de la evaluación, hay que evaluar distintas cosas en los diferentes momentos del proceso. Es decir, las preguntas sobre “qué queremos hacer” y “para qué vamos a hacerlo” debe formularse al evaluador antes y durante la intervención y no después de la misma. De hecho, insiste Kant, lo que distingue la ciencia de la pseudo-ciencia es el momento del proceso de investigación en el que se formulan tales preguntas y propone, para el futuro, un modelo de ciencia empírica, en el que la formulación previa de las tres preguntas nos permita contextualizar los hallazgos y las evidencias. Es más, añade, no hay hallazgo ni evidencia válida sin estas preguntas previas<sup>13</sup>.

El avance de la ciencia y la posterior constitución del área de conocimiento de la evaluación se olvidaron de las aportaciones de Kant en una gran medida porque, hasta fechas recientes, en Estados Unidos, los “filósofos europeos” (O como dicen los anglosajones: “los filósofos continentales”) eran considerados como señores raros dados a la especulación creativa y muy poco científicos (aunque ahora todo ha cambiado y en los ámbitos anglosajones de la tradicional “ciencia inductiva” se ha descubierto de pronto el enorme potencial de estos “filósofos”). Como consecuencia, la evaluación fue considerada siempre una tarea post-facto, principalmente porque los “científicos” no leían a Kant, al que se consideraban un “pensador” que hablaba de temas que carecían de interés para ellos. Por su parte, los filósofos siempre trataron de olvidar la conexión entre Kant, Hume y Newton, porque ellos a fin de cuentas eran filósofos y por tanto propietarios de un Kant al que dotaron de un aire melancólico de trascendencia, que para nada debía preocuparse de estos temas pertenecientes a otras áreas del conocimiento. Menciono esta cuestión como un ejemplo de los inconvenientes de las burbujas especializadas y las ventajas del trabajo multidisciplinar.

Sin embargo, la evaluación fuerte recupera la estrategia kantiana (y algo parecido debería hacer toda la investigación científica) y hace preguntas en diferentes momentos.

Las preguntas sobre lo que queremos hacer aparecen en una primera fase de planificación y en dicha fase conviene ya pensar en evaluación, una evaluación que es el análisis de la realidad, el diagnóstico y la evaluación del diseño. Se trata de tipos de evaluación que nos informarán en torno a este “querer hacer”, a si es más o menos correcto, a si deberíamos modificarlo y en qué sentido, y que en todo caso nos permiten justiciar de manera adecuada porqué hemos elegido ésta y no otra misión o tarea.

La pregunta sobre “para qué vamos a hacerlo” parece cabalgar entre la planificación y la intervención, pero para Kant el “para qué” de la evaluación no es el “para qué” de la intervención. El “para qué” de la intervención son los objetivos de la misma, el “para qué” de la evaluación se refiere también a sus objetivos, pero desde la perspectiva de la evaluación los objetivos de la intervención son variables independientes, mientras que los objetivos de la evaluación son variables dependientes.

Un ejemplo para entender esta especie de galimatías. Supongamos un programa de movilidad de un municipio que incluye campamentos y “granja-escuelas” en España, campamentos internacionales de intercambio y gestión de viajes internacionales para jóvenes. Los objetivos del programa pueden referirse a cosas como “facilitar el contacto con la naturaleza a jóvenes urba-

<sup>13</sup> El gran guru del management, Peter Druke, resulta, aunque a muy pocos se les haya ocurrido un practicante estricto de la estrategia kantiana. Sus famosas cinco preguntas previas, que tanto han influido en la conformación del actual pensamiento empresarial (así como la gestión pública y la de las ONGs) son precisamente: 1) ¿cuál es nuestra misión?, 2) ¿quién es nuestro cliente?, 3) ¿qué valora el cliente?, 4) ¿qué resultados hemos alcanzado?, 5) ¿cuál es nuestro plan? (Druke, 1995), son Kant puro y duro.

nos”, “realizar aprendizajes de convivencia con otros adolescentes de las mismas edades”, “tener experiencias en participación y toma de decisiones colectivas”, “conocer otras culturas y otras realidades” y “desarrollar el sentido de la solidaridad y la tolerancia cultural”.

Para este momento de la evaluación tales objetivos vienen dados, aunque el trabajo evaluativo anterior, por ejemplo el análisis de la realidad, haya ayudado a formularlos en términos de “necesidades de los jóvenes” o “deseos no alcanzados de los jóvenes”. Pero una vez formulados tales objetivos son “variables independientes”, es decir, inamovibles en su formulación, mientras el programa continua en vigor y no se revisa. Lo cual puede ocurrir en cualquier momento a consecuencia de un cambio político (y en algunas ocasiones por los resultados de la evaluación), pero mientras tanto siguen siendo variables independientes.

En cambio, las variables de la evaluación se corresponden con indicadores dependientes, porque sólo se pueden formular a partir de los propios objetivos del programa, por ejemplo “*¿Cuántos jóvenes han participado en el campamento X, de qué edades y de que extracción social?*”; pero también “*¿Han cambiado las actitudes de los participantes en relación a la auto-imagen de otras culturas?*”. Cada uno de estos indicadores se desarrolla a través de otro “indicador” (bien cualitativo o bien cuantitativo) que establece un estándar de medida, por ejemplo en el primer caso sería: “*incrementar anualmente (quizá en un x%) el número de jóvenes participantes*”, y en el segundo: “*mejorar las puntuaciones de tolerancia cultural en una escala pre-establecida*”. Finalmente estos indicadores se formulan a modo de “estándares medibles” mediante algún método de recogida y análisis de datos, por ejemplo en el primer caso mediante un registro y en el segundo mediante una encuesta. Parece un lío, pero tenemos todo el Manual para repetirlo y entenderlo.

En este punto debemos limitarnos a considerar que si observamos bien estos dos indicadores, y utilizando el lenguaje de los evaluadores, nos daremos cuenta de que el primero nos va a facilitar información sobre la implementación y el proceso, mientras que el segundo nos va a facilitar información sobre los resultados o el impacto.

Se supone que el primero mide las cosas mientras están ocurriendo y se limita a proporcionar datos sobre el cumplimiento cuantitativo de los objetivos. Por ejemplo, “conseguimos mandar al campamento un total de 140 chicos y chicas entre 15 y 17 años, en la primera tanda fueron 30, en la segunda 60 y en la tercera 40”. Mientras que el segundo mide las cosas después de que han ocurrido y nos proporciona una valoración de la intervención. Por ejemplo, “un 49% indicó que su imagen de la cultura (o las culturas o las etnias) tal, había cambiado profundamente, un 26% que había cambiado algo, un 20% no habían modificado sus creencias previas y un 5% no contestó”.

Pero esta es una clasificación muy confusa porque podemos decir que evaluamos el proceso cuando comprobamos o constatamos el número de adolescentes que acude al campamento. Pero ¿no podría ser esto también un resultado del programa? Pues sí podría serlo y, de hecho, la mayor parte de las evaluaciones de programas en el ámbito de juventud y casi todas las memorias llaman “resultados” a estas cifras de “audiencia”. Pero para la mayoría de los evaluadores no son resultados, sino procesos. ¿Por qué? Pues porque nos limitamos a constatar que la actividad se ha realizado de acuerdo con lo previsto en los objetivos (y seguramente en los presupuestos que preveían dinero para 140 desplazamientos). Para el técnico de juventud es, en cambio, un “buen resultado” porque ha conseguido ajustar la asistencia a los recursos presupuestados y además no le ha sobrado nada.

Tomemos el segundo indicador, el relacionado con el cambio de actitudes, y ahí parece que no hay problema porque está claro que son datos relacionados con resultados o bien con impacto. Sin embargo, el ejemplo elegido (con un poco de mala intención por mi parte), nos conduce hacia una cierta paradoja, ya que ¿el cambio de actitudes es hacia mayor tolerancia o hacia mayor intolerancia? Porque puede que los jóvenes involucrados procedan de un ámbito en el que predominan una serie de estereotipos negativos hacia esta otra cultura, entonces seguramente la idea del cambio se refiere a la superación de tales estereotipos a partir de una aproximación coti-

diana. Pero también puede ser, y de hecho es un caso cada vez más frecuente, que la convivencia sea con un extraño al que se tiene idealizado. Como en general la idealización del otro es una exageración de nuestros propios valores que cultivan los medios de comunicación (como parte de un compromiso para evitar el racismo), la cercanía cotidiana puede suponer la construcción de una imagen negativa de este otro.

Luego la evaluación no nos dice nada en relación al objetivo, quizá implícito de “mejorar el conocimiento y la imagen de jóvenes de otra etnia”, porque quizá ha mejorado el conocimiento pero en cambio la imagen positiva se habrá deteriorado.

¿Dónde está el problema? Pues la respuesta está en otra pregunta ¿Cómo podemos diferenciar los tipos de evaluación? ¿Simplemente diciendo que se trata de evaluación de procesos y de una evaluación de resultados? Pues no, porque si **sólo** hacemos una evaluación, los “procesos” y los “resultados” son fácilmente intercambiables. Dependerá del uso que queramos darles, de la función que les atribuyamos. Pero si convenimos que la evaluación es “una estrategia”, la estrategia Kantiana, la estrategia de Scriven, ocurre que “procesos” y “resultados” nos parecen cosas diferentes porque en la estrategia hemos tenido que diferenciar ambos momentos, ambos tipos de evaluación. Es decir, si evaluó de acuerdo con la lógica lineal que aparece en la columna de la evaluación clásica, no puedo distinguir los tipos de evaluación que se superponen y se muestran abiertos a cualquier categoría. En cambio, si aplico el procedimiento de la evaluación fuerte, cada tipo se definirá en relación a los otros.

Podemos ahora contestar a la pregunta de ¿dónde está el problema? Diciendo que el problema no está en la formulación del indicador “¿Han cambiado las actitudes de los participantes en relación a la auto-imagen de otras culturas?”, sino en lo que significa dicho indicador, algo que sólo podemos valorar si nos lo hemos planteado en las evaluaciones previas de diagnóstico, análisis de la realidad y diseño, en las deberemos contestar a la pregunta “¿Qué queremos cambiar, por qué y cómo vamos a hacerlo?”. Esto es lo que se preguntó Kant antes de comenzar y así evitó confundirse con los indicadores al evaluar a Newton.

## BIBLIOGRAFÍA Y DOCUMENTACIÓN COMPLEMENTARIA.

Toda esta parte histórica, conceptual y teórica puede completarse con los tres siguientes textos, aunque el tercero es una aplicación específica a las políticas públicas españolas de los años 80:

- HOUSE, E.R. (1980), **Evaluación, ética y poder**, Madrid, Morata, 1997.
- SHAW, I.E. (1999), **La evaluación cualitativa**, Barcelona, Paidós.
- SUBIRATS, J. (1989), **Análisis de políticas públicas y eficacia de la administración**, Madrid, INAP.

También puedo recomendar dos biografías recientes de Issac Newton e Inmanuel Kant, que han accedido a documentación inédita (hasta ahora resguardada por razones testamentarias) y que ofrecen una versión distinta, y más atractiva en términos humanos, de ambas figuras de la historia de la ciencia. En concreto se pueden completar los argumentos manejados en este capítulo y comprender cómo la fundación de la ciencia (la ilustración y la revolución industrial) en el siglo XVIII fue una tarea multidisciplinar. Lo que nos permite especular sobre cómo habría evolucionado la civilización si se hubiera continuado por esta vía y no se hubiera producido el repliegue hacia los campos de conocimiento particulares.

- WESFALL, R.S. (1993), **Isaac Newton: Una vida**, Cambridge, University Press.
- KUEHN, M. (2003), **Kant**, Cambridge, University Press

### 2.1.- El método científico y la evaluación.

La evaluación es un procedimiento científico que se ajusta a las “Condiciones de Ciencia”, que han sido resumidas y sistematizadas de una forma muy clara por Mario Bunge. Representamos en la tabla 1.3 este resumen, a partir del contenido de propuestas más detalladas de este mismo autor (Bunge, 2001 y Bunge, 2003). Se trata de un esquema que vamos a utilizar en otros momentos de este texto pero que además puede servir como orientación para las propias tareas de evaluación.

También cabe añadir que si bien la evaluación utiliza de forma tradicional las técnicas y los instrumentos del campo de conocimiento de las ciencias sociales que se exponen en la parte II del Manual, el método científico nos permite imaginar otras posibilidades.

En todo caso esta dependencia instrumental al área de las ciencias sociales proyecta sobre la evaluación algunos debates, en particular en lo relativo a las relaciones entre la investigación y la evaluación, en la confrontación entre lo cualitativo y lo cuantitativo y sobre la necesidad de optar entre una misión empírica y una misión racional. Se trata de tres debates cuyas conclusiones particulares para el campo de conocimiento de la evaluación le otorgan una parte sustancial de su identidad y algunas peculiaridades. En todo caso, son debates que aparecen en otros campos, aunque el primero de ellos (investigación versus evaluación) parece más propio y exclusivo de la evaluación, pero el segundo se ha producido en el campo de conocimiento de la sociología, mientras que el tercero, muy frecuente en el territorio de las ciencias sociales, se proyecta sin embargo sobre el contexto más generalista de la filosofía de la ciencia.

A pesar de que los tres debates puedan parecer un tanto innecesarios para “la práctica” de la evaluación, lo cierto es que resultan bastante imprescindibles para comprender bien en que consiste la evaluación.

#### 2.1.1.- Investigación versus evaluación.

Las relaciones entre investigación y evaluación son muy claras y se deducen de la propia definición de evaluación. Así, la evaluación **también** utiliza el método científico, pero mientras la investigación pretende alcanzar un fin abstracto de carácter universal: el conocimiento, la contrastación de hipótesis teóricas o su falsación destinados a la comunidad científica; la

Cuadro 2.1. **Campos de conocimiento: (Cc) = < C, S, G, D, F, E, P, K, O, M>**  
**Condiciones de ciencia y pseudo-ciencia en cada componente + <V, T>**

	<b>CONTENIDO</b>	<b>CONDICIÓN CIENCIA</b>	<b>SEUDO-CIENCIA</b>
<b>C</b>	La comunidad de sujetos que conocen Cc.		Una comunidad de creyentes
<b>S</b>	La sociedad que hospeda a Cc.		Tolerada por razones prácticas
<b>G</b>	La concepción general o presupuestos filosóficos de Cc.	Una ontología de cosas inmutables, una gnoseología crítica y no ingenua y la libertad para buscar la verdad	Uso excesivo de imágenes inmateriales y autoritarismo interno
<b>D</b>	El dominio, universo o territorio del discurso de Cc.	Entidades reales	Entidades irreales
<b>F</b>	Un trasfondo de formas: las herramientas que se emplean en Cc.	Un conjunto de teorías lógicas o matemáticas actualizadas	No existen o son ad-hoc
<b>E</b>	El conjunto de presupuestos que provienen de otros Cc.	Una colección de datos, hipótesis y teorías razonablemente bien confirmadas	No existen o son muy reducidas
<b>P</b>	Los problemas que se abordan en Cc.	Referidos al territorio (D)	Sólo incluye problemas prácticos cotidianos
<b>K</b>	El fondo específico de conocimiento acumulado en Cc.	Compatible con (E)	Estático y con evidentes falsedades
<b>O</b>	Los objetivos y metas de Cc.	Incluyen el descubrimiento de leyes relativas a (D) y el perfeccionamiento de (M)	Se limitan a consideraciones prácticas
<b>M</b>	Los métodos que pueden utilizarse en Cc.	Son escrutables y justificables	No son controlables
<b>V</b>		Hay al menos un Cc. contiguo y que confluye con Cc.	No hay otro Cc. cercano, salvo pseudociencia
<b>T</b>		De (G) a (M) cambian con el tiempo como resultado de la investigación en Cc o en otros campos relacionados	No cambian con el tiempo

Elaboración propia a partir de Bunge, 2001 y Bunge, 2002.

evaluación en cambio pretende alcanzar un fin práctico: mejorar una intervención difundiendo sus características particulares y reales entre lo interesados, los afectados y aquellos profesionales que realizan intervenciones similares (Bustelo, 1999).

Durante mucho tiempo ambos procedimientos se confundieron de forma interesada y aún hoy en día algunos autores afirman que “son lo mismo”, pero esto ocurrió en una gran medida porque los primeros evaluadores se sentían “débiles” y para legitimar y afirmar la naciente área de conocimiento utilizaban términos como “investigación evaluativa” (Bustelo, 2003). En la actualidad, aunque no entre los evaluadores, aún sigue siendo muy común confundir investigación

y evaluación para poder afirmar que “sólo hay que aplicar los programas positivamente evaluados”. Pero esto, como ya hemos explicado, refleja un cúmulo de errores en torno a la función de la evaluación y el deseo de transferir el poder corporativo de determinadas áreas de conocimiento e investigación hacia la gestión de programas.

Un ejemplo muy interesante del debate se produjo en el ámbito de Servicios Sociales (Cohen y Franco, 1993). Cuando comenzaron a expandirse los Servicios Sociales en España a partir del Plan Concertado a finales de los años 80, los psicólogos fueron los primeros en hablar de “evaluación” refiriéndose, por las razones explicadas en el capítulo anterior, a la “evaluación del caso” (algunos incluso a la “evaluación continua”) como sustitutivo del “psico-diagnóstico”. Desde el trabajo social se reaccionó y se reclamó la existencia de una “evaluación o un diagnóstico social”, lo que parecía muy lógico. En ambos casos se señalaba (por parte de algunos profesionales) que “el registro de estas evaluaciones permitía realizar proyectos de investigación”, lo cual es cierto. En este sentido, evaluación e investigación se confundían porque en realidad nadie estaba haciendo evaluación, sino sólo diagnósticos individuales o sociales (que pueden ser parte de la evaluación pero no la evaluación).

Un poco más adelante algunos profesionales comenzaron a plantearse la necesidad de hacer “Estudios de Seguimiento”<sup>14</sup> con la finalidad de comprobar la eficacia de las intervenciones a largo plazo (Comas, 1992; Comas, Roldan y Garcia, 1996) ), lo que provocó algunas confusiones con la evaluación, porque, en general, los Estudios de Seguimiento tratan de comparar “una muestra de casos de usuarios” antes y después de la intervención, lo que permite conocer la efectividad diferencial de los servicios o programas y a partir de esta comparación introducir mejoras en los mismos, pero en ocasiones se trata de medidas a muy largo plazo y en general se utilizan sólo para programas psicoterapéuticos y sanitarios de alta intensidad.

Además, se puede afirmar que, por ejemplo, “completar la intervención con un programa de terapia familiar mejoraba el pronóstico”, lo cual constituye una información científica muy válida, pero que no nos dice nada sobre cómo se ha conseguido implementar dicho programa complementario, cómo lo llevan los distintos profesionales, qué casos (y porqué) permanecen excluidos del mismo, qué desplazamientos internos se han producido y qué consecuencias han generado, cómo se consigue que participen las familias que forman parte del programas y, en última instancia, qué convendría hacer para mejorar la intervención. El Estudio de Seguimiento nos facilita datos sobre alguna de estas preguntas y en este sentido es muy útil, pero a la vez no es la respuesta evaluativa precisa a ninguna de ellas y se requiere un trabajo posterior de carácter (aunque sea una mera reflexión en el equipo) para contestarlas.

En todo caso, conviene tener en cuenta que este es un problema que no aparece en los ámbitos de juventud, porque en los mismos no se ha hecho ningún Estudio de Seguimiento y el carácter colectivo y grupal de la mayor parte de las acciones y los planes propicia las estrategias evaluativas más participativas. Pero no estaría de más, para ciertos programas como por ejemplo los de empleo, realizar algún Estudio de Seguimiento.

Finalmente no podemos dejar de citar otra realidad: la acumulación de evaluaciones contribuye a la mejora y ampliación del conocimiento (Zufiaurre y Albertin, 2006), lo que a la vez produce “competencia por el saber” entre la investigación y la evaluación y a la vez crea sinergias, porque el conocimiento adquirido es en todo caso conocimiento, entre ambos territorios.

### **2.1.2.- Cualitativo versus cuantitativo.**

Ya hemos explicado que, entre las diferentes tradiciones que han conformado la evaluación, las ciencias sociales han aportado los métodos y técnicas para la recogida de datos. Esta elección se justifica sobre un argumento irrefutable: la evaluación trabaja sobre el difuso territorio

<sup>14</sup> En algunos casos se llaman Estudios Longitudinales y Estudios de series temporales. La noción de “Estudios de seguimiento” es típica de las áreas administrativas de salud y servicios sociales, un ámbito sectorial donde este tipo de estudios ha tenido un amplio desarrollo. En todo caso lo importante es no confundirlos con los sistemas de seguimiento de casos o programas.

de la realidad social, en el cual resulta imposible (e inadmisibles) reproducir condiciones estrictas de control (salvo algunas estrategias de modelización post reservadas a la investigación básica y que requieren una alta preparación y muchos recursos), aunque a lo largo del siglo XX se han conformado metodologías de trabajo sobre la realidad social y que a la vez garantizan el carácter empírico de los datos obtenidos.

Las ciencias sociales trabajan con colectivos sociales en cuanto a tales. No trabajan con individuos, ni con agregados de individuos, sino con poblaciones definidas y delimitadas que en su totalidad constituyen los sujetos únicos de su intervención. Tales poblaciones se pueden fragmentar en subgrupos a partir de variables, pero siempre serán sujetos colectivos. Las políticas sociales también se dirigen a colectivos sociales y no tanto a los individuos que los conforman, aunque todos ellos, o una parte, pueden resultar beneficiarios de la misma. Pero una política democrática (y las intervenciones que se derivan de la misma) es una acción colectiva. Las políticas que identifican y, por tanto, discriminan a individuos concretos dentro de colectivos sociales, no son políticas democráticas y son moralmente inaceptables (Bourdieu, 1997). Las evaluaciones de políticas, planes, programas y acciones de juventud sólo se realizan sobre políticas democráticas y por tanto sobre acciones colectivas, por este motivo los métodos de recogida de datos son los de las ciencias sociales. La “evaluación de casos” corresponde a otros campos de conocimiento y no tiene nada que ver con la noción autónoma de evaluación que se aplica en la evaluación de programas.

Como consecuencia, uno de los problemas que tiene que afrontar la evaluación se refiere a la herencia de los debates que envuelven el tema de los métodos y técnicas de recogida y producción de datos que se han producido en el ámbito de la sociología. En particular, el debate entre los métodos cualitativos y los cuantitativos.

El debate entre los métodos cuantitativos (singularmente la encuesta) y los cualitativos (singularmente los grupos de discusión) apareció a finales de los años 60 y se generalizó en la década de los años 70. El debate se produjo tras más de un cuarto de siglo de absoluto dominio de la encuesta (ver capítulos 11 y 12), que además había creado una potente industria en el mundo desarrollado a través de las empresas de Estudios de Mercado, Marketing y Publicidad y los Institutos de Opinión Pública y Estudios Electorales. El movimiento metodológico en torno a los grupos de discusión alegaba, con poderosos argumentos, que las encuestas al precodificar las respuestas limitaban las opciones a aquellas que resultaban aceptables para los que detentaban el poder de su gestión y administración, limitando el ámbito de lo posible a lo aceptable. Si se quería conocer la realidad no se podía limitar previamente aquello que se quería conocer. Por su parte los partidarios de los métodos cuantitativos alegaban que sin una base numérica y estadística no se podía tratar de forma objetiva ningún fenómeno, al analizar grupos no representativos y al no recoger los datos para transformarlos en cifras, se decía, lo que en realidad se hace es proyectar la subjetividad del analista.

Sin embargo, desde principios de los años 90 el debate se fue volviendo un tanto artificial. La oposición entre ambos métodos dejó paso a la noción de complementariedad y la idea de que una “dualidad metodológica” ya sólo la sostenían aquellos que tenían interés en que el debate perviviera. En una investigación publicada hace ya muchos años por el propio INJUVE, mostré la posibilidad de combinar ambas metodologías de forma muy productiva apelando a la complementariedad metodológica (Comas, 1993). La referencia y la propuesta comenzó a ser recogida en diversos manuales y la idea de la complementariedad comenzó a tomar cuerpo (Valles, 1997; Rubio y Varas, 1996 y González, 1997).

En este contexto fue Alfonso Ortí quien dio el paso definitivo, poniendo en evidencia las falacias de la supuesta dicotomía: *“Así pues, la dicotomización radical de ambos enfoques, en cuanto pretenden una absolutización excluyente, en lugar de reconocer sus respectivas limitaciones y mutua complementariedad (por deficiencia), tiende a concluir -por ambas partes- en el refugio en un metodologismo estéril, tanto más irrelevante para la praxis de la intervención social cuanto más*

*riguroso –en un plano abstracto- se pretende. Y en este sentido, inspirado en un modesto realismo metodológico, y en la investigación social realmente existente, lejos de toda pretensión imperialista, la adecuada comprensión de las posibilidades y límites, tanto de la perspectiva epistemológica y de las técnicas cuantitativas, como de la propia perspectiva epistemológica y de las prácticas cuantitativas de investigación social, pasa por el honesto reconocimiento de su radical deficiencia en la representación y análisis de la realidad social” (Ortí, 1995).*

La idea de la **complementariedad por deficiencia** entre ambos enfoques metodológicos que plantea Alfonso Ortí va a ser ampliamente desarrollada en el capítulo tres. De hecho, este es otro de los componentes esenciales de la evaluación fuerte tal y como la describimos en este manual.

### **2.1.3.- Racional versus empírico.**

Ésta es una cuestión compleja cuya comprensión siempre ha exigido un cierto entrenamiento en filosofía del conocimiento o en epistemología. Sin embargo, se puede realizar una aproximación sintética y al mismo tiempo útil para los usuarios del Manual.

El debate entre lo racional y lo empírico hunde sus raíces en la filosofía griega, pero aún es un tema no resuelto en el campo de conocimiento de la epistemología que fundamenta la condición de la ciencia. Incluso en la actualidad perviven posturas extremas como la que sigue enfrentando la ciencia cognitiva de Chomsky (racional) con el conductismo experimental (empírico) de Skinner. Los primeros, que incluyen a Platón, a Descartes y a los constructivistas modernos, consideran, expresado de una forma muy elemental, que la realidad está en la mente del sujeto (o pertenece a una categoría superior y trascendente sea Dios, el bien, la bondad o la verdad) y que con el mero ejercicio de la reflexión podemos llegar al óptimo de conocimiento.

Los segundos, los empiristas, entienden en cambio que la mente del sujeto es un “gabinete vacío” (Locke) que la experiencia y el conocimiento del mundo van estructurando. Así, el conocimiento pertenece a la realidad del mundo y sólo podemos “aprender” a través de las experiencias. La ciencia experimental, y no la reflexión, es la única manera de acercarnos al mundo. Los citados Hume, Newton y Kant representan esta línea de pensamiento. Debe quedar claro que tanto racionalistas como empiristas apelan a la “búsqueda de la razón” (y lo razonable y las certezas o las evidencias), sólo que cada uno la busca en lugares diferentes (el propio sujeto/el mundo).

La evaluación en si no entra en este debate, que tal como lo he expuesto parece la respuesta a un examen de bachiller, pero la cuestión forma parte del “contenido” de la evaluación, en una gran medida porque, de una manera tautológica, la evaluación es por si misma y de una forma inevitable una práctica empírica. Expresado en otros términos, ocurre que por su propia definición y funciones la evaluación pertenece al campo de las prácticas empíricas, lo cual la ubica de forma clara dentro del sistema conceptual de los empiristas.

¿Se puede ser racionalista y a la vez realizar evaluaciones? Sin duda, aunque para ello es necesario respetar la práctica empírica de las mismas. Se puede por tanto ser racionalista, pero hay que aceptar un “empirismo” de lo que se llama “segundo orden”, es decir, el empirismo práctico del que se alimenta la evaluación. En todo caso, conviene aclarar que para los evaluadores no resulta aceptable del racionalismo su “afición” por proyectar “mundos ideales”, sociedades utópicas o perfectas, tan queridas por los racionalistas, desde “La Republica” de Platón a “Walden” de Henry David Thoreau, pasando por “La Ciudad de Dios”, de San Agustín, “La Ciudad del Sol”,

de Campañella o la propia “Utopía”, de Tomas Moro<sup>15</sup>. Aunque también es cierto que el mismo Skinner, deseando emular (y quizá poner en evidencia que no queda claro en el texto) a los racionalistas, escribió “Waldem II”.

Todas estas referencias cultas pretenden poner en primer plano un componente importante de la evaluación: no se trata de proponer ideales utópicos para valorar si se han alcanzado, sino sólo de saber lo que ocurre como consecuencia de determinadas acciones humanas, en las sociedades reales. Expresado en otros términos, los objetivos ideales o muy ambiciosos no son un buen criterio para determinar el alcance o el mérito de un programa, porque las intervenciones sociales se valoran por lo que hacen, no por lo que dicen que deberían hacer.

## 2.2.- Terminología.

En los apartados anteriores se han utilizado algunos términos de uso común y que se supone conocemos, pero me ha parecido interesante facilitar algunas definiciones de los mismos que ayuden a clarificar el texto. Pero además aparecen otros términos propios de la evaluación que también resulta conveniente conocer. Hay que aprender a utilizarlos y a saber qué significan, porque el manejo de esta terminología es ya una exigencia profesional y de competencia personal. Conviene también señalar que los “tipos de evaluación” van a ser ampliados en próximos capítulos. También podemos resaltar la importancia de tres términos (eficiencia, efectividad y eficacia), que están en el centro de todos los debates en torno a evaluación y cuya adecuada comprensión resulta esencial. Como esta comprensión en Español no es fácil, y además no coincide con la semántica formal de estos tres términos, se ha realizado un cierto esfuerzo para evitar “que nos alcancen dogmatismos terminológicos falsos y rígidos”.

- **CAPACIDAD:** [DRAE: Propiedad de una cosa de contener otras dentro de ciertos límites. Aptitud, talento, cualidad que dispone a alguien para el buen ejercicio de algo]. Lo mismo en evaluación, aunque en general referido a la posibilidad que tiene un programa para conseguir alcanzar sus objetivos.
- **COBERTURA:** [DRAE: Cantidad o porcentaje abarcado por una cosa o una actividad]. Consiste en evaluar hasta qué punto un programa llega a la población beneficiaria del mismo, es decir, a qué proporción de la población diana han alcanzado los beneficios del programa y a cuál no, cuáles son las características de ambas y si hay explicaciones a esta discriminación. La “cobertura” ha sido descrita como una “premisa del programa” ya que si no se logra el objetivo de la cobertura para nada vale el programa (Ligero, 2001).
- **DIMENSIÓN:** [DRAE: Aspecto o faceta de algo. Cada una de las magnitudes de un conjunto que sirven para definir un fenómeno], [SALVÁ/1865: solo lo define como “tamaño o volumen”] [MOLINER: En sentido no material, aspecto o cualidad que pueda ser considerada en algo]. En EVALUACIÓN se utiliza solo en plural “las dimensiones” de un sistema de información, de un conjunto de estándares, de variable so de indicadores. Una definición muy personal sería, “las dimensiones son aquellos trozos que contienen información relevante para elaborar el mapa de la totalidad del asunto”.

---

<sup>15</sup> Los empiristas han reaccionado contra esta idealización racionalista, tratando de poner en evidencia las sombras que se mueven tras toda forma de utopía social, desde los “Viajes de Gulliver”, de Johnathan Swift, hasta “Un Mundo Feliz”, de Aldoux Huxley, pasando por “Erewhon o Tras las montañas”, de Samuel Butler. Un ejemplo especialmente recurrente de esta estrategia de ridiculización del idealismo racionalista es el opúsculo, “Una modesta proposición para evitar que los hijos de los pobres de Irlanda sean una carga para sus padres o su país, y para hacerlos útiles al público”, de Johnathan Swift. El cine y en particular la ciencia ficción ha convertido el idealismo de los racionalistas en una fuente de inspiración, que en ocasiones adopta un tono apologético y en otras crítico. Las películas sobre el tema son muchas, pero quizá la que mejor nos ayuda a comprender porque los evaluadores deben rehuir el idealismo racionalista (aunque en otros aspectos de su vida o sus intereses intelectuales o sus creencias nadie le dice que dejen de ser racionalistas), con una especial referencia a “los jóvenes”, es “La fuga de Logan” (1976), de Michael Anderson, que por cierto no tiene nada que ver con una horrorosa serie de TV que también se titulaba así.

- **EFFECTIVIDAD:** [DRAE: Capacidad de lograr el efecto que se desea o se espera], [SALVÁ/1865: Que se hace realmente y llega a tener efecto]<sup>16</sup>, [MOLINER: Cualidad de efectivo, (efectivo: se aplica a la cosa que produce efecto), (efectivo en MOLINER ORIGINAL: Real y verdadero en oposición a lo quimérico, dudoso o nominal)], [Del latín (VALBUENA REFORMADO): *efficio, efficus, effeci, efecctum, efficiere*: ejecutar, llevar a cabo, obtener, producir y efectuar], [En inglés (OXFORD CONCISE): *effectiveness*]<sup>17</sup>. EN EVALUACIÓN: Se refiere a la capacidad para conseguir, alcanzar o satisfacer nuestros deseos o los objetivos marcados. Mide por tanto la posibilidad de hacer o alcanzar algo. Podemos imaginar que tales deseos u objetivos puedan ser las propuestas **políticas** (y las retóricas y los enunciados de las políticas y planes de juventud), la medida estricta de la efectividad sería sólo, y de acuerdo con la definición original de María Moliner, plantearnos de forma previa las posibilidades de alcanzar estas metas, es decir, la noción de efectividad supone plantearse ¿es realista conseguirlo? Porque si nos planteamos medir la efectividad una vez aplicado el programa y obtenidos los resultados se confunde de forma inevitable en nuestro idioma con la medida de la eficacia.
- **EFICACIA:** [DRAE: Capacidad de lograr el efecto que se desea o se espera], [SALVÁ/1865: Activo, poderoso y fervoroso para obrar], [MOLINER: Cualidad de eficaz (eficaz: se aplica a las cosas que producen el efecto o prestan el servicio al que están destinadas), (Eficacia en MOLINER ORIGINAL: fuerza y poder para obrar / validez)], [Del Latín (VALBUENA REFORMADO): *eficaz, eficacis*: que puede producir el efecto deseado]<sup>18</sup>, [En inglés (OXFORD CONCISE): *efficacy*]. EN EVALUACIÓN: es la medida de la capacidad para alcanzar las metas y los resultados propuestos en un periodo determinado y sin considerar los costes en los que se incurre para obtenerlos. Mide por tanto la relación entre objetivos y resultados en una situación ideal y sin considerar, por tanto, los recursos adecuados y necesarios (o aplicados) para conseguirlo. Algunos la interpretan como una medida absurda porque no nos explica nada, pero otros en cambio, entienden que es la “verdadera evaluación de las políticas”, ya que sólo tiene en cuenta las propuestas (las reivindicaciones y las necesidades) y los resultados (las respuestas).
- **EFICIENCIA:** [DRAE: Capacidad de disponer de alguien o de algo para conseguir un efecto determinado], [SALVÁ/1865: Virtud y facultad para hacer alguna cosa y la misma acción con la que se ejecuta], [MOLINER: cualidad de eficiente (eficiente: se aplica a lo que realiza cumplidamente la función a que esta destinado), (eficiencia en MOLINER ORIGINAL: Poder y facultad o aptitud especial para lograr un efecto determinado / acción con que se logra este efecto)], [Del Latín (VALBUENA REFORMADO): *efficientia*: virtud del producto], [En inglés (OXFORD CONCISE): *efficiency*]. EN EVALUACIÓN se refiere al criterio “económico” que relaciona los costes y los beneficios obtenidos, esto es, se trata de determinar cuál es la combinación óptima de recursos financieros, materiales, técnicos, naturales y humanos para maximizar los resultados a obtener. Será mayor cuanto más cercana al óptimo sea la aportación de insumos para maximizar el producto. Se define como impacto global a los efectos, tanto buscados como no previstos. En tal sentido, será necesario discernir entre aquellos efectos provocados por el proyecto y aquellos que

16 El Nuevo Diccionario de Don Vicente Salvá se editó en París en 1865 por los hermanos Garnier y es sin duda el mejor diccionario del castellano cotidiano hablado en el siglo XIX y el que mayor influencia ha tenido en la conformación del idioma real que hablamos hoy en día.

17 Algunos autores comienzan a utilizar el término *responsiveness*, que siempre se ha traducido por conformidad o simpatía, pero que para estos autores parece incluir la noción de que no se trata de “efectos ideales”, sino más bien de los “derechos sociales y humanos” reales. Creo que en los próximos años se va a hablar mucho de *responsiveness*, que algunos traducen ahora por “responsabilidad social”.

18 En la Edad Media los Tomistas debatieron sobre la posibilidad de preservar este término como uno de los atributos de Dios. Porque sólo la divinidad podía ser plenamente “eficaz”.

son producto de acontecimientos externos. Mide por tanto la relación entre recursos y resultados en la realidad.

- **EMICIDAD:** [DRAE: no figura] procede de “emic”, término que escenifica en antropología la perspectiva del “criterio cultural del nativo”. No quiere decir ni subjetivo, ni personal, sino que se refiere a lo que resulta razonable y lógico para la cultura de estos “nativos”. En el caso de la evaluación, la emicidad reflejará la perspectiva cultural de los usuarios y del equipo que realiza la intervención.
- **EMPATÍA:** [DRAE: Identificación mental y afectiva de un sujeto con el estado de ánimo de otro]. La capacidad del evaluador ilustrado por colocarse en el sitio, en el lugar social y cultural del otro. Se supone que en evaluación, la empatía implica manejar una buena base de eticidad, pero a la vez poder captar, asumir, aceptar y entender los componentes emic de la situación.
- **ESTÁNDAR:** [DRAE: Que sirve como tipo, modelo, norma, patrón o referencia], [MOLINER: Se aplica a lo que es conforme a un determinado modelo]. Se trata de un anglicismo que lleva pocos años incorporado al DRAE, tienen el mismo origen que el galicismo “estandarte” que se incorporó al español en el siglo XVI, esto explica que en inglés estandarte y estándar sean lo mismo. En el manual del EFQM se define como: Resultado deseado de un proceso que se puede medir mediante el uso de indicadores. Otras definiciones adecuadas son: Normas de desempeño definidas para una actividad, un proceso, un producto o un servicio. Propiedad que garantiza la uniformidad en los métodos de capturar, representar, almacenar y documentar la información. Es un nivel o referencia de calidad establecida por alguna agencia, organismo de acreditación o institución. Los estándares de calidad o de excelencia sobre instituciones o programas se establecen previamente y, de forma general, por una agencia de acreditación. En resumen es un término muy polisémico que se utiliza para acreditar algo. Lo cual exige que tenga apellidos, es decir, que sean “estándares de algo”. En nuestro caso “estándares de evaluación”. En general se requiere además una institución que garantice que los estándares hayan sido construidos con precisión y conocimiento. En nuestro caso vamos a elegir a Naciones Unidas.
- **ETICIDAD:** [DRAE: no figura acepción]. Procede de “etic”, término que en antropología trata de definir la perspectiva “cultural, técnica y teórica” del colectivo de los antropólogos. Incluye tanto sus conocimientos académicos como su acervo cultural como miembro de una determinada cultura, que actualmente se supone que es la “cultura democrática y científico técnica” del occidente post-ilustrado. En evaluación, la perspectiva etic se corresponde con la de la “evaluación fuerte” que desarrollamos en estas páginas, aunque en la misma también incluimos la perspectiva “emic”.
- **EVIDENCIA.** [DRAE: Certeza clara y manifiesta de la que no se puede dudar]. Se trata de un resultado empírico relevante en términos de conocimiento científico y que además resulta especialmente pertinente en el contexto del conocimiento. La evidencia supone, por tanto, una constatación empírica con pretensiones de valor universal, es decir, que trasciende el contexto particular del hallazgo. La evidencia es por tanto una certeza que además nos aporta conocimiento más allá del contexto de la evaluación. Pero a la vez se ha convertido en una palabra fuerte que se utiliza como una definición operativa en casi todas las disciplinas (desde salud hasta justicia), e incluso en áreas tan dispares como la teología o la ufología. El evaluador no busca evidencias, porque esta es la tarea del investigador, pero de la misma manera que al jugar a la lotería cabe una cierta posibilidad de que te toque el premio gordo, en la evaluación pueden aparecer evidencias. En todo caso, como es bien sabido, si abandonamos todos nuestros proyectos de vida esperando que nos toque el gordo, por puro cálculo de probabilidades, acabaremos mal.

- **FUNDAMENTACIÓN:** [DRAE: Establecer, asegurar y hacer firme algo] Se refiere a los componentes teóricos y conceptuales que justifican los criterios adoptados para definir y diseñar tanto el programa como la evaluación.
- **HALLAZGO:** [DRAE: Dar con alguien o algo que se busca. Dar con alguien o algo sin buscarlo. Descubrir con ingenio algo hasta entonces desconocido. Ver, observar, notar. Descubrir la verdad de algo]. En evaluación es lo mismo.
- **INDICADOR:** [DRAE: INDICAR, Mostrar o significar algo con indicios y señales] [SALVÁ/1865: Dar a entender o señalar alguna cosa con indicios o señales] [MOLINER: INDICAR, Servir para hacer saber o conocer una cosa]. En el capítulo 16 se describe una parte del debate que existe sobre este concepto, pero aquí podemos adelantar que la definición más tradicional de indicador corresponde a la identificación de una magnitud numérica referida a un hecho o acontecimiento. Sin embargo hoy en día se utilizan también indicadores cualitativos. El tratamiento y análisis de la información aportada por el indicador permite establecer la explicación, evolución y predicción de un fenómeno. Es por tanto el instrumento que pueden permitir evaluar de manera objetiva aspectos particulares del proceso de ejecución o de los resultados de un programa ó proyecto a través de mediciones de carácter cualitativo o cuantitativo.
- **INFRAESTRUCTURA:** [DRAE: Conjunto de elementos o servicios que se consideran necesarios para la creación y funcionamiento de una organización cualquiera]. En evaluación es lo mismo.
- **OPORTUNIDAD:** [DRAE: Sazón, coyuntura, conveniencia de tiempo y de lugar]. En evaluación es más concreto y se refiere a la conveniencia óptima para aplicar el programa.
- **PERTINENCIA:** [DRAE: Que viene a propósito]. Es la adecuación, en todas sus dimensiones, de los objetivos de la acción (o política o programa) al contexto del territorio y la población diana sobre en los que se realiza la intervención. La pertinencia también implica la optimización de las posibilidades técnicas por parte de la institución y su relevancia con respecto a los supuestos en los que se basa la definición de la política de juventud.
- **PROCEDIMIENTO:** [DRAE: Método de ejecutar algunas cosas]. En evaluación es lo mismo.
- **RELEVANCIA:** [DRAE: Sobresaliente, destacado, importante, significativo]. Se trata de un criterio cultural que otorga valor a las cosas en términos de importancia. Obviamente esto significa que algo es relevante desde una determinada perspectiva (y no tanto desde otra). Por ejemplo, un objetivo puede ser muy relevante desde la perspectiva de los técnicos, pero no serlo para la población diana y a la inversa. Asimismo, un fuerte grado de conocimiento por parte del público de la existencia de un dispositivo puede ser muy relevante para los planificadores pero nada relevante para este mismo público que no lo va a utilizar jamás.
- **SATISFACCIÓN:** [DRAE: Cumplimiento del deseo o del gusto]. Indica el grado de satisfacción de los usuarios como un valor de juicio sobre el programa.
- **VIABILIDAD:** [DRAE: Dicho de un asunto: que, por sus circunstancias, tiene probabilidades de poderse llevar a cabo]. Se asocia a las perspectivas de duración y sostenibilidad de los efectos, es decir, el grado en que los efectos positivos derivados de la intervención continúan una vez finalizado el flujo de la ayuda externa.
- **VISIBILIDAD:** [DRAE: Que se puede ver. Tan cierto y evidente que no admite duda]. Analiza que el proyecto es conocido por los diferentes agentes implicados y por la población beneficiaria y se identifica a sus financiadores.

### 2.3.- Las relaciones de la evaluación con la planificación y la gestión, la calidad y la satisfacción.

Vamos finalmente a tratar de explicar las relaciones entre la evaluación, la planificación y la gestión de programas, la calidad y la satisfacción. Aunque, salvo a lo referido a calidad las otras cuestiones van a ser tratadas en diversos momentos del texto, nos ha parecido conveniente citarlas en un apartado específico, para afrontar de una forma explícita la cuestión de las relaciones entre todos estos niveles.

Así en el primer tema, **planificación y gestión**, los apartados precedentes nos muestran cómo la evaluación forma parte del proceso de planificación de una política, un plan o un programa. Es decir, no es algo que viene “después” o se desarrolla de una forma paralela y sin contacto con los programas, todo lo contrario. Cuando se planifica y diseña un programa tienen que considerarse de entrada las necesidades y las características de la evaluación, a la vez que la evaluación de diagnóstico, de realidad y de diseño son muy útiles para la planificación. Como afirma Olga Nierenberg en una afortunada frase *“la evaluación es un espejo de la planificación y la planificación un espejo de la evaluación”* (Nierenberg, 2000). Pero además esta relación debe mantenerse a lo largo de toda la gestión del programa y considerar que la evaluación no es algo ajeno, sino una parte integral del mismo.

Algunos autores no piensan así, aunque son pocos y la mayoría situados en el ámbito politécnico de las ingenierías que sigue manejando el criterio clásico de primero planificación y después evaluación, los cuales les conduce a diferenciar la evaluación de la planificación y gestión. Pero ésta no es la opción de la evaluación fuerte y, por tanto, en los próximos capítulos desarrollaremos de forma detallada esta cuestión.

En cambio, las cosas están más abiertas en torno a **la relación entre evaluación y calidad**, sobre la que hay muy poca literatura a pesar de que ambos han emergido de forma similar y continua en todos los ámbitos sociales y en particular en juventud. En el fondo porque entre ambos conceptos se producen muchas sinergias y continuidades e incluso en algunos casos ambos territorios se confunden, pero a la vez cada uno trata de conquistar un espacio propio. Para aclarar la cuestión vamos a utilizar un modelo que distingue cuatro narraciones en torno la noción de calidad (Fantova, 2001):

1. La narración que procede de la gestión de calidad desarrollada en el mundo industrial y, más en general, en el sector empresarial convencional, y que se ha convertido en la más frecuente en los discursos de presentación y legitimación de las propuestas de mejora de la gestión de las intervenciones sociales. Se trata de las corrientes de pensamiento y acción que giran en torno a las normas ISO 9000 o al modelo EFQM<sup>19</sup>. Conviene retener, en todo caso, que ambas opciones exigen evaluación y participación del conjunto de miembros de la organización.
2. La narración que surge desde los discursos y las prácticas de la evaluación de programas sociales y, en general, de las herramientas relacionadas con estándares de referencia para dichos programas sociales. Así, en ocasiones, el enmarque es más bien el de la evaluación de programas (Nierenberg, 2000), así como en los procesos de acreditación u homologación de servicios (Comas, 1992).
3. La narración que está vinculada a las propuestas de modernización de las administraciones públicas (Osborne, 2003). En general, se trata de propuestas que pretenden introducir la filosofía de la denominada *orientación al cliente* y las técnicas de gestión empresarial convencional en las administraciones y servicios públicos.

<sup>19</sup> Por su parte numerosos autores consideran que mientras las normas de la ISO-9000 se sitúan en el territorio conceptual de la mera calidad, el modelo EFQM trata de compaginar calidad y evaluación, o al menos facilita la evaluación en el proceso de calidad para alcanzar la puntuación que otorga el certificado EFQM (Ligero, 2004), algunos incluso consideran que el modelo EFQM aplicado a las ONGs (y en particular a las Organizaciones Juveniles) representa una forma de evaluación implícita que resuelve muchas de las dificultades de emprender evaluaciones reales de programas y acciones.

4. Por último, aparece la relación ligada con el concepto de *calidad de vida* (que desarrollaremos en el capítulo 16), que suele ser presentado como guía para los servicios con un énfasis en la vivencia subjetiva de las usuarias y usuarios. No se trata de atender sólo a las condiciones objetivas en las que vive la persona, sino también a sus percepciones, valores, preferencias y satisfacción. Las propuestas que giran en torno al concepto de calidad de vida suelen aparecer vinculadas a modelos de planificación y evaluación basados en las preferencias y deseos de las personas destinatarias de los servicios y, en general, modelos de gestión de calidad (Schalock y Verdugo, 2003).

Estas narraciones, que Fernando Fantova describió en su citada tesis doctoral, se ha completado recientemente con la idea de que las narraciones y las tradiciones se mezclan, influyen las unas en las otras y cristalizan en las diversas concepciones de la gestión de calidad que cabe encontrar en la vida cotidiana de las instituciones y en los textos que se escriben más o menos a partir de dicha vida cotidiana (Fantova, 2004). Se trata de una realidad compleja y casi inabarcable con cientos de modelos (específicos y generales), normas y códigos éticos cuya mera relación ya produce publicaciones completas (Vidal, 2007). Listar una clasificación de las mismas y además tratando de establecer una relación entre la calidad y la evaluación, produce una cierta sensación de mareo, aunque si no conseguimos sistematizar un poco el tema corremos el riesgo de “elegir” y perdernos en el dogma. Las principales narraciones serían:

1. Se puede entender la gestión de calidad, en primer lugar, como gestión basada en la evidencia o gestión con datos, en cuyo caso es necesaria una estrategia de evaluación formulada en términos de evaluación fuerte.
2. En otros casos, por calidad se entenderán los compromisos que la organización o sistema adopta y cumple en relación con sus clientes, en nuestro caso los jóvenes. En este caso la evaluación no tiene ninguna relación con la calidad, ya que ésta se construye con herramientas como códigos éticos, carteras de servicios y con el esfuerzo por el establecimiento de derechos y obligaciones de los diversos agentes o instancias involucradas en el gobierno, la gestión y la provisión de los servicios.
3. Por gestión de calidad se entiende, también, en términos de gestión en red, cuando se alude a la necesidad de afrontar las nuevas situaciones de complejidad, utilizando las tecnologías de la información y la comunicación y también de nuevas realidades sociales ante las que la gestión se propone como necesariamente participativa. En este caso no hay una relación directa con la evaluación, pero resulta evidente que sin los resultados y los instrumentos de la evaluación poco se puede avanzar en esta línea.
4. Como consecuencia de lo anterior se habla de una gestión de calidad en la que el acceso a y por esas tecnologías y la participación de todas las personas se potencian mutuamente para construir organizaciones cada vez más capaces de aprender. En este caso, calidad y evaluación son casi lo mismo.
5. Se entiende también la gestión de calidad como gestión por procesos, en la medida en que se insiste en la importancia de identificar, analizar y mejorar los procesos que realmente tienen lugar en las organizaciones y sistemas. Esta es exactamente también la tarea de la “evaluación de diseño”.
6. La propuesta más común es la de dimensión estratégica, uno de los lugares comunes de la actual teoría de la gestión y de muchas de las propuestas de gestión de calidad, como el modelo EFQM. Hablar de dimensión o perspectiva estratégica supone intentar prever los ciclos largos e intentar anticipar y gobernar los cambios que afectan de forma relevante al posicionamiento sostenible de la organización o sistema que estamos gestionando. En este caso la evaluación puede ser un instrumento de la gestión de calidad, pero un instrumento que no acaba de dar cuenta de las necesidades a largo plazo, aunque puede servir para medir el grado de coherencia de las organización en relación a tales objetivos.

7. Finalmente aparece una gestión de calidad relacionada con la participación. Se trata de la mirada clásica de la teoría organizacional, muy vinculada al mundo industrial, aunque atribuyendo una gran importancia a la comunicación interpersonal y a fenómenos como el liderazgo o el empoderamiento. Ésta es, posiblemente, una de las intuiciones principales del movimiento de la calidad total (Municio, 2002), a la que ya hemos atribuido (Comas, 2003) la esperanza de una futura imbricación con la evaluación. En todo caso, no es más que una esperanza.

Esta claro, con todas estas narraciones y cristalizaciones, que las relaciones entre evaluación y calidad son muy complejas y cambiantes, porque aunque la evaluación parece que se va condensando en un campo de conocimiento (el de la evaluación fuerte), la calidad muestra un rostro muy poliédrico. En cualquier caso, estamos ante dos prácticas en construcción que no sólo caminan en paralelo, sino que se interfieren y se apoyan.

Aunque en esta relación el futuro dirá, parece lógico que los ámbitos de juventud apuesten por una cierta sinergia entre ambas (sin confundirlas), en parte porque la posibilidad de asumir ambas tareas no es fácil en un ámbito caracterizado (o identificado) por el “agobio cotidiano” (Comas, 2007). También es cierto que la “calidad” permite superar estos “inadecuados estilos de trabajo”. Conviene también tener en cuenta que muchas cosas de las que aprende un profesional en un Plan de Calidad (especialmente en modelos participativos), son luego de aplicación en las evaluaciones. Y a la inversa.

En cualquier caso es muy importante, especialmente en los ámbitos de juventud en los que la noción de efectividad es especialmente relevante (aunque aparezca muy ligada a las retóricas políticas), no confundir la calidad de los organismos de juventud, con criterios de eficacia, de eficiencia y de mercado), con la calidad de las políticas de juventud, con criterios sociales de efectividad (Zufiaurre y Albertín, 2006).

Una vez presentada esta compleja relación entre la evaluación y la calidad podemos referirnos a **las relaciones entre evaluación y satisfacción**. Porque la cuestión de la satisfacción en particular, la “satisfacción del cliente”, se ha convertido en uno de los temas centrales de la gestión de calidad en casi todas las narraciones aludidas (Evans y Lindsay, 2005).

Pero a la vez, como hemos visto en el apartado precedente, la medida de la satisfacción puede ser un componente específico de la evaluación, porque se supone que la satisfacción se limita a medir, por ejemplo, si los jóvenes están satisfechos con un determinado servicio. Se supone que si lo están la evaluación es “positiva” y si no lo están es “negativa” y algo hay que cambiar.

Medir la satisfacción es además algo relativamente fácil y hay excelentes manuales para hacerlo (Sanz 1994). Pero en realidad las relaciones entre ambos conceptos son algo más complejas. En una gran medida porque entre los motivos de satisfacción de los usuarios y los objetivos del programa se pueden producir diversas contradicciones. Además, los diversos actores sociales pudieran tener motivos de satisfacción muy diferentes.

Expresado en términos sencillos se puede decir, que la satisfacción de los usuarios podría no tener que ver con los objetivos del programa. Un ejemplo clásico de esta situación se refiere a las relaciones entre profesores y alumnos, el objetivo de los primeros es que los segundos aprendan, el objetivo de los segundos puede ser el de aprender y aprobar, pero también el de aprobar sin esfuerzo y sin necesidad de aprender. En este segundo caso la satisfacción de los estudiantes no se corresponde con los objetivos del programa, así puede que hayan aprobado todos (y por ello se sienten muy satisfechos), pero los objetivos de aprendizaje no se han cubierto. Volveremos a este tema en el capítulo 12.

## 2.4.- Preparados para comenzar a despejar algunos equívocos.

Antes de seguir, y con los elementos que hemos aportado en los párrafos precedentes sobre la evaluación, podemos ya entender cómo algunos tópicos producen una imagen equívoca de la

función de la evaluación. En este sentido resulta muy frecuente encontrar afirmaciones como la siguiente: ES NECESARIO EVALUAR LOS PROGRAMAS DE JUVENTUD PARA **APLICAR SÓLO AQUELLOS QUE HAN DEMOSTRADO SU EFICACIA.**

Este eslogan, que se ha convertido en un lugar común a partir de su continua reiteración, refleja lo poco que se sabe de evaluación y el habitual uso político y corporativo que se hace de la misma. Pero a la vez nos resulta una frase útil porque condensa, como pocas, todos los errores que envuelven la noción de evaluación.

PRIMER ERROR:

**Sostener la fantasía omnipotente de que el mundo social las evidencias son universales.**

En el mundo físico, por ejemplo en el funcionamiento del cuerpo humano o en funcionamiento del cosmos, cuando se establece una evidencia se supone que es universal (aunque los físicos teóricos, los que construyeron la noción de evidencia universal ya no están de acuerdo con esta afirmación). Pero el mundo social real no obedece a leyes universales y aunque fuera así estamos muy lejos de descubrirlas. En el ámbito social los intentos de producir evidencias empíricas universales son, como hemos venido explicando, muy escasos y hasta ahora se han limitado a dejar constancia de que en determinadas condiciones, que son sólo las de un presente y unos espacios sociales concretos, ocurren algunos fenómenos. Lo demás son proyecciones ideológicas y creencias, que se toleran por razones prácticas, que representan espacios de poder (en ocasiones de campos de conocimiento que prestan sus presupuestos), que no son controlables y que no cambian con el tiempo. Es decir, si observamos la tabla 1.3, pensar en evidencias universales a partir de la evaluación de programas reúne todas las condiciones de la pseudo-ciencia.

SEGUNDO ERROR:

**Olvidarse de la variable tiempo histórico.** Mientras el universo físico se mantiene relativamente estable, el mundo social se halla sometido a continuos procesos de cambio. El cambio social no sólo es imprevisible, sino que además establece una cierta dialéctica con las propias previsiones y las transforma. El ideal utópico de una explicación biológica del comportamiento humano y social, que en la actualidad representan muy bien tanto el racionalismo de Chomsky como el empirismo de la socio-biología, no es más que una propuesta hipotética cuyos avances son más que discretos. Mientras tanto, todas nuestras explicaciones sólo pueden ser post-facto.

Esto significa que las evidencias obtenidas en la evaluación de un programa o actividad tienen una proyección limitada. Su validez, en términos de proyección, depende de la capacidad para integrar los nuevos (e inéditos) componentes sociales y culturales, que van conformando los hechos sociales. La validez de un programa social se limita a un contexto histórico, más allá del cual es sólo una “experiencia” que nos puede aportar un simple “mejor entendimiento” de posteriores experiencias y evaluaciones.

Todo programa social tiene un tiempo limitado, **una fecha de caducidad, transcurrida la cual no conviene utilizarlo si no queremos intoxicarnos.** Es decir, si consumimos sin precauciones un programa social considerado muy eficaz hace apenas un par de décadas, se pueden garantizar graves trastornos gastro-intestinales.

TERCER ERROR:

**Olvidarse de la variable condiciones de aplicación.** Cada equipo profesional es diferente cuando interacciona con personas. No sólo existen titulaciones y experiencias diferentes, sino una relación de poder entre ellas particular en cada equipo, pero además las biografías no son las mismas, y factores como la edad determinan tanto el contenido de la comunicación como el de la comprensión. Los equipos que trabajan en lo social proceden de campos de conocimiento con orientaciones diferentes y en ocasiones antagónicas. Los valores manifiestos no coinciden con los latentes, que además pueden modificarse en un corto periodo de tiempo y por razones que en ocasiones son muy personales (un

malestar, una separación, un luto). El establecimiento de protocolos cerrados no evita todos estos problemas, porque no puede contemplarlos y además provoca reacciones de rechazo propias de cada contexto y estructura profesional.

Asimismo, aunque esto tiene que ver mucho con el siguiente error, las poblaciones diana no son nunca las mismas aunque parezcan idénticas. La lista de variables que definen una población diana, como la edad, el género y la identidad sexual, los vínculos familiares y la estructura de la familia, las experiencias escolares, el estilo de vida de la familia de procedencia, la clase social, el hábitat residencial, la ideología y los valores religiosos y cívicos, la identidad étnica o el universo simbólico más o menos subconsciente, resulta infinita y con el tiempo y el avance de la investigación los ítems de cada variable son cada vez más. El mundo social es, en la percepción de los investigadores, cada vez más complejo y plural, y a la vez los contenidos de la cohesión social van variando.

#### CUARTO ERROR:

**No reconocer la especificidad de los territorios sociales reales.** Como consecuencia, cada territorio social es distinto y no resulta fácil determinar las características que lo distinguen de los demás. Un programa social que ha funcionado en un sitio no va a funcionar nunca de la misma manera en otro, hay que adaptarlo y para ello hay que hacer un diagnóstico o un análisis previo de la realidad. En este tipo de adaptaciones los “programas de éxito” son sólo ejemplos de “buenas prácticas”, pero nunca guías a las que debemos ajustarnos de una manera automática.

#### QUINTO ERROR:

**Imaginar un mundo sin intereses culturales, políticos, corporativos y económicos.** El cambio histórico y social, la variabilidad de los equipos y los contextos, así como la transformación cultural, son el resultado de procesos complejos en los que operan todas las variables citadas (y algunas más). Se pueden analizar los procesos de cambio e incluso prever algunas transformaciones y en este sentido proponer incluso adaptaciones de “programas de éxito”.

Pero una parte sustancial del cambio social responde a la influencia de los intereses culturales, políticos, corporativos y económicos. Debemos por una parte aprender a entender estos intereses y, por otra, debemos considerar que la evaluación no es inmune a la misma. Una parte sustancial de los “programas que han demostrado su eficacia” responden a alguno de estos intereses y la demostración, aunque sea metodológicamente correcta, puede, por ejemplo, ser resultado de una inversión cuantiosa con capacidad para producir ciertos efectos que sin la misma no se producen (ver más adelante la noción de evaluación de relaciones públicas).

¿Nos impiden todos estos errores trabajar? Más bien no, sino que nos invitan a trabajar de otra manera, nos invitan a practicar un tipo de evaluación realista (es decir, ajustada a la realidad social) que tenga en cuenta los aspectos técnicos e instrumentales de la misma, pero también los componentes políticos, ideológicos, de contexto e historia de los programas. Pero ¡ojo! la evaluación no “denuncia” estos componentes, sino que los integra en el análisis porque son parte de la realidad, de una realidad inevitable y que no podemos ignorar. A partir de esta constatación vamos a tratar de entender cuál es la estrategia de trabajo de la evaluación fuerte.

## BIBLIOGRAFÍA Y DOCUMENTACIÓN COMPLEMENTARIA.

- ECHEVARRÍA, J. (2003), **La revolución tecno-científica**, Madrid, Fondo de Cultura Económica.
- SANZ, E. (1994), **Manual de estudios de usuarios**, Madrid, Fundación Sanchez Ruperez / Pirámide.

### 3.1.- Los tres momentos de la evaluación y los tipos de evaluación.

En el capítulo anterior dejábamos bien establecido que la evaluación fuerte reúne procedimientos operativos que se distinguen según los momentos del desarrollo del programa, de la política o de la acción. En cada uno de estos momentos la evaluación obedece a una lógica distinta y tiene por tanto una finalidad diferente (Comas, 2001). Vamos a retomar desde la perspectiva de los procedimientos de evaluación el contenido de cada uno de estos tres momentos. Comencemos por describirlos.

- El primer momento se refiere al **“antes de”** y se plantea objetivos relacionados con la necesidad de realizar un análisis previo de la realidad sobre la que se va a desarrollar el programa. Su finalidad es la de averiguar cuáles deberían ser los contenidos y los objetivos del programa.
- El segundo momento se produce **“durante”** la realización del programa o la intervención e implica averiguar lo que estamos haciendo para valorarlo y tratar de modificarlo. La finalidad de este tipo de evaluación es revisar los procesos y los procedimientos para comprobar cómo se utilizan los recursos y estimar cómo se van a conformar los resultados.
- El tercer posible momento de la evaluación se refiere al **“después de”**, es decir, cuando el programa o la actividad ha concluido, para medir los resultados y el impacto causado. Su finalidad es la de determinar las transformaciones que ha sufrido la realidad.

Cada uno de estos tres momentos requiere procedimientos evaluativos propios que han sido denominados de forma particular, porque su finalidad es distinta. Cada uno de estos procedimientos se puede realizar de forma diferente, utilizando distintas técnicas de recogida, producción y análisis de datos. Pero, a la vez, no son momentos independientes, ya que sin un buen conocimiento de la realidad no podemos evaluar correctamente los procesos y sin una evaluación de procesos no podemos determinar el impacto. A la vez, la evaluación de impacto es imprescindible para captar la evolución de la realidad tras la aplicación del programa.

¿Significa esto que en una evaluación debemos evaluar cada uno de estos momentos? Pues en buena lógica evaluativa esto resulta imprescindible. Pero hacerlo no es fácil, no tanto por razones técnicas, sino por falta de recursos y porque los tiempos de intervención pueden estar

Cuadro 3.1. **Tipos de evaluaciones según finalidad y momento de su aplicación**

TIPO DE EVALUACIÓN	MOMENTO	FINALIDAD
ANÁLISIS DE LA REALIDAD	Antes	Sirve para conocer el contexto social sobre el que va a ser aplicado el programa.
DIAGNÓSTICO ("necesidades")	Antes	Se utiliza para conocer las necesidades o los problemas de la ciudadanía en relación a un tema concreto.
DISEÑO o CONCEPTUALIZACIÓN	Antes	Nos permite conocer la adecuación del programa de intervención social a las necesidades de la población.
EVALUABILIDAD	Antes	Determina las posibilidades de evaluar el programa que se va a realizar.
COBERTURA	Antes	Permite conocer si el programa puede llegar de verdad a la población objeto del mismo.
IMPLEMENTACIÓN, MONITORIZACIÓN Y SEGUIMIENTO	Durante	Proporciona la información para saber cómo estamos aplicando el programa. No confundir con los Estudios de Seguimiento, ni tampoco con el registro de información para la gestión del programa
PROCESOS	Durante	Describe de qué manera, con qué medios y con qué resultados (relativos a la aplicación) se está desarrollando el programa.
PERSONAL	Durante y después	Permite constar el rendimiento del equipo y sus diferentes profesionales y especialmente establecer si la calificación es adecuada. Se trata de una tipo de evaluación muy delicada que puede interferir con aspectos laborales.
RESULTADOS	Después	Sistematiza los logros obtenidos en el programa en relación a los objetivos del mismo.
IMPACTO	Después	Permite saber qué efectos/consecuencias ha tenido la intervención social (programa) sobre el contexto social intervenido.
ECONÓMICA	Después	Se trata de medir el coste/beneficio (evaluación monetaria de los efectos del programa), el coste/efectividad (casi imposible) y el coste/utilidad (exige categorías preestablecidas de utilidad)

Fuente: elaboración propia.

muy ajustados. Tampoco pasa nada porque, como veremos, existen estrategias de compensación y atajos.

En las políticas sociales en general, y en las particulares de juventud, la duración de una intervención es limitada y desde la toma de la decisión política hasta su ejecución no aparecen demasiadas posibilidades para evaluar. Asimismo, una vez concluida la intervención se demandan con urgencia "los resultados". En general, todas las intervenciones se ajustan a los mecanismos presupuestarios anuales y cuando se trata de programas consolidados a lo largo de años, su propia consolidación administrativa impide, en muchas ocasiones, la revisión y adaptación de los mismos. Además, son muchos los técnicos que sostienen (aunque sea de forma implícita) que aquellos programas que están consolidados "y funcionan", es mejor no tocarlos ni evaluarlos. En realidad tienen razón, porque, como hemos explicado, la "mala prensa" de la evaluación tiene que ver con su utilización como "recurso de control", pero también de "justificación". Como consecuencia, muchos técnicos poseen la experiencia de que cuando el programa consolidado, estable y que cumplía con sus objetivos ha sido eliminado (por razones presupuestarias, personales o ideológicas), el proceso de eliminación comenzó en realidad "encargando una evaluación".

Por este motivo, dos de cada tres evaluaciones (o más) responden a un mismo esquema de momentos de evaluación: no hay evaluación “antes de”, pero se produce una evaluación “durante” (en general una evaluación de procesos), aunque los datos de la evaluación de procesos son facilitados “después de” y titulados “resultados”. La buena lógica de la evaluación es sin embargo otra, ya que requiere que se realice evaluación en cada uno de los momentos indicados. El cuadro 3.1 muestra los tipos de evaluación que se pueden realizar en cada uno de estos momentos.

Solo una aclaración para los más expertos en la materia: se puede jugar con los tiempos, es decir, el esquema de los tres momentos no es rígido, pero esto es algo para manejar con mucha precaución y sólo cuando se tiene mucha experiencia en ajustar tiempos y tipos de evaluación según este esquema. También hay que considerar el problema de los tiempos (en particular el “después”) cuando los programas son estables. En este caso hay que desagregar las intervenciones en “unidades temporales” cuyo límite superior estable el inicio del después para un grupo de casos o acontecimientos ocurridos “durante” la unidad temporal.

### 3.2.- Las fases de la evaluación.

Con independencia del tipo de evaluación que estemos realizando, toda evaluación requiere una planificación que se establece a través de una serie de fases (concordes con lo que hemos llamado “estrategia kantiana), cada una de las cuales supone una parte del procedimiento a desarrollar. En el cuadro 3.2 se desarrolla de forma muy sintética la estructura de estas fases y los aspectos que incluye cada una de ellas. No vamos a incluir más explicaciones en este apartado

Cuadro 3.2. Estructura y fases de la evaluación

FASES	ASPECTOS A INCLUIR
PLANIFICAR LA EVALUACIÓN	<ul style="list-style-type: none"> <li>• Conceptualizar la naturaleza del programa.</li> <li>• Conocer la literatura previa afín o análoga.</li> <li>• Clarificar y jerarquizar los contenidos de la evaluación (¿Qué se va a evaluar y por qué?).</li> <li>• Seleccionar los instrumentos de evaluación.</li> <li>• Definir objetivos, criterios, requisitos y estándares.</li> <li>• Seleccionar a los evaluadores.</li> <li>• Delimitar el campo de la observación.</li> <li>• Establecer las matrices de evaluación.</li> <li>• Diseño del plan de análisis.</li> </ul> <p>EL ORDEN PUEDE VARIAR DE UN PROYECTO A OTRO</p>
APLICAR EL PLAN Y RECOGER LOS DATOS	<p>(VER PARTE II)</p> <p>Definir el proceso a seguir en la implementación del tipo de evaluación diseñada. Incluye recogida de datos, control y supervisión de los mismos, así como verificación, grabación y depuración de los mismos.</p>
ANALIZAR LOS DATOS OBTENIDOS	<p>(VER PARTE III)</p> <ul style="list-style-type: none"> <li>• Aplicación del plan de análisis y juzgar los resultados y efectos.</li> <li>• Formular juicios de valor con los resultados.</li> <li>• Estimar resultados, efectos e impactos</li> </ul>
DAR A CONOCER LOS RESULTADOS	<p>(VER PARTE III Y ANEXO II)</p> <ul style="list-style-type: none"> <li>• Elaborar un Informe.</li> <li>• Preparar extractos institucionales.</li> <li>• Difundir los resultados a los afectados.</li> <li>• Difundir los resultados a la Comunidad Científica.</li> <li>• Remitirlo a registros y a bases de datos y documentales.</li> <li>• Preparar artículos científicos.</li> </ul>

Fuente: elaboración propia.

ya que el contenido de los diferentes aspectos señalados en el cuadro se explica en diferentes lugares del Manual. Así los aspectos relacionados con la Planificación aparecen en esta primera parte y el resto de aspectos según se indica en el propio cuadro.

### **3.3.- Determinar los objetivos (¿Qué queremos saber?).**

Una parte fundamental del procedimiento de la evaluación se refiere a proporcionar una respuesta a la pregunta de Kant sobre qué queremos saber. De hecho, en mi experiencia como evaluador es, en ocasiones, la tarea más difícil y que además retiene y retrasa las tareas siguientes hasta conseguir una respuesta consistente<sup>20</sup>, porque es una respuesta que no podemos obviar y seguir adelante como si nada.

Aún en los casos en los que la respuesta nos viene impuesta, por ejemplo, en un Pliego de Condiciones Técnicas, donde la propuesta puede ser muy abstracta (al estilo de “evaluar la eficacia del Plan de Juventud”), es necesario fijar objetivos más concretos que nos proporcionen una respuesta más precisa sobre lo que queremos saber. Con alguna frecuencia, los Pliegos de Condiciones (y otras demandas) no incluyen objetivos precisos, pero sí citan en cambio indicadores que se supone debemos utilizar, lo cual nos obliga a concebir los objetivos en un momento procesal inadecuado, es decir, después de que componentes esenciales de la evaluación ya hayan sido definidos.

Cuando resulta complicado proporcionar la respuesta a la pregunta esencial de Kant, la única salida es darle muchas vueltas, leer el material, entrevistar a los diversos agentes y actores, reservando toda esta información al modo de Malinowski: recoger todo hasta que el grado de acumulación mental de toda esta información nos ofrezca la posibilidad de responder a la pregunta de Kant. También es cierto que esta manera de trabajar no gusta demasiado a las instituciones públicas y privadas que están sometidas al problema de los plazos y que exigen un cierto grado de claridad antes de comenzar la evaluación.

#### **3.3.1.- El objetivo general.**

El objetivo general de una evaluación se deriva del objetivo general de la intervención, que a su vez es una consecuencia de la misión de la entidad o el dispositivo.

Si no existe una definición precisa de la misión y el objetivo es muy abstracto, no resulta fácil determinar el objetivo de la evaluación, aunque esta carencia ya es un resultado para la evaluación. Para determinar un objetivo general se requiere entonces un trabajo colectivo que incluya una especificación del objetivo general real del programa e incluso una propuesta para la misión de la entidad o el dispositivo.

Esta carencia suele ser corriente en los ámbitos de juventud, ya que la propia misión de los organismos de juventud contiene muchas ambigüedades (y las políticas muchas complejidades y desacuerdos), puesto que en ocasiones no está ni tan siquiera claro cuáles son las competencias de ejecución de las mismas. Además juventud es, en ocasiones, un saco en el que se acumulan necesidades diversas y coyunturales. Trataremos en el capítulo 16 de clarificar algunas cuestiones relacionadas con esta complejidad.

En todo caso, el objetivo general debe ser realista y preciso. Realista en el sentido de alcanzable con los recursos disponibles y las metodologías propuestas, y preciso en el sentido de que enuncia de una forma concreta lo que se quiere hacer.

#### **3.3.2.- Los objetivos específicos.**

Aunque el objetivo general sea muy preciso, también se requiere la formulación de objetivos específicos o particulares relativos a aspectos más concretos o parciales de la evaluación.

---

<sup>20</sup> Lo cual, en más de una ocasión, produce un problema grave porque impide cumplir los plazos, pero es mejor hacerlo así (o no hacerlo de ninguna manera) que presentar una supuesta evaluación que no sabemos porque la hemos hecho.

Por ejemplo, en la evaluación (sea de diagnóstico, de proceso o de impacto) particular de un Centro de Información Juvenil el objetivo general podría ser: comprobar el grado de cobertura, accesibilidad y fiabilidad que posee el servicio para los jóvenes de la localidad. Pero, además, se podrían añadir otros objetivos más específicos del tipo:

- Determinar cuáles son los contenidos que gozan de más credibilidad y porqué.
- Valorar las dificultades de acceso de algunos colectivos.
- Sugerir nuevos campos de información.
- Medir el grado de aceptación de los distintos materiales.
- Determinar si las instalaciones y su ubicación son las adecuadas.
- Valorar la adecuación de los horarios,...

El establecimiento de todos estos objetivos específicos resulta imprescindible para realizar la siguiente tarea: establecer las matrices de evaluación.

### **3.4.- Las matrices de evaluación (¿Cómo vamos a hacerlo?).**

¿Qué es una matriz? Según el DRAE “matriz” es muchas cosas, pero entre ellas es *“un molde en el que se funden objetos de metal que han de ser idénticos”* y también *“un molde de cualquier clase con el que se da forma a algo”*. De manera particular, en matemáticas es *“un conjunto de números o símbolos algebraicos colocados en líneas horizontales y verticales y dispuestos en forma de rectángulo”*. Pues en evaluación una matriz es ambas cosas: el molde que define los procedimientos de la evaluación y que además adopta la forma de un rectángulo de dos ejes en el que se describe el plan de actuación.

La idea de la matriz de evaluación surgió en el ámbito de la educación, especialmente con el impulso de los modelos “comprehensivos” que tratan de valorar lo que los escolares aprenden efectivamente. El éxito en este ámbito ha sido tal que incluso existen páginas WEB con modelos de matrices para aplicar en cualquier asignatura, actividad y nivel educativo (por ejemplo: <http://rubistar.4teachers.org>). Desde el ámbito educativo esta manera de trabajar ha tenido una fuerte expansión en ámbitos como las ingenierías, así en áreas tan diversas como en los Estudios de Impacto Ambiental, el “Strategic Management” o la Gestión de Emergencias y, por supuesto, en la mayoría de modelos de calidad. Esta expansión ha producido múltiples “herramientas” de gestión de matrices (MDI, EFE, EFI, DOFA, PEYEA, BCG, I/E, MCPE,.... así como muchas “empresas” que lanzan su propia herramienta (con su nombre), pero esto, aparte de que el infinito número de alternativas es ya una locura, resulta incluso contraproducente para un correcto acercamiento al conocimiento de la evaluación, porque proporcionan elaboraciones demasiado cerradas y en ocasiones incompatibles entre sí, que en nada benefician a una buena comprensión de lo que es la evaluación y en particular el modelo de evaluación fuerte.

En todo caso, en paralelo y a partir de la experiencia en el ámbito de la educación (en parte porque muchos de los primeros evaluadores eran pedagogos), comenzaron a utilizarse las matrices de evaluación y en la actualidad la mayor parte de evaluaciones se realizan a partir de esta metodología.

De manera muy simple podemos decir que las matrices se encabezan (en la primera fila horizontal) con cada uno de los componentes del diseño metodológico de la evaluación, por ejemplo los objetivos, las preguntas a formular, los indicadores que se van a utilizar, las fuentes de información, los criterios metodológicos para obtener los datos, los responsables de las diversas tareas, los resultados esperables, la temporalización de cada tarea,...

También deben incluir todos los objetivos previstos (en la primera columna vertical), lo que de nuevo refuerza la idea de tener bien claros los objetivos específicos. Después hay que completar la matriz en todos sus recuadros, de tal manera que se proporcione una respuesta concreta, en cada objetivo a cada uno de los componentes del diseño metodológico.

Así, en el ejemplo citado más abajo vemos cómo cada fila de la matriz debe incluir un objetivo, el cual se desagrega en una o varias preguntas. Cada pregunta debe incluir al menos un indicador y cada indicador debe incluir al menos una fuente de verificación y la fuente debe quedar bien especificada de acuerdo con una técnica de recogida o producción de datos. El proceso, de izquierda a derecha, produce datos y los datos regresan, de derecha a izquierda, para dar cuenta del objetivo. Una matriz es, por tanto, “un molde” que supone una “forma” que se rellena adoptando dicha forma. Una matriz es como el molde de una figura de Botero que rellenas de metal para que, una vez roto el molde, tengamos la correspondiente figura.

El dato, como iremos viendo, es una puntuación, o un conjunto más o menos complejo de información. Aunque también es cierto, en especial en el campo de la economía y las ingenierías, que los datos son cada vez más valoraciones cerradas, porque los indicadores son digitales (no/si) o porque son escalas de puntuaciones. Esto, como veremos, no parece fácil de trasladar a las Políticas de Juventud (ni a ninguna política social), salvo si asumimos que tenemos un grado de conocimiento absoluto y definitivo y nuestros recursos nos permiten ser perfectamente efectivos.

Mientras que esta posibilidad sigue en estudio, la opción más realista es la de matrices abiertas. Vamos a elaborar un ejemplo de matriz abierta y simple, para facilitar la comprensión de la misma.

Se trata de evaluar la eficiencia (ver definición) del ciber del Centro de Juventud que lleva funcionando unos años. Se trata por tanto de una evaluación de resultados que tiene como objetivo general “considerar cuáles son las mejoras más adecuadas para aumentar la eficiencia del servicio”. El cuadro 3.3 nos muestra la matriz de evaluación con los contenidos más frecuentes de la misma (objetivo, pregunta, indicador, fuente), aunque no hemos incluido otros tan importantes como “responsable” y “tiempo” porque son fáciles de entender y alargarian demasiado el cuadro para el tamaño del libro. En la práctica una hoja de “Excel” es magnífica para establecer la matriz.

Siguiendo el criterio de sencillez del ejemplo nos hemos limitado a fijar seis objetivos, aunque a cualquier equipo en una reunión un par de horas se les ocurren al menos otros seis. Pero cuidado no vale sólo con fijar objetivos, sino que hay que formular preguntas, establecer los corres-

**Cuadro 3.3. Un modelo de matriz de evaluación.**

OBJETIVO	PREGUNTA	INDICADOR	FUENTE
Conocimiento del ciber entre los jóvenes	¿Se conoce la existencia del CIBER, su dirección, los horarios y las normas de uso?	Si/No a cada cuestión	Pregunta en encuesta general sobre servicios municipales.
Nivel de utilización del ciber	¿Cuántos jóvenes diferentes han utilizado el ciber y con qué frecuencia?	% de jóvenes que utilización del Ciber y número de veces que a hacen a lo largo del año	Registro del CIBER
Características de los usuarios del ciber	¿Quiénes son los que utilizan y los usuarios más frecuentes?	Datos sociodemográficos, educativos y de lugar de procedencia (barrio)	Ficha Registro para acceder a la condición de usuario CIBER
Percepción del personal técnico sobre la eficacia del servicio.	¿Piensan los técnicos del Ciber que la oferta es adecuada a las necesidades de los jóvenes de la localidad?	Valoración en una escala de diferentes aspectos del servicio: Publicidad, facilidad, ergonomía,...	Cuestionario (o grupo de discusión) a los técnicos de juventud y del CIBER.
Comparación con otros servicios similares	¿Se trabaja y se oferta el servicio mejor o peor que otros similares en otras localidades?	Posición relativa en relación a otras evaluaciones.	Información y documentación sobre otras evaluaciones.

pondientes indicadores, las fuentes y los otros contenidos entre los mencionados más arriba (si es el equipo quien decide sería muy pertinente fijar el responsable de cada tarea) y además hay que tener en cuenta que para el análisis de los resultados de la matriz se requerirán otros datos como por ejemplo, costes, personal, horarios,... cuestiones nos vamos a ocupar más adelante.

En todo caso este modelo simple, pero flexible en el Excel, que trata de ser realista en relación con nuestras capacidades y recursos, puede, según la literatura de la evaluación resultar muy complejo. Así por ejemplo Robert Stake propone un modelo con dos matrices y una base lógica. La base lógica es la filosófica del programa a la que debe ajustarse el contenido de las matrices. Una de las matrices es descriptiva (como la que acabamos de proponer para el caso del Ciber), pero la otra es valorativa y se refiere a los enjuiciamientos (podrían ser los dos últimos objetivos del ejemplo, pero Stake se refiere más bien a las “normas”, es decir al “reglamento” de funcionamiento del Ciber y a las valoraciones que les atribuyen los diversos agentes involucrados).

A la vez Stake divide cada una de estas matrices en tres campos o niveles diferentes: objetivos o normas relacionados con los antecedentes ¿porqué se hizo lo que hizo?, procesos (¿qué se está haciendo?) y resultados (los efectos del programa). Al final este complicado modelo nos proporciona una visión muy clara del grado de congruencia del programa, lo cual, para Robert Stake es la verdadera evaluación. Se trata de un modelo que además tiene otras muchas complejidades, que fue elaborado en los años 60, que se ha desarrollado desde una perspectiva cualitativa y casi excluidamente en los ámbitos educativos, pero que he citado para mostrar como de complejas pueden ser algunas matrices. Aunque creo que en los ámbitos de juventud, debido a su propia complejidad hay que apostar, en un primer momento por matrices sencillas.

### 3.5.- Los requisitos de la evaluación.

Los epígrafes precedentes nos han permitido entender las líneas básicas y el concepto de la evaluación fuerte. Alcanzar esta comprensión de la evaluación ha requerido un cierto esfuerzo, porque, como ya hemos explicado, hace unos pocos años, al menos en España, las cosas eran un tanto confusas y nos hemos visto obligados a evolucionar desde la visión metodológica contenida en la fase III del cuadro 1.1, referida a la “visión de la administración” (GrupoGID, 1994; Comas, 1992) hasta la fase IV, referida a la “visión ética, política y social” (Comas, 2001). Como parte de este proceso tuvieron que asimilarse ciertos requisitos (cuadro 3.4), expresados como estándares y creados por Comités de Evaluación.

Una de las propuestas más consensuadas y que por ello ha gozado también de gran predicamento ha sido la realizada por el *Joint Committee on Standards for Educational Evaluation*. En 1974, un comité nombrado por la *American Educational Research Association*, la *American Psychological Association* y el *National Council on Measurement on Education* y presidido por D.L. Stufflebeam, sugirió a estas organizaciones la elaboración de un conjunto de estándares con los cuales valorar la práctica de la evaluación educativa. El trabajo del comité fue complejo, debido a sus divergencias corporativas, pero al final el *Joint Committee* recogió los resultados de sus reuniones en la obra *Estándares para la evaluación de programas, proyectos y materiales educativos*, publicada en 1981<sup>21</sup>.

Estos estándares fueron sometidos a un proceso de contrastación por parte de especialistas en el campo de la evaluación y de las Ciencias Sociales, así como por profesionales de la educación, antes de que su utilización fuera recomendada (Joint Committee, 1981 y 1994).

Tales estándares definen la concepción del Comité sobre los principios y requisitos que debería reunir toda evaluación. Dentro de su concepción de la evaluación, estos Estándares (30

---

<sup>21</sup> Conviene tener en cuenta que estos estándares educativos no son los únicos y que en el propio ámbito de educación aparecen diversos proyectos de éxito, por ejemplo en España se han difundido muy bien los estándares OTL (Opportunity To Learn), aunque no parece fácil proyectarlos a otros ámbitos y además los estándares del Joint Committee gozan de un merecido prestigio internacional (Fernandez, 2001).

en total), sintetizan toda una serie de requisitos que, sin duda alguna, influyen y optimizan la validez de las evaluaciones. Es evidente que estos Estándares por sí solos no pueden garantizar o asegurar la calidad de toda evaluación, pero sí constituyen un instrumento-guía imprescindible para la práctica de la evaluación en el ámbito educativo, mucho más si tenemos en cuenta que se trata de Estándares de evaluación que son el resultado de un proceso permanente y extensivo de actualización y revisión.

Según dicho Comité Conjunto, los Estándares se organizan en torno a cuatro atributos principales que validan la calidad de la evaluación: los estándares de Utilidad, los de Viabilidad, los Decoro (o Propiedad) y los estándares de Precisión, cuya definición aparece en el cuadro 3.4. En el anexo 1 aparecen detallados los mencionados 30 estándares propuestos por el Comité.

#### Cuadro 3.4. Requisitos para la evaluación.

---

**UTILIDAD:** *Hay que asegurar que una evaluación satisfará las necesidades de información práctica de los usuarios.*

---

**VIABILIDAD:** *Hay que asegurar que una evaluación será realista, prudente, diplomática y frugal.*

---

**DECORO:** *Hay que asegurar que una evaluación es llevada a cabo de manera legal, ética y con el conveniente respeto hacia el bienestar de las personas involucradas en la evaluación, así como de aquellas afectadas por los resultados de la misma.*

---

**PRECISIÓN:** *Hay que asegurar que la evaluación revelará y expresará información técnicamente adecuada sobre los rasgos que determinan el mérito o valor del programa evaluado.*

---

Fuente: JCSEE (1981/1994).

Como ya hemos dicho estos cuatro requisitos básicos (que conviene aprenderse bien), aparecen desarrollados de forma completa en el anexo 1. Pero también se ha añadido al texto como anexo 2 los estándares recientemente aprobados por Naciones Unidas para las evaluaciones realizadas en su seno y que representan, tanto en términos éticos como técnicos, el más avanzado consenso actual, desde las referencias que ofrece la comunidad y el sistema de organismos internacionales, en torno a este tema. Como parte de las utilidades de este Manual conviene revisar atentamente estos estándares, ya que de hecho, todo el contenido del mismo trata de prepararnos para que podamos entenderlos. La posible comparación entre los estándares del *Joint Committee on Standards for Educational Evaluation* (que ya tiene 25 años) y los estándares propuestos por Naciones Unidas nos ayudan a comprenderlos mejor.

En cualquier caso adoptamos estos cuatro requisitos del Joint Committee porque esperamos que la evaluación sea útil y práctica, ayude a mejorar los programas, permita cambiar las condiciones sociales, aumente el grado de información, cambie las expectativas de los participantes y finalmente ayude a tomar buenas decisiones a los responsables técnicos, políticos o administrativos a los que se ha encomendado la gestión del programa. Es decir, evaluamos porque pensamos que **es una buena manera de conseguir que las cosas adquieran un cierto valor para la colectividad.**

A esta forma de entender la evaluación también se la ha llamado “evaluación realista”, porque implica una permanente retroalimentación entre los aspectos teóricos (Por ejemplo: ¿Porqué hacer una política de juventud afirmativa?), la hipótesis (¿Podemos impulsarla ofreciendo servicios a los jóvenes?) y los datos recogidos en el trabajo de campo, así como su análisis en el contexto específico del programa, para volver a reformular la teoría.

Podemos afirmar que la evaluación realista es una posición radical de la evaluación fuerte porque implica una serie de reglas (cuadro 3.5), que parecen una condensación de los estándares de evaluación del JCSEE, citados más arriba, pero adoptando una posición “radical, crítica y política”. Esta postura ha sido defendida por diversos autores (Pawson y Tilley, 2000) y conforma una opción metodológica que va más allá de los presupuesto de la evaluación fuerte, la cual contrasta de una forma absoluta con los viejos ideales de la evaluación experimental. En este Manual también nos posicionamos fuera del modelo de la evaluación experimental (aunque respetamos

**Gráfico 3.1:**  
**La retroalimentación lógica de la evaluación realista**



como técnica el diseño experimental), pero no asumimos sin más y de forma estricta las reglas de la evaluación realista.

Sí asumimos, en cambio, un precepto básico de la evaluación realista que afirma que la evaluación supone un “rechazo a cualquier forma de ortodoxia metodológica, trazando un camino hacia la adecuación metodológica como un criterio básico para juzgar la calidad de la metodología empleada” (Patton, 1990), aunque asumiendo, como otros autores, que la evaluación realista es lo mismo que la evaluación pragmática, para que el realismo no nos conduzca hacia un posible relativismo en las conclusiones. Las reglas que aparecen en el cuadro 3.4 tratan de evitar este problema, pero a costa de convertir al evaluador en una especie de militante de la acción social.

### 3.6.- El papel del evaluador.

En el desarrollo de las reglas de la evaluación realista de Pawson y Tilley aparece una oposición frontal al inductivismo al que califican de “deficiencia experimental” de los evaluadores positivistas, así como a la “generalización burocrática” de los resultados. Para ellos la “lección realista” reduce la importancia de los instrumentos y las metodologías experimentales para fijar la mirada sobre quién y porqué ha realizado la evaluación, es decir, no se trata tanto de una cuestión de “ciencia positiva”, sino de saber quiénes son los evaluadores y quiénes les han encargado la evaluación (Pawson y Tilley, 2000).

Pero esta opción, que escenifica una reacción pendular frente a los excesos del modelo positivista de la ciencia experimental de la fase II de la historia de la evaluación (cuadro 1.1), contiene un toque de exageración. Porque al definir la evaluación hemos dicho con claridad que se trata de “prácticas fundadas en el método científico”, lo que implica la necesidad (una necesidad esencial) de reconocer y utilizar el inductivismo. Aunque la opción realista de Pawson y Tilley contiene una crítica acertada a los excesos de empirismo positivista y del inductivismo (del modelo experimental de evaluación), su conclusión no es la adecuada.

En una gran medida porque otorga un gran protagonismo a las capacidades y competencias teóricas e intelectuales del evaluador, para que a la vez, y de forma paradójica, éste “otorgue” protagonismo a los evaluados, es decir, sitúa la responsabilidad de la participación en el propio evaluador que “debe proporcionar la palabra a profesionales y usuarios”. La paradoja implica la presencia de un profesional idóneo que no sólo domina un conocimiento instrumental, sino que

#### Cuadro 3.4. **Requisitos para la evaluación.**

---

*Regla nº1: Capacidad de cambio.*

*Los evaluadores deben prestar atención al potencial de cambio de los programas sociales.*

---

*Regla nº2: Profundidad analítica.*

*Los evaluadores necesitan profundizar más allá de la superficie, más allá de la información de entrada y de los resultados observables del programa.*

---

*Regla nº3: Mecanismos causales.*

*Los evaluadores necesitan averiguar cuáles son los mecanismos causales que generan problemas sociales y de comportamiento y cómo pueden ser eliminados o contrarrestados a través de procesos alternativos introducidos en un programa social.*

---

*Regla nº4: Contextos.*

*Los evaluadores necesitan entender los contextos dentro de los cuales los problemas se activan, así como los mecanismos de los programas que pueden eliminarlos de manera exitosa.*

---

*Regla nº5: Resultados.*

*Los evaluadores necesitan entender cuáles son los resultados de una iniciativa y cómo se producen.*

---

*Regla nº6: Configuraciones C.M.R.*

*Para desarrollar lecciones transferibles y acumulativas de las investigaciones, los evaluadores necesitan orientar sus opiniones sobre la configuración del modelo contexto-mecanismo-resultados (Configuraciones C.M.R.).*

---

*Regla nº 7: Procesos Profesor-alumno.*

*Para crear y probar las explicaciones del modelo contexto-mecanismo-resultados los evaluadores se deben comprometer en una relación profesor-alumno con los que diseñan la política del programa, practicantes (profesionales/trabajadores) y participantes.*

---

*Regla nº8: Sistemas abiertos*

*Los evaluadores necesitan reconocer que los programas se implementan en un mundo social cambiante y permeable y que la efectividad del programa puede ser así subvertida o mejorada a través de una intrusión no anticipada de nuevos contextos o nuevas fuerzas de causa.*

---

en términos de Cronbach “posee una mentalidad crítica, erudita e ilustrada”. Es decir, ni más ni menos, se propone para la evaluación el modelo leninista de “vanguardia consciente” para promocionar la participación y “otorgar el poder” a la nueva clase de los profesionales.

Pero además, como la idoneidad de la condición de evaluador no viene establecida por una determinada titulación académica, ocurre que los evaluadores se autentican a sí mismos como tales a partir de sus peculiares capacidades, lo que añade leña a la paradoja. Porque el evaluador, que es inevitablemente un sujeto externo, asume un “rol intelectual”, que contiene rasgos jerárquicos en la relación con los profesionales (y no digamos los usuarios) que va a evaluar. Así, cuando un organismo de juventud vaya a realizar una evaluación, se “debe” entregar a un “agente externo” que “movilizará la participación” de los profesionales del organismo, que son los que realizarán la evaluación. Cualquier responsable de un programa, plan u organismo de juventud sabe que esto no es cierto y que los agentes externos que llegan de repente y adoptando el rol de “sabios”, requieren un tiempo de adaptación (y conocimientos específicos) para motivar al equipo.

Como consecuencia esta auto-atribución del rol no es la más adecuada o no lo es del todo. En realidad se ha producido una idealización “democrática” que trata de evitar los rasgos más extremos del viejo modelo jerárquico, cuando el evaluador era alguien que unía a condición de sujeto externo la utilización de metodologías experimentales, casi sagradas y para las que sólo él poseía conocimientos suficientes para poder aplicarlas.

Tal idealización explica la presencia de una cierta obsesión para tratar de evitar que el evaluador vuelva a adoptar un papel jerárquico, lo que ha introducido múltiples cautelas en la definición de quien debe ser el evaluador y como debe actuar (Garnier y Przewerski, 2004), que se condensan en las reflexiones en torno a lo que significa ser “crítico” y ser “ilustrado”. Dos términos cuyo contenido nos permiten superar los riesgos de la idealización y de un planteamiento leninista (que nunca supo, una vez la vanguardia comenzó a administrar el poder en nombre del proletariado, cómo devolvérselo de forma efectiva). El evaluador es por tanto un agente externo, que realiza

una tarea bien definida, lo que supone, durante un tiempo, adoptar un rol con algunos contenidos jerárquicos, pero que tiene que auto-limitarse por el establecimiento de un mecanismo consensual y participativo, asumiendo una condición crítica e ilustrada.

Para Ernest House **ser crítico** supone formar parte de un consenso en torno a valores democráticos fundamentales, a manejar un sentido preciso de la justicia y la equidad, así como a apostar por el cambio y el progreso social. Pero cuando se trata de desarrollar, de forma operativa, los diferentes parámetros de esta definición, las cosas no son tan sencillas. Así, por ejemplo, el concepto de equidad (cuadro 3.5) ha sido expresado de una forma clara y sencilla. Pero los otros términos resultan un poco más complicados.

**Cuadro 3.5. Las condiciones de equidad en evaluación.**

- *Ausencia de coerción.*
- *Racionalidad en la aplicación.*
- *Aceptación colectiva de los términos y condiciones de aplicación.*
- *Desinterés personal o grupal, universalidad y utilidad colectiva.*
- *Información transparente, completa e igualitaria.*
- *Ausencia de riesgos individuales y colectivos.*
- *Posibilidad (suficiencia de recursos sociales y tecnológicos).*
- *Considerar todas las opiniones y optimizar la participación.*

Adaptación de House, 1994.

Por su parte, Elliot Eisner plantea la cuestión del “**evaluador ilustrado**” como el del sujeto reflexivo que traslada a lo público la esencia de la intervención especializada, proporcionándole un nuevo sentido que trasciende tanto a la representación social como a las prácticas profesionales (Eisner, 1991). Este es sin duda el papel que puede desempeñar un evaluador externo en relación a un Plan o un programa de juventud, es decir, puede adoptar la propuesta de Ernest House en el sentido de “*tratar de compartir la virtud pública*”. De algún modo de se trata de asumir que el evaluador sólo mira por encargo, trabaja para la colectividad, no determina cuáles son los valores por sí mismo, sino que ayuda a esta colectividad a satisfacer su necesidad de atribuir valor a las cosas (House, 1994). Cuando el evaluador se sitúa al margen de la colectividad, considera que sus valores y los de la sociedad van a ser distintos, ya no evalúa, ya no da valor a las cosas para todos, sino que juzga y al contraponer su juicio al de la colectividad se equivoca. Aunque, lógicamente, retiene algo de poder.

En cualquier caso, la evaluación realista ha provocado un debate sobre la ética de la evaluación y el papel de los distintos actores en la misma, que nos ha permitido finalmente definir un perfil adecuado para el evaluador. No voy a describirlo aquí porque aparece en el anexo 2, de una forma muy clara, en forma de un consenso sobre “competencia y ética”, que agrupa un total de 23 estándares muy clarificadores. Hay que repasarlos, asumirlos y aplicarlos. No va a ser fácil, pero es una de las prioridades de los próximos años.

### **3.7.- El ajuste de los tiempos: el desarrollo de la intervención y el cronograma de la evaluación.**

Entramos en la cuestión de los tiempos, un tema delicado al que se le presta poca atención, pero que ocasiona algunos problemas en la evaluación de las intervenciones. En general las evalua-

ciones realizadas son postfacto porque no han sabido resolver el problema del ajuste temporal. Es decir, una buena evaluación requiere realizar diferentes tareas mientras se va desarrollando la intervención, lo que implica una previsión previa que forma parte de la planificación del programa. Esto significa que cada tarea hay que realizarla en el momento preciso y si se retrasa implica una pérdida irreparable para el procedimiento de evaluación.

Por este motivo resulta siempre necesario realizar un cronograma de la evaluación y su relación con los diferentes aspectos y momentos del programa. El cuadro 3.6 muestra, en forma de matriz, un cronograma típico, en el eje vertical aparecen las diferentes fases del programa y en el horizontal las de la evaluación.

Para mostrar un cronograma muy sencillo (pero típico) imaginemos que para 2008 vamos a poner en marcha un nuevo programa de convocatorias para el apoyo a la promoción de la participación de los jóvenes a través de las organizaciones juveniles y que pretendemos evaluarlo para saber hasta que punto la vía elegida responde al objetivo de “promocionar la participación de los jóvenes”.

Veamos el cronograma: en la primera columna del programa aparecen en mayúsculas los principales hitos del mismo, incluidos dos (en minúsculas) que se corresponden con la evaluación. En la segunda columna aparecen las fechas en las que se desarrollan las diferentes fases del programa que se extienden en un periodo de casi dos años. Aparece una fase previa de consulta a las organizaciones (no del texto de la convocatoria sino de cómo debería ser el programa, porque consultar un texto ya elaborado es un acto más administrativo que evaluativo), que ya forma parte de la evaluación (recogida de datos, análisis e informe).

La fase de planificación es común y durante la misma se pueden realizar evaluaciones de evaluabilidad, cobertura y diseño. Después, en enero de 2008 se realiza la publicación en del correspondiente Boletín Oficial, comienza el plazo de presentación y después el de resolución. Con la publicación hay que preparar los instrumentos y en el momento de la selección de los proyectos hay que registrar los mismos (y sus principales características), mientras se realizan los proyectos a lo largo de 2008, se puede realizar algún tipo de observación participante (se explicará en que consiste en la parte siguiente), los resultados de la observación pueden ser analizados al final de la fase y dan lugar a un informe parcial en diciembre de 2008.

Una vez realizados los proyectos aparece una fase de justificación (que incluye las correspondientes memorias de dichos proyectos) que debería realizar cada organización subvencionada. A partir de los datos de las memorias se podría establecer una muestra de nuevos afiliados en cada organización y en enero/marzo de 2009 realizar una pequeña encuesta (actitudes previas, como recibieron la actuación, motivación inducida, satisfacción expectativas,...) entre los nuevos afiliados. Los análisis de todo este material se realizarían en febrero y marzo de 2009, el 30 de abril estaría concluido el informe, que se podría presentar el 20 de mayo y a partir de este momento se comenzaría con la difusión.

La perfección de la evaluación se habría alcanzado si antes de iniciar este programa, a mitad de 2007 se hubiera realizado una encuesta (o unos grupos) al conjunto de la población joven de la localidad (o la Comunidad) en torno a su opinión sobre el tema de la participación y las organizaciones juveniles y que el enero/febrero de 2009 se hubiera repetido la consulta para comparar el antes y el después.

Todo esto cuesta un dinero que se supone no se aplica al objetivo, pero a la vez hay que tener en cuenta:

- Que la difusión de la evaluación visibiliza y prestigia la imagen de la entidad territorial (el municipio o la Comunidad) con tanta o mayor intensidad que otras actuaciones (publicitarias, turísticas o culturales) cuyo coste será seguramente mayor.
- Que el coste de la evaluación se amortiza en un incremento de la eficacia de las posteriores convocatorias.

Cuadro 3.6. **Un modelo de matriz de evaluación.**

PROGRAMA	FECHAS	PLANIFICACIÓN EVALUACIÓN	RECOGIDA DE DATOS	ANÁLISIS	INFORME	PRESENTA DIFUSIÓN
<b>CONSULTA A LAS ORGANIZACIONES</b>	SEP/OCT 2007		Sep/oct 2007	Nov 2007	20 dic 2007	
<b>PLANIFICACIÓN Y PREPARACIÓN DE LA CONVOCATORIA</b>	NOV/DIC 2007	Nov/dic 2007				
<b>Medidas de diseño, evaluabilidad o cobertura</b>		Nov/ dic 2007			Nov/dic 2007	
<b>PUBLICACIÓN</b>	ENE 2008	Registro				
<b>PLAZO DE PRESENTACIÓN</b>	20 FEBRERO 2008					
<b>SELECCIÓN PROYECTOS</b>	20 MARZO 2008		Registrar			
<b>REALIZACIÓN DE LOS PROYECTOS</b>	MAR/DIC 2008		Observación participante	Oct 2007	Dic 2007	
<b>JUSTIFICACIÓN Y MEMORIA DE LOS PROYECTOS</b>	20 ENERO 2009		20 de enero de 2009			
<b>Encuesta a nuevos afiliados</b>			En/feb 2009			
<b>Informe de evaluación</b>				Feb/mar 2009	30 de abril 2009	20 mayo 2009

Para completar esta reflexión sobre los cronogramas podemos imaginar un proceso monitorizado muy cerrado y podemos pensar en un proceso abierto en el que el cronograma se va ajustando a los diversos momentos y cambios del programa. Lo primero exige un trabajo previo muy intenso y la principal dificultad reside en la dificultad de afrontar cualquier cambio. Lo segundo parece más fácil pero requiere la existencia de una figura (individual o grupal; interna o externa), con una información precisa y continua y una cierta autoridad para tomar decisiones sobre el proceso (lo que más adelante llamaremos evaluación formativa).

En general, los programas de juventud están formalmente establecidos por las administraciones públicas, pero a la vez, en el nivel local se producen cambios relacionados con demandas o con reacciones sociales, lo cual nos obliga a mantener un cronograma “entreabierto”.

En todo caso y a modo de ejemplo extremo unos colegas con los que mantengo una relación muy antigua diseñaron para una evaluación muy abierta y reprodujeron post-facto el verdadero cronograma de la evaluación muy peculiar y no- reproducible en este texto (Quintero y Perez, 1988). El programa se refería a una formación para el profesorado, durante el curso académico y la evaluación incluía cuatro fases: planificación, diagnóstico, procesos, resultados e impacto. El cronograma muy similar al del cuadro 3.6, incluía casillas semanales y se representaba mediante líneas que ponían en relación abscisas y ordenadas. La planificación era una espiral que partía de la primera casilla y acababa en la última; el diagnóstico aparecía al principio pero volvía a salir, como el Guadiana, en algunas otras; los procesos si eran una línea recta pero esto era en realidad porque los procesos de la evaluación se habían

subsumido en la intervención, los resultados y el impacto mostraban líneas titubeantes que volvían al punto de partida.

¿Por qué pongo este ejemplo?, para mostrar dos cosas, la primera que el establecimiento de un cronograma es muy importante y lo segundo que su efectivo cumplimiento no es fácil y que hay que adoptar una cierta flexibilidad porque los programas sociales (incluidos los de juventud) exigen una cierta pericia adaptativa en cuanto a plazos y ajuste al programa.

## **BIBLIOGRAFÍA Y DOCUMENTACIÓN COMPLEMENTARIA.**

Muchos de los contenidos de este capítulo aparecen en los libros recomendados con anterioridad, las dos recomendaciones siguientes reflejan, de una parte, una visión más directa y práctica de los conceptos manejados en este capítulo y, de otra parte, el punto de vista, siempre peculiar, de los economistas.

- AGUILAR, M. J. y ANDER-EGG, E. (1992), **Evaluación de servicios y programas sociales**, Madrid, Siglo XXI.
- PASCUAL, J. (1999), **La evaluación de políticas y proyectos: criterios de valoración económicos y sociales**, Barcelona UAB / Icaria

**PARTE II**  
**LOS MÉTODOS Y LAS TÉCNICAS DE RECOGIDA**  
**Y PRODUCCIÓN DE LOS DATOS**

*“El universo, real e inteligible, puede ser presentado desde un número infinito de perspectivas, y el número de posibles sistemas del saber humano es tan amplio como el de estas perspectivas”*

Denis Diderot, artículo “*Enciclopedia*” en “**ENCYCLOPÉDIE ou dictionnaire raisonne des sciences des arts et des métiers**”, Vol V, 1755.



#### 4.1.- Estrategias empíricas.

Fue Alfonso Ortí quien, a mitad de los años 90, realizó una propuesta articulada de un diseño metodológico completo, multidimensional y concreto para la investigación, análisis e intervención social, para el cual distinguió diferentes momentos que representan, a la vez, todas las posibles estrategias empíricas de aproximación a la realidad social (Ortí, 1993). Debates posteriores nos permiten clasificar estos momentos en ocho tipos de estrategias que podemos jerarquizar de acuerdo con categorías de exigencia profesional.

1. Análisis secundario de fuentes de datos.
2. Observación y observación participante.
3. Entrevistas.
4. Registro y análisis de casos.
5. Informantes clave y cualificados.
6. Narraciones e historias de vida.
7. Encuestas por cuestionario.
8. Grupos de discusión.

Más adelante, el mismo autor, en una cita que hemos incluido en el capítulo anterior, explicó que todos estos momentos metodológicos eran en si mismos “deficientes”, acuñando la idea de que los diferentes métodos y técnicas de recogida y producción de datos eran “**complementarios por deficientes**”. Se trata además de una idea, la complementariedad, que lleva de forma inevitable aparejada la idea de **articulación**, porque todo esquema que pretende establecer sinergias entre sus componentes requiere algún tipo de articulación entre los mismos (Ortí, 1995). Tal articulación se expresa sobre un esquema con sucesivos niveles, momentos o prácticas, de aproximación empírica a la realidad. Desde la perspectiva de las necesidades de la evaluación voy a seguir el esquema propuesto por Alfonso Ortí, pero modificando la atribución jerárquica entre los momentos empíricos y las técnicas de recogida de datos. No tanto porque no crea en la bondad de dicha jerarquía, sino porque considero que en el ámbito de la evaluación conviene una visión más lineal de la relación entre los diferentes métodos y técnicas.

Esta opción lineal obedece a una lógica muy particular, ya que en la evaluación se vive la permanente contradicción entre, por una parte, la razón instrumental de las políticas sociales y, por

otra, los intentos de los diferentes actores sociales por manipular estas mismas políticas (un tema del que hemos hablado amplia y provocativamente en el prólogo). Por este motivo, el diálogo entre los momentos y los instrumentos empíricos no se fundamenta en la epistemología, sino en los componentes del estilo de trabajo de la evaluación, lo cual supone, como hemos explicado en las páginas precedentes, una visión muy pragmática del método científico, lo que a su vez implica que el valor de las diferentes técnicas y métodos no debe situarse sobre un esquema pre-establecido de carácter general, sino sobre el contexto particular de cada práctica evaluativa. Es decir, cada método, en relación a la evaluación de cada programa o intervención social en particular, sirve para lo que sirve en el contexto del propio programa.

Por tanto, esta opción lineal implica una especie de “subjetividad operativa” que en sí misma supone dotar de funcionalidades concretas de los diversos métodos y técnicas de investigación, en relación al contexto del programa a evaluar. Es decir, el Grupo de Discusión o la Encuesta no son “una categoría para”, sino que para evaluar el programa X se requiere contar con estas u otras técnicas concretas. Parece lo mismo, pero no lo es. Porque no son las técnicas las que eligen el programa, sino el programa el que elige las técnicas. No existe por tanto el evaluador experto en Grupos, en Encuestas o en Observación Participante, sino que el evaluador debe poder elegir cualquier técnica en función del programa y los objetivos de la evaluación. Un evaluador que “tenga” técnicas preferidas, no es un buen evaluador.

Conviene también considerar que tal “subjetividad operativa” de las técnicas y métodos de recogida y producción de datos representa el necesario correlato técnico a la noción de evaluación fuerte (y a la visión pragmática y realista de la misma). En la definición de evaluación (capítulo I) hemos hablado de “prácticas fundadas en el método científico”, pero también hemos explicado que la evaluación no es investigación, sino una práctica social y política con una finalidad diferente: mejorar la intervención social.

Se trata por tanto de una práctica “particular y diferente de la investigación”, lo cual nos obliga a revisar la relación entre los términos del método científico y los términos logísticos de las técnicas de recogida de datos. En el ámbito de la investigación por “metodología” se entiende, de forma habitual “la aplicación rigurosa y exclusiva” de una técnica, lo cual no es del todo cierto, porque siendo verdad no es toda la verdad. Por este motivo, en evaluación debe entenderse por “metodología”: una “adecuación a las necesidades” para obtener una información fiable y válida, en relación a las preguntas formuladas por Kant. Es decir, cada una de las técnicas de recogida de datos y su asignación a una determinada metodología, responden a lógicas políticas, sociales o culturales, más que a argumentos conceptuales y metodológicos, porque tales argumentos pertenecen a otra fase del diseño de la evaluación y no al momento de la recogida de los datos. Por este motivo, convertir la técnica de recogida de datos en un argumento teórico supone menospreciar los criterios científicos de la evaluación (Babbie, 1995).

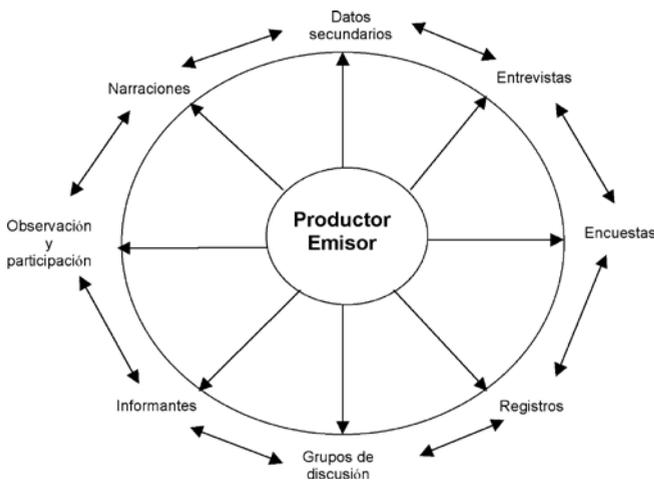
Esto significa que el carácter científico de la evaluación no se justifica sobre el rigor en la recogida de los datos (aún siendo necesario), sino sobre un adecuado planteamiento del diseño de la misma (que a la vez debe ser pragmático). No se trata de una opción novedosa que me acabe de inventar, porque las relaciones entre teoría e instrumentos, entre método y logística, es más la historia de los artesanos que “descubrieron estos métodos” que de los teóricos que los propusieron a posteriori (Taylor y Bogdan, 1984).

#### **4.2.- La “tabla redonda” de los métodos y las técnicas de recogida y producción de datos.**

En la propuesta original Alfonso Ortí hablaba de una “jerarquía metodológica”, aunque después añadió la idea de la complementariedad por deficiencia. Si en evaluación debemos renunciar a la jerarquía pero manteniendo la idea de la complementariedad, la mejor manera es recurrir a una “tabla redonda” en la cual cada método ocupa un lugar equivalente según la tradición de

la retórica artúrica, aunque Arturo el Rey (la encuesta) y Lancelot, el segundo caballero pero el más poderoso (el grupo de discusión), no son exactamente como los demás caballeros aunque se sienten en la misma mesa.

La figura 4.1 nos muestra los componentes de dicha tabla redonda. En el centro de la misma figura el emisor (o el productor), es decir un colectivo social más o menos amplio que tiene ciertas opiniones y creencias, que adopta actitudes y que realiza diversos comportamientos que hay que recoger para analizar. En el borde de la mesa se distribuyen las diversas técnicas de recogida de datos, técnicas que implican la producción, elaboración, gestión y codificación de estos datos. Sin estas técnicas el emisor se mantiene mudo e invisible. Las técnicas implican por tanto procedimientos de mediación imprescindibles para visualizar y escuchar al emisor. Cada técnica representa una forma, un arte, una tradición distinta, para recoger y orientar la producción del emisor.



**Figura 4.1:**  
**La tabla redonda de los métodos de investigación**

Alrededor de esta tabla redonda se mueven los evaluadores que recogen los datos producidos por todos estos métodos. Pero, por regla general, existen dos tipos de investigadores sociales. De una parte los que **se sientan inmóviles** en un puesto prefijado y ocupan siempre el mismo lugar frente a una técnica concreta y “se especializan” en la misma, ignorando a las otras técnicas de recogida de datos. Delante de ellos se yergue un sistema cerrado de procedimientos que dominan con mucha precisión y con un paradigma de rigor que confunden con el método científico en su conjunto. Son investigadores fijos, muy competentes en relación al dominio de su método, pero inadecuados como evaluadores.

Otros investigadores prefieren en cambio moverse, **pasear** alrededor de la tabla, quizá observar desde lejos, sentarse en ocasiones y durante un tiempo limitado en un puesto fijo, pero sin renunciar a una actitud **peripatética** (es decir, quizá paseándose como Aristóteles o quizá saltando como un “mono extravagante”, según la otra acepción del DRAE), que les permite combinar los diferentes métodos y técnicas de recogida de datos. Se trata de un paseo que puede parecer azaroso según el modelo positivista, pero que está orientado por criterios científicos, porque utiliza los diferentes instrumentos según los necesita, es decir, según le convengan al diseño previo de la evaluación. Estos son investigadores **pedestres**, muy adecuados como evaluadores.

El recorrido de un evaluador puede iniciarse con la simple mirada, seguir formulando preguntas y acabar interactuando con el colectivo social para acabar eligiendo las informaciones que le proporcionan los informadores cualificados. Un camino **cuyo carácter elíptico conforma la integridad de los instrumentos de recogida de datos** y en el que cada uno de los trayectos que lo constituyen resulta necesario, al mismo tiempo que inevitable, y que hay que recorrer una y

otra vez. Un trayecto que pone en evidencia cuál es el territorio de la evaluación: una senda operativamente cerrada que deberíamos ser capaces de recorrer con agilidad, en cualquier sentido y en la que podemos aprovechar cualquier estrategia metodológica. Una senda bien definida que nos obliga a reconocer la integridad topológica de un espacio plural.

De hecho, está es la clave que los profesionales deberían captar para relacionarse de una forma efectiva con las evaluaciones de los programas. Las técnicas de recogida de datos (y las metodologías concurrentes) no son tecnologías independientes y aisladas, sino que conforman un procedimiento a la vez integral y a la vez plural llamado evaluación. Cuando se comprende este hecho, el grado de dominio de cada una de las técnicas resulta secundario, aunque obviamente y desde una perspectiva formal, cuantas más técnicas domine dicho profesional mucho mejor. Pero si no es capaz de interiorizar esta clave, aunque sea el mejor experto en registros del mundo, no va a ningún sitio.

Lo que no puede ser nunca el evaluador es un experto inmóvil, anclado en un único conocimiento especializado que le impide percibir la utilidad de otros métodos. Da lo mismo el tipo de especialización, la técnica de referencia que se utiliza como exclusiva, porque todas “se cierran” en una militancia (y en un conjunto de referencias conceptuales y argumentos sistémicos), que bloquean las prácticas evaluativas. Porque no hay ninguna técnica que no sea deficiente y, por tanto, ninguna técnica por si misma ofrece a la evaluación un grado de certidumbre total.

Veamos a continuación cada una de estas técnicas.

## **BIBLIOGRAFÍA Y DOCUMENTACIÓN COMPLEMENTARIA.**

Vamos a realizar una recomendación general para todos los capítulos de esta segunda parte. No por falta de posibles referencias, sino más bien por todo lo contrario, ya que existen muchos manuales de métodos y técnicas de investigación social. Cada departamento de métodos y técnicas de cada facultad ha publicado al menos un par de ellos en los últimos años. Resulta difícil recomendar alguno en concreto. Por este motivo, aparte de las referencias que aparecen en los próximos capítulos (que se recogen en la bibliografía general) los textos de referencia podrían ser:

En inglés un reciente diccionario en 2 volúmenes (nivel alto):

LEWIS-BECK, M.S.; BRYMAN, A. y FUTING LIAO, T. (2004), **The SAGE Encyclopedia of Social Science Research Methods**, Londres, SAGE.

En español la mejor documentación complementaria la constituye la colección de Cuadernos Metodológicos del CIS (nivel intermedio, aunque quizá alguno es algo complejo), a la que puede accederse a través de [www.cis.es](http://www.cis.es), publicaciones, catálogo, colecciones y cuadernos metodológicos.

# 5

## La mirada del observador

### 5.1.- La observación científica.

El primer momento de la investigación social, el acto empírico primordial, supone **dirigir la mirada para observar** aquello que está ocurriendo para retenerlo y registrarlo como una experiencia directa de los acontecimientos.

Observar es un proceso humano esencial con unas profundas raíces biológicas, de tal forma que observar es, en primer lugar, percibir las cosas de una manera analógica, intuitiva y reactiva, el despliegue de la vida, la reacción de la flor que se cierra ante los rayos del sol, la de la gacela que huye cuando percibe al leopardo y la de ser humano asustado por el ruido del trueno. En segundo lugar, aparece la observación interactiva, presente en la acción social natural, en la expresión de creencias y preferencias, en el uso de los lenguajes simbólicos y especialmente en la interpretación de los signos que denotan la presencia de estructuras sociales. Los seres humanos para vivir en sociedad necesitan observar, calificar lo observado y actuar según les oriente dicha observación.

Pero hay una tercera manera de observar. Aquella que utiliza las reglas y procedimientos científicos persiguiendo, con el método de alguna ciencia particular, responder a las preguntas que el objeto de dicha ciencia le plantea. La vigilia vital, la observación natural y la observación científica responden a la misma capacidad humana, pero a la vez son profundamente diferentes.

En especial, hay que distinguir muy bien entre el segundo y el tercer tipo de observación, que se diferencian tanto por el método como por el objetivo. Ciertamente ambas formas de observar pueden coincidir en la profundidad y pertinencia de sus observaciones, ambas pueden utilizar métodos y técnicas similares, pero mientras para la una (la observación natural) esto es sólo una posibilidad, para la otra (la observación científica) es una necesidad que debe aplicar sistemáticamente. La observación científica supone, por tanto, el conocimiento y el aprendizaje de unas técnicas que se deben aplicar de forma clara, transparente y sin excepciones, lo que implica que la observación científica contenga exigencias, las cuales la diferencian de la observación natural. Sin embargo, aquellos que no han sido entrenados en la observación científica pueden confundir ambas formas de observación y cuando las reglas, como ocurre en el caso de las ciencias sociales, se manejan como actitudes endógenas del observador, la confusión es fácil y el resultado peligroso. Al ser la observación el momento fundacional de la investigación, la confusión entre ambas maneras de observar invalida todos los procedimientos subsiguientes.

Dicho aprendizaje incluye, entre otras exigencias:

- Especificar el lenguaje de la observación, que no puede ser un lenguaje neutro, pero sí verdadero,
- Identificar con claridad aquello que tenemos que observar,
- Determinar la forma de hacerlo, tanto en el plano teórico de las categorías y conceptos como en el plano técnico de los métodos, las medidas, los códigos o los sistemas de registro y
- Poder evaluar la calidad de las observaciones, su validez, fiabilidad y las posibles fuentes de error.

Aunque en las Ciencias Sociales una parte sustancial de los estudios más relevantes para la conformación del área de conocimiento han recurrido a la observación directa, de manera progresiva es una técnica que se utiliza cada vez menos, en parte porque ha sido una técnica que se ha aplicado de una forma poco rigurosa y no se han extraído datos verdaderamente interesantes. Al crearse el modelo de la evaluación fuerte, la observación ya parecía una cosa del pasado y no se ha utilizado demasiado (o casi nunca) como técnica de recogida de datos.

Porque la observación es una técnica muy exigente, quizás la más exigente de todas las técnicas de investigación. Requiere un gran entrenamiento y mucha práctica. La creciente formalización de los proyectos de investigación ha hecho que muchos investigadores se hayan visto tentados a abandonar la observación por otros métodos de recolección de datos más rápidos y más fáciles. La observación es una actividad artesanal que comparada con las otras técnicas de recogida de datos exige mucho más tiempo de dedicación con una productividad aparentemente inferior a la que ofrecen las técnicas de producción más estandarizadas.

Sin embargo, tales exigencias se traducen, en un plano logístico, en un diseño especialmente adecuado para la evaluación, ya que la observación:

1. Tiene bien formulado un objetivo de investigación concreto.
2. Debe plantearse y registrarse de forma sistemática.
3. Se somete a comprobaciones y controles sobre su validez y fiabilidad (Sellitz, 1964).
4. Requiere definir un escenario en el que ocurrirán los acontecimientos e identificar a los personajes que vayan a intervenir en los mismos (Taylor y Bogdan, 1984).

Estas cuatro exigencias se completan con la idea de que la observación ni es patrimonio de ningún método ni de ninguna teoría, ni tampoco es un "proceso natural", como de alguna manera pretende en el territorio de la evaluación Egon Guba. Se trata de algo más complejo, que implica valores, conocimientos previos, un cierto grado de control y la "confección" de un cuerpo ideográfico de "observaciones" (Ruiz de Olabuénaga, 1996). La observación es por tanto una técnica compleja, que requiere formación, entrenamiento y conocimientos teóricos adecuados. También es cierto que cuanto más "natural" sea la situación observada, al menos en determinados temas, más productiva va a resultar, pero no debemos confundir "*observación natural*" con la "*observación científica de situaciones naturales*".

Esta distinción es muy importante en el terreno de las políticas y los programas de juventud, donde se produce un cierto rechazo a lo teórico y se trata, en ocasiones, de "dar la palabra" a los jóvenes, lo cual constituye una estrategia participativa muy coherente, pero que a la vez supone refugiarse en el "naturalismo" extremo de la realidad cotidiana, sin otros elementos de contraste. Si este joven "observador" no se aplica "*las conclusiones de lo observado a sí mismo*" (Luhmann, 1990), algo poco probable ya que no posee un lugar alternativo, conceptual y teórico desde donde hacerlo, la observación natural será sólo información sin elaborar sobre esta vida cotidiana, lo que exige un segundo observador que realice la "observación científica".

## 5.2.- Procedimientos para la observación.

La observación se refiere a las apariencias y, por tanto, su percepción es directa por parte del observador. Pero a la vez, el registro de las observaciones requiere la presencia de soportes que, en general, implican el uso del lenguaje, bien porque lo observado es una conversación, o bien porque los hechos o acontecimientos se registran en forma de palabras. Esto significa que no se analiza nunca el hecho en sí mismo, sino las palabras que describen el acontecimiento. Esta limitación comienza a ser superada con las nuevas tecnologías de la imagen, que permiten introducir formas de registro utilizando sistemas muy discretos de video que no se interfieren en la acción y que posibilitan un análisis directo del gesto, la posición, las distancias y, en general, el lenguaje no verbal.

Debemos también tener en cuenta que muchos programas de juventud, especialmente los relacionados con el ocio y el deporte, incluyen por sí mismos procedimientos habituales de registro, porque el registro de la actividad forma parte del mismo acontecimiento, lo cual permite al evaluador realizar directamente la observación, sin interferir en el comportamiento ordinario.

Este tipo de observación, en la que el observador no se ve obligado a participar activamente en la situación y a registrarla personalmente, permite centrar la atención exclusivamente en la situación a observar y ordenar sistemáticamente la totalidad del fenómeno y el registro de los resultados. Pero es una observación limitada a conductas que son posteriormente interpretadas, lo que tiene sus ventajas y sus inconvenientes. La gran ventaja es la “naturalidad” de lo observado, que además transcurre con su propia lógica; la gran desventaja que desconocemos es el significado que tiene para el sujeto y no podemos inducir a la aparición de determinadas secuencias de comportamiento.

Cuando la situación no es pública debe ser planificada y autorizada por los observados, lo que implica que “*aunque uno se siente al fondo se percibirá su presencia*” (Woods, 1986), lo cual supone que el observador introduce algún tipo de rol en el campo de observación, con lo que aparece el peligro de modificar el suceso observado.

Dicho riesgo induce, a una mayoría de autores, a dar por supuesto que la observación no participante sólo puede llevarse a cabo si se sabe que la presencia del observador no va a modificar sensiblemente el proceso social o si las posibles modificaciones no son controlables, lo que en realidad sólo se puede conseguir manteniendo la presencia del observador en secreto y fuera del campo de observación, lo que nos enfrenta a serios problemas éticos. En realidad, la solución aparece por otro lado, ya que la influencia del observador sobre la realidad observada es algo tangencial, medible y en todo caso equivalente a aquello que ha sido definido en las Ciencias de la Naturaleza como el principio de indeterminación, el cual no ocasiona tampoco demasiados quebraderos de cabeza a este tipo de científicos. Si en los laboratorios la incidencia de la presencia del observador se “estima” sin demasiados prejuicios, lo mismo podemos hacer en la observación de situaciones naturales. El único problema real para la observación se refiere a que el observador no debe interferir conscientemente en el proceso de observación de los acontecimientos naturales para obtener aquellas informaciones que cree que la observación no le está proporcionando.

Las diversas técnicas e instrumentos diseñados para realizar observación sistemática suelen medir, por lo general, la forma, duración, frecuencia, antecedentes y consecuencias de comportamientos individuales y de estructuras sociales. En determinadas ocasiones también recogen información sobre objetos y artefactos presentes, así como sobre los usos que se les da a los mismos (Goetz y LeCompte, 1988). Además, en la investigación de tipo sistemático el investigador sabe qué aspectos de la actividad son relevantes para su objetivo de investigación y, por tanto, tiene la posibilidad de desarrollar un protocolo de las observaciones antes de comenzar la recogida de datos.

La observación sistemática supone que los sucesos que se observan son seleccionados, anotados y codificados en unidades significativas para que pueda analizarlos alguien que no está

presente en la observación. Como es físicamente imposible registrar absolutamente todo lo que está ocurriendo en un campo de observación, ni tan siquiera utilizando medios audiovisuales, se debe tomar una decisión relativa a los sucesos a observar, así como al período de tiempo durante el cual se realizarán las observaciones, lo que implica algún tipo de muestreo sobre los acontecimientos para garantizar su posterior generalización.

Una vez seleccionados los hechos que se van a observar hay que proceder al registro de los mismos, tomando notas, haciendo fotografías, con vídeo o marcando las frecuencias en una hoja precodificada. Decidir qué procedimiento de registro se va a utilizar depende de las hipótesis establecidas y de los recursos disponibles.

### 5.3.- La actitud del observador durante la observación y la cuestión de la observación participante.

Hay un acuerdo casi unánime en que una de las decisiones más importantes que el investigador tiene que tomar antes de iniciar un trabajo de observación consiste en definir la manera en la que va a relacionarse con las personas objeto de la observación. Tradicionalmente se ha concebido el papel de observador como un continuo que va desde el ideal del observador puro, es decir aquella situación, como por ejemplo un acto público, en la que las personas observadas ni conocen ni saben que están siendo sometidas a este tipo de observación; hasta el ideal del participante completo, es decir, aquella situación en la que las personas observadas interactúan con el observador, pero no saben nada sobre sus intenciones (Babbie, 1996).

En cualquier investigación es necesario especificar las condiciones de la observación, porque la condición pública del observador influye en el comportamiento de las personas observadas de manera diferente según el papel que adopte. Esto ha conducido a algunos observadores a ocultar su trabajo, para evitar que la conducta de los sujetos pueda resultar influida por el conocimiento de saberse observados. Pero la investigación encubierta no está autorizada por los códigos éticos.

Pero éste no es un problema para la **observación participante**, la cual puede definirse como *“un conjunto de métodos de observación donde el investigador forma parte de los sucesos que se observan. El observador está en una relación “frente a frente” con los observados, recogiendo datos al mismo tiempo que participa con ellos dentro del marco de su vida natural. La participación directa del observador en la situación le permite obtener una imagen más detallada y completa de la situación observada, pero, a su vez, tiene el inconveniente de que al estar inmerso en la situación que observa, está expuesto a la pérdida de objetividad”* (Cicourel, 1982).

El concepto de observación participante procede del ámbito de la antropología y describe el tipo de relación que viven los etnógrafos con la población que están estudiando.

A Malinowski, en sus estudios sobre los habitantes de las islas Trobriand, le corresponde el mérito de haber dotado a la observación participante del rigor científico y eficacia, que la han convertido en una herramienta mítica y cargada de romanticismo, ya que *“cuando el corazón del observador haya comenzado a latir al mismo ritmo que el de las personas a quienes está observando y no como latiría el del enviado de un lejano laboratorio, entonces habrá conquistado el título de observando participante”* (Madge, 1953). Una metodología que todos los científicos sociales desearían haber utilizado en alguna ocasión, aunque pocos lo han hecho. Pero el exceso de deseo impide captar el verdadero mensaje de la observación participante: no debe importar tanto que los sujetos se sientan observados, como el carácter que atribuyen a esta observación (Shwartz y Jacobs, 1984).

La ventaja de la observación participante reside en situar tanto al observador como a la persona observada en el mismo bando, convirtiendo al observador en un miembro del grupo para que pueda sentir lo que ellos sienten y pueda trabajar dentro del marco de referencia del propio grupo, para lo que tiene que renunciar a un rol de entrevistador con intenciones de recabar cier-

tas informaciones pertinentes. La observación funciona no tanto porque el observador esté más o menos identificado con el grupo, sino porque adopta una posición pasiva sin interrogar. **Luego la estrategia empírica de la observación, de cualquier tipo de observación, se resuelve en el silencio del investigador.**

El propio sistema tradicional de registro de la observación participante confirma este aserto. El “diario de campo”, las “notas de campo” o el cuaderno donde se registran las anotaciones del día, contienen anotaciones de comportamientos visibles, como los participantes (edad, género, lazos,...), el ambiente, los motivos explícitos que orientan la conducta, los componentes de la misma, su frecuencia y duración, pero también se anotan interpretaciones de lo que se ha observado. Si bien el objetivo de las notas de campo es que no se pierda la información observada y que ésta se encuentre en todo momento accesible a nuevos análisis e interpretaciones, lo que en absoluto significa que sirva para orientar la acción del observador, porque si pregunta para conseguir respuestas a las dudas anotadas la observación dejará de ser válida (García Jorba, 2000).

Así, aunque la observación participante implica la interacción social, por la que el observador entra en contacto con los actores sociales, con los protagonistas de la situación objeto de interés, la técnica de recogida de información garantiza el control de que los datos recogidos han sido producidos **sólo** por los observados. Después vendrá la tarea científica de sistematizar, tipificar y analizar la información recogida con la finalidad de poder captar el significado de la situación observada, para interpretar una situación, elevando a categoría teórica la información almacenada a través de la observación, pero la observación en sí misma, **la auténtica observación no implica ningún tipo de dinámica interpretativa.** Quizás, por este motivo, es tan importante la elaboración de un informe de campo que, al margen del registro de la observación, describe, para incorporar al análisis, las incidencias de la misma (Taylor y Bogdan, 1984).

#### 5.4.- La observación y la evaluación.

En el territorio de la evaluación de programas e intervenciones sociales la observación es una técnica muy poco utilizada, aunque, de forma paradójica, ocurre que la observación es una técnica muy desarrollada para evaluar en los ámbitos educativos y también en la evaluación de otras conductas individuales. También es una técnica muy citada en los manuales y en los programas académicos, que siempre indican que es muy adecuada a los fines de la evaluación. Especialmente si se habla de observación participante. Pero a la vez no hay ninguna práctica de evaluación, al menos ninguna publicada y difundida en España, que haya utilizado, de forma consistente, esta técnica de recogida de datos.

Esto ocurre porque la observación (y en particular la observación participante) se tiene que enfrentar a un dilema delicado y un tanto complejo si quiere ser utilizada en la evaluación de proyectos y programas. Ocurre que los profesionales que desarrollan esta intervención están, de forma continua, cotidiana y permanente, realizando labores de observación y reciben por tanto un gran volumen de información (de datos) sobre la realidad de dicha intervención, así como sus resultados y las reacciones que suscita. Para cualquier evaluador externo alcanzar este grado de información requiere una gran dedicación, ya que de alguna manera el conjunto de datos obtenidos por un evaluador externo tienen que alcanzar un nivel de información equivalente, al menos, al obtenido por los profesionales que realizan la intervención. Alcanzar este objetivo requiere su tiempo y además requiere hacerlo en los tiempos (y los lugares) donde se desarrolla la intervención. Como consecuencia se suele considerar que otros métodos, como las entrevistas a informantes clave, los grupos de discusión y la aplicación del método Delphi a los profesionales que realizan la intervención, resultan más eficientes que la observación participante.

A la vez, evaluar mediante observación requiere acceder a lo que Clifford Geertz llama “una descripción densa”, lo cual solo puede lograrse, como se ha explicado, combinando una fuerte

sistematización, realizando los oportunos registros, pero también adoptando una mirada externa y seleccionando aquellos acontecimientos verdaderamente relevantes. Pero todo esto también requiere su tiempo y especialmente una preparación específica que suele escasear. Es decir, los profesionales de la observación de los acontecimientos sociales deben poseer unas habilidades específicas, como una gran capacidad de empatía, saber diferenciar en lo concreto los rasgos étnicos de las proyecciones étnicas, así como una cierta capacidad para ordenar, sistematizar y jerarquizar las propias informaciones, algo que sólo manejan unos pocos antropólogos, algún psicólogo social o pedagogo muy comunitario, algunos sociólogos con mucha experiencia cualitativa y quizá algún educador social. El resultado final es que estamos ante una técnica con pocos profesionales competentes, cuya aplicación requiere además mucho tiempo y dedicación.

Como consecuencia tales exigencias suponen un coste elevado y una planificación muy compleja de la intervención y la evaluación, para conseguir, al final, un tipo de información que no goza de demasiado reconocimiento institucional.

Pero por otra parte ocurre que alguno de los grandes hitos de la investigación social, tanto en el logro de evidencias como en el desarrollo teórico y conceptual, son el resultado de la aplicación de la metodología de la observación participante, lo que nos permite suponer que para la evaluación el uso de esta técnica podría también producir avances importantes.

Por este motivo parece necesario seguir apostando por la validez de dicha técnica y buscar las maneras de realizar evaluaciones utilizando la metodología de la observación participante. Quizá lo más sensato, en un primer momento, sería utilizar esta metodología para realizar análisis de la realidad (siguiendo el modelo de la llamada monografía de campo), así como evaluación de diagnóstico y diseño, pero también para evaluar la implementación y los procesos. Es decir, podemos tratar de utilizar, en un primer momento, la observación para evaluar el antes y el durante, para, cuando tengamos más experiencia, utilizarla en el después.

# 6

## Los datos producidos por otras observaciones o investigaciones

### 6.1.- Una maraña organizada.

Tanto la observación natural como la observación científica producen datos, pero siendo, como he dicho, dos tipos de datos muy distintos, ambos tienen en común la posibilidad de que otro observador los descodifique para producir una observación de segundo grado. Este procedimiento empírico es común, tanto si se trata de datos producidos por la observación natural directa (una carta al novio, la descripción de unos acontecimientos en un diario o el testimonio ante un juez), institucional (las noticias y los comentarios aparecidos en los medios de comunicación o las respuestas a las preguntas en los exámenes) o bien por la observación científica (las estadísticas del INE o la reciente eclosión de ensayos fallidos de los físicos sobre el problema del tiempo). Obviamente el análisis de cada una de estas fuentes de observación será distinto, pero el procedimiento empírico que nos conduce a este tipo de análisis secundario será siempre el mismo.

Toda comunidad produce una gran cantidad de informaciones que guardan relación con el comportamiento de sus miembros. Aunque estos datos han sido producidos con diversos fines, tienen un considerable interés para la evaluación. Olvidarnos de su existencia supone desestimar una información que es muy relevante o, si dedicamos la evaluación a la recogida de datos equivalentes a los ya existentes, una considerable pérdida de tiempo y otros recursos. En este sentido, los documentos empleados en las investigaciones casi siempre incluyen información sobre trabajos previos realizados por otros investigadores en el campo de estudio en cuestión.

Pero la utilización de estas fuentes implica una serie de riesgos, primero, porque aunque aparezcan de una forma perfectamente organizada y clasificada, en realidad están muy enmarañados; segundo, porque en ocasiones no conocemos el significado original de los mismos, por ejemplo, si son o no representativos incluso cuando afirman serlo (Royo, 1997); tercero, porque la intención con la que fueron registrados es un tanto particular (Ander-Egg, 1989); y en cuarto lugar, pueden existir errores o incluso falsificaciones de las que no somos conscientes o no estamos informados (González, 1996).

Todas estas posibilidades nos obligan a adoptar algunas precauciones para manejar las fuentes de datos secundarias:

- Que la metodología y los conceptos utilizados para realizar la recopilación original estén descritos de forma amplia, clara y precisa.
- Que se trate de una fuente garantizada, reconocida y creíble.
- Que el procedimiento de recogida haya sido bien definido y tengamos acceso a los instrumentos utilizados en el mismo.

En todo caso, hay que diferenciar entre la utilización de los datos secundarios ya elaborados y la construcción de un “sistema de indicadores” propio, tema al que dedicaremos el capítulo 16.

En el primer caso nos limitamos a dirigir la mirada sobre datos ya existentes, en el segundo, en cambio, se trata de construir el sistema de recolección vía censal o muestral y, por lo tanto, pertenece a otro momento empírico distinto. En general, el que tiene que ver con la encuesta. La habitual confusión entre la “utilización de datos secundarios” y la “construcción de sistemas de indicadores” es esencialmente técnica y metodológica y supone no diferenciar bien entre lo estático y lo dinámico, entre lo puntual y lo continuo.

Centrándonos en este capítulo en los datos ya elaborados, el principio básico para el uso de datos secundarios se refiere a la necesidad de establecer una adecuada correspondencia entre las preguntas que planteamos en la matriz de la evaluación y las respuestas que ya nos ofrecen los datos elaborados. Lo ideal es que una idea o una hipótesis pueda ser planteada de forma tal que la información disponible contenga la respuesta exacta a la pregunta, tratando de evitar que la respuesta “sólo se parezca o pueda ser asimilada” a la pregunta. Es, por tanto, una técnica en la que el investigador también permanece silencioso, porque tiene que aceptar que sólo puede hacer las preguntas que se formularon al recoger los datos originales. Claro que esto, como le ocurrió a Durkheim con “*El suicidio*”, no impide obtener excelentes resultados.

En cualquier caso, la inadecuación entre la pregunta de la matriz y la respuesta de la fuente de datos secundaria no es fácil y la ambigüedad es frecuente. Por ejemplo, la serie de encuestas de juventud nos muestra la creciente insatisfacción de los jóvenes con su capacidad de gasto para el ocio. Pero este dato puede interpretarse de dos maneras distintas: porque tienen menos dinero o porque las exigencias del mercado del ocio son crecientes. Obviamente, tales ambigüedades pueden eliminarse en una gran medida triangulando los datos (ver capítulo 12) e introduciendo otras variables, por ejemplo la evolución del nivel de gasto o el volumen de consumo de servicios de ocio por parte del público juvenil.

Por otra parte, la utilización de la información de fuentes secundarias implica una serie de ventajas:

- Se trata de un procedimiento muy económico que permite trabajar con presupuestos mínimos.
- proyecta las garantías de calidad que ofrece la fuente original.
- Muchos datos secundarios forman largas series estadísticas, lo que permite visualizar un largo periodo de tiempo, realizar las oportunas comparaciones y reemprender análisis en diferentes momentos.
- Se evitan todos los problemas prácticos del trabajo de campo.
- Desde la irrupción de las nuevas tecnologías de la información y la comunicación las fuentes son cada vez más accesibles. Aunque esto produce un efecto paradójico, por ejemplo, desde que el nuevo Padrón Continuo resulta directamente accesible por Internet ya no tiene justificación que los proyectos, planes y programas de juventud no utilicen, mes a mes, los datos demográficos de los jóvenes edad por edad. Sin embargo, son muchos los equipos que dicen ignorar la pirámide de edades actual de los jóvenes de su localidad.

## 6.2.- Algunos tipos de fuentes secundarias.

### 6.2.1.- Las estadísticas oficiales.

La fuente de datos secundaria más importante para cualquier evaluador la constituyen las estadísticas oficiales. Se trata de la fuente de datos más fiable, de mayor calidad y más transparente, ya que es la más observada y criticada, de tal forma que ante cualquier estadística oficial siempre podemos acceder a una descripción de sus límites metodológicos para saber hasta que punto se adecuan a los contenidos de las preguntas que formulamos en la matriz.

Sin embargo, la cultura popular (y la prensa) mantiene una actitud de rechazo hacia las estadísticas oficiales, aunque luego todo el mundo las utiliza. Existe una cierta “leyenda urbana” en torno a las estadísticas, condensada en una famosa frase que unos atribuyen a Benjamin Disraelei y otros a Mark Twain (últimamente también a Goethe) que afirma: *“Hay tres clases de mentiras: mentiras, condenadas mentiras y estadísticas”*, pero que en realidad fue formulada en 1985 por Leopard H. Courtney, un correoso (e ilustrado) político inglés que llegó a ser presidente de la Royal Statistical Society, y que la pronunció en un debate sobre “información privilegiada y corrupción”. En dicho debate Courtney no se refería a las “estadísticas”, sino a la compra de terrenos colindantes al trayecto de una nueva vía de ferrocarril por parte de aquellos que conocían de antemano el proyecto. Por una “extraña” coincidencia esta frase se ha hecho muy famosa en aquellos países y ámbitos muy tocados por la corrupción social y política. En todo caso, muestra el temor que suscita la evaluación y el uso de estadísticas fiables en algunos sujetos “cargados de opinión”.

Ciertamente, sabemos que las estadísticas oficiales son en ocasiones poco precisas en la definición de los términos, que no desagregan suficientemente los datos, que cambian las definiciones sin avisar y que en algunos casos los métodos y técnicas para recoger la información dejan mucho que desear. Pero por saber justamente todo esto podemos utilizarlas de una forma mucho más razonable que aquellas fuentes de datos que parecen más adecuadas porque no sabemos nada sobre ellas. Quizás hacemos esto porque las fuentes estadísticas son muy accesibles.

Se supone que el problema crónico de las estadísticas oficiales en España reside en la dispersión y el retraso de algunas. En todo caso, hay una tendencia clara hacia la actualización y muchas de ellas están hoy al día. A la vez se han concentrado en algunos organismos, como el INE (y los correspondientes en las Comunidades Autónomas e incluso algunos Municipios importantes) y en el caso de juventud en el INJUVE. En los últimos diez años, gracias a la informática y las nuevas tecnologías de la comunicación y la información, el panorama ha cambiado totalmente. En realidad el problema actual reside en ser un usuario lo suficientemente competente de las nuevas tecnologías y al mismo tiempo poseer una visión adecuada de los significados y limitaciones de los datos proporcionados. En todo caso, el evaluador debe saber moverse con una cierta seguridad en toda esta maraña.

Retomaremos, desde otra perspectiva, alguna de estas cuestiones en el capítulo 8.

### 6.2.2.- Los documentos oficiales.

Los documentos oficiales, tanto aquellos que reflejan la estructura, funciones, fines, valores de una organización, como aquellos que son producidos por una institución como parte de una política de comunicación con el exterior (entre los que podemos citar memorias, revistas, folletos y boletines), son fuentes de datos muy adecuadas para conocer la perspectiva oficial del organismo o institución.

En el caso de las políticas de juventud, este tipo de documentos son muy frecuentes (Comas, 2007) y reflejan una cierta preocupación por difundir “lo que se hace o lo que se va a hacer”. En este sentido recopilarlos y analizarlos supone una buena práctica evaluativa.

### 6.2.3.- Los documentos personales.

Los documentos personales son *"autobiografías, cartas, diarios, relatos, ciertos tipos de redacciones escolares, relatos premiados etc., en general documentos en forma tangible -escritos o grabados- que han sido escritos por la propia iniciativa del escritor o, en otro caso, en forma tal que su contenido introspectivo haya sido determinado enteramente por el autor, y documentos que se refieren a la experiencia personal del autor"* (Selltiz y Wrightsman 1975). Es decir, todo aquello que muestre cualquier experiencia personal, bien en primera o en tercera persona, que tienen valor para la evaluación, en la medida en que las actitudes, opiniones, conductas reflejadas contengan información pertinente para los objetivos de la misma.

Un ejemplo histórico muy importante de esta técnica de recogida de datos la ofrece la conocida obra de Isaiah Thomas y Florian Znaniecki, *The Polish Peasant in Europe and America* (1918), que constituye una lectura muy recomendable para cualquier evaluador, especialmente después de una reciente traducción (Thomas y Znaniecki, 2004).

Sin duda, en cualquier programa se producen abundantes documentos personales que representan una fuente extraordinaria de hipótesis e ideas que aguardan que alguien los utilice, aunque en general, aquellos que los producen no suelen tener una buena disposición a facilitarlos al menos hasta pasado un tiempo, lo cual les invalida como fuente de datos para una evaluación.

### 6.2.4.- Los contenidos de los Medios de Comunicación.

Por su parte, los Medios de Comunicación de Masas (MCM), es decir, los periódicos, películas, emisiones de radio y televisión, revistas y en los últimos años las páginas Web de Internet, que se pueden utilizar bien como una fuente de documentación general, bien como una fuente de documentación sobre ciertos acontecimientos, o bien como una fuente de información sobre ellos mismos.

La relación entre la investigación social (que puede extrapolarse a la evaluación) y los MCM se ha sostenido sobre una clara dualidad. Por una parte ha existido una tradición crítica o "apocalíptica", en los términos de Umberto Eco, en la que aparecerían figuras como Alexis de Tocqueville, Ortega y Gasset, Dwight MacDonald, T.S. Eliot o Edward Shills, la cual considera a los Medios de Comunicación de Masas como una amenaza para el conocimiento y el avance de la ciencia. Pero, por otra parte, el análisis de contenido de la comunicación, desde la perspectiva de los "integrados", representa uno de los capítulos más brillantes de la historia de las ciencias sociales, que se inicia con Lasswell en 1915 y alcanza su cenit con Berelson en la década de los 50. Aunque, ciertamente, desde entonces el llamado "análisis de contenido de los MCM" no ha seguido avanzando (Bardin, 1977).

De alguna manera, la versión apocalíptica puede ser cierta si ubicamos a los MCM en el lugar de una realidad observable, en vez de situarlos en la categoría de productores de datos. Pero, a la vez, como fuente de documentación general resulta muy útil para poder caracterizar acontecimientos concretos, tanto desde la perspectiva de los hechos conocidos como desde las actitudes públicas manifestadas con relación a los mismos. En el ámbito de las políticas de juventud, que casi siempre suelen tener una proyección mediática, parece posible utilizar, aunque sea a modo de contraste, esta fuente de datos.

### 6.2.5.- Internet.

En quinto lugar, en los últimos años se ha ido configurando como una fuente de primer orden Internet y en particular los buscadores y las bases de datos específicas. Más recientemente se han incorporado a esta posible fuente de datos secundarios las imágenes institucionales y aquellas que se obtienen con el vídeo doméstico. Se trata de un conjunto de informaciones de diferente valor, que hay que utilizar con cautela, aunque a la vez todas las fuentes de información fiables han acabado por adoptar un formato electrónico y formar parte de Internet.

En evaluación no existe mucha experiencia del uso de Internet, salvo en lo relativo a la evaluación de aquellos programas que utilizan o tienen que ver con las nuevas tecnologías de la comunicación y la información. Estas primeras experiencias nos indican la aparición de numerosas facilidades, pero también algunos riesgos relacionados con la fiabilidad de los datos. En cualquier caso, un evaluador bien formado podrá discriminar con facilidad el origen de las fuentes y determinar su grado de fiabilidad. En todo caso, parece necesario realizar algún desarrollo metodológico de esta cuestión y aunque han sido los ámbitos de juventud los que han realizado las primeras aportaciones significativas (Ver la Revista de Estudios de Juventud), no podemos citar un texto de referencia útil que explique las posibilidades de Internet (y en general de las TIC) en la evaluación de programas.

#### **6.2.6.- La bibliografía especializada.**

Finalmente, disponemos de las propias fuentes de datos obtenidos mediante la investigación social y la evaluación, en las cuales se recogen desde los Estudios de Juventud de carácter general hasta los estudios monográficos. En España y en los ámbitos de Juventud, la fuente de datos más consistente es el INJUVE, pero aparte del mismo disponemos de varias entidades que, con independencia de las universidades, mantienen fondos documentales vivos de investigación social, a los que sería necesario añadir docenas de organismos sectoriales públicos (CIS, INCE, Instituto de la Mujer, Seguridad Social...) o privados, en especial fundaciones que se limitan a recoger un tipo especializado de investigación. Conviene indicar que numerosas fuentes (de las cuales el INJUVE es buen ejemplo) han situado sus datos sobre páginas Web, lo que hace esta información mucho más accesible.

Volveremos a este tema en el capítulo 8.

#### **6.3.- El uso de fuentes secundarias en evaluación.**

En las evaluaciones de programas y acciones desarrollados por un ámbito local no es fácil encontrar evaluaciones que utilicen fuentes secundarias, en cambio, en las grandes evaluaciones de ámbito nacional (y en algunas autonomías) los programas relacionados con políticas generales de carácter sectorial, pueden evaluarse más fácilmente a partir de indicadores procedentes de fuentes estadísticas normalizadas.

El ejemplo más amplio de la utilización de las fuentes de datos secundarias en evaluación es el de Educación, donde la mayor parte de los indicadores proceden de las estadísticas educativas generales. Para alcanzar esta situación se han adaptado, a la vez, los estándares de la evaluación educativa, pero también las estadísticas. El proceso seguido para realizar este proceso por parte de educación (en España puede consultarse a través del Instituto de Evaluación del MEC, <http://www.institutodeevaluacion.mec.es/>) es un modelo que podría extrapolarse a otros ámbitos.

Sin embargo, fuera del campo educativo, las experiencias que utilizan datos secundarios en evaluaciones son escasas. La razón es clara, el grado de segregación (de detalle) de las estadísticas ha sido, hasta ahora, bastante bajo y además la duración (el tiempo de implantación) de los programas en el ámbito local relativamente corto. En muchos casos se requería una evaluación rápida para tomar decisiones en los siguientes presupuestos y las estadísticas suelen tener una periodicidad anual y se publican, en ocasiones, con mucho retraso. El acceso tampoco era fácil, aunque, como hemos visto, todo esto está cambiando. No obstante, la evaluación de los sistemas y las reformas educativas se refiere a procesos a largo plazo en los que resulta no sólo fácil, sino aconsejable utilizar modelos de indicadores basados en las estadísticas generales. De hecho, estas estadísticas se diseñan teniendo en cuenta los indicadores adecuados para la evaluación.

Por otra parte, como veremos en el capítulo 16, son muchos los indicadores para la evaluación de programas y planes de juventud que tienen una plasmación estadística a nivel nacional o au-

tonómico. Sin embargo, salvo quizá los economistas que, en este sentido, suelen tener una mentalidad más pragmática, el resto de profesionales trata de evitar el uso de tales fuentes de datos. En una gran medida porque “no se fian” (aunque esta desconfianza suele tener un cierto matiz ideológico) y en parte porque prefieren desarrollar sus propios instrumentos (que se supone redundante en prestigio profesional), a lo que dedican un esfuerzo que en ocasiones supone reducir los recursos destinados al análisis. Quizá ha llegado el momento de cambiar de mentalidad.

### 7.1.- El dominio intencional.

La observación directa y la utilización de datos secundarios representan un primer plano empírico caracterizado por la falta de intencionalidad en la recolección de los datos. En ambos casos, aún pudiendo ser datos inferidos, nunca son el resultado de la acción emprendida por el propio evaluador. Cuando el observador interviene para que la fuente produzca no sólo determinado tipo de datos, sino en un determinado orden y dándoles un sentido y un significado, entramos en otro plano empírico que refleja la transición desde las técnicas pasivas a las activas.

El primer nivel en las técnicas activas de recolección de datos lo constituye **la entrevista**, que adopta la forma de una conversación entre el sujeto que produce las observaciones y el que las registra. Sin embargo, contrariamente a lo que ocurre en la simple observación, el observador **dirige** mediante preguntas, y siguiendo un plan preconcebido, aquellas cuestiones o temas sobre las que desea obtener una respuesta del informante, adoptando así un rol dependiente.

La teoría y la práctica de la entrevista se formalizó en los años que transcurren entre 1954 y 1957, período en el que aparecen primero el clásico y canónico texto de Hyman, "**Interviewing in social Research**" (1954), después "**The Focused Interview**", de Merton, Fiske y Kendall en 1956 y finalmente, en 1957, "**The Dynamics of interviewing**", de Khan y Cannell (Cicourel, 1964). Siendo estos últimos los que justamente describieron, por primera vez, la entrevista como "*una conversación con un propósito*".

Se revela así no sólo la naturaleza social originaria de dicha técnica, sino especialmente su componente intencional, ya que hasta ahora, en las técnicas anteriores, el investigador se limitaba a ser un notario, pero a partir de ahora, al investigar actúa sobre el sujeto de la investigación, buscando o provocando producir determinadas reacciones. Por ello, una entrevista se compone de una parte sustantiva formada por preguntas y respuestas y un tipo particular de relación social en la que cada uno de los participantes poseen roles bien definidos, que no se solapan, que uno es el que hace las preguntas, -el entrevistador-, y otro el que responde, -el entrevistado-, (Fowler y Mangione 1990).

Se define así la entrevista por la conformación de dos roles bien diferenciados, en la que el elemento crucial va más allá de la simple interacción, porque uno de los sujetos manipula como a un objeto al otro. La información fluye de forma asimétrica entre ambos, porque el uno se limita a preguntar y el otro a responder. El entrevistador se reserva adoptando un papel jerárquico, no

tiene porque facilitar ninguna información o explicaciones al entrevistado, mientras que éste le proporciona las respuestas que el primero necesita.

Es por tanto obvio que se trata de una forma de intercambio desigual en la que uno, el entrevistado, entrega su punto de vista, al otro, el entrevistador, a cambio de nada. Por ello, la entrevista aparece a partir de un momento histórico determinado, cuando los ciudadanos asumen que pueden dejarse entrevistar, porque sus respuestas, por las que nada reciben a cambio, **aportan algo al bien común**, un tipo de información que permitirá mejorar las cosas. En este sentido, la generalización de las técnicas de recogida de datos mediante entrevista representa uno de los pasos en la conformación de la ciudadanía.

La entrevista es, en definitiva, una conversación, un acto social que implica una interacción entre dos individuos, pero el intercambio que se produce en esa interacción social orientada no sigue las reglas de una conversación cotidiana, sino que obedece a un planteamiento técnico que controla el entrevistador y, por supuesto, aquellos que han diseñado la entrevista.

## 7.2.- Tipos de entrevista.

Pero la entrevista requiere también otros elementos cuya combinación producirá muchos tipos diferentes de entrevistas, cada uno de los cuales será el apropiado para un tipo de evaluación. El primero de estos elementos se refiere al grado de estructuración de la información que se va a recoger y a la intensidad de la directividad que ejerza el entrevistador sobre las respuestas del entrevistado. Tal elemento establece un continuo que va, desde la **entrevista estructurada** y estandarizada en un extremo, hasta la entrevista **no estructurada** en el otro, permaneciendo en una posición intermedia la entrevista **focalizada** o “centrada” (Eckhardt, 1977).

Estos tres tipos de entrevistas, que podemos situar sobre un eje con tres posiciones que, sin embargo, no se corresponde con el grado de directividad del entrevistador. De hecho, he realizado muchas entrevistas no estructuradas pero si muy directivas, porque aunque parezca paradójico, necesitaba ser directivo, ya que estaba persiguiendo un determinado objetivo, pero no quería que el entrevistado se percatara y contestara lo que el suponía que yo “quería oír”.

Esto implica que estamos hablando de dos ejes distintos, cuya relación conformaría diversos tipos de entrevista, lo cual nos conduce a sostener que el tipo de entrevista no se refiere tanto a “modelos alternativos” como a un continuo de posibilidades que hay que aprender a utilizar. Lo mismo cabe decir de los resultados esperables. La entrevista será más o menos exploratoria o descriptiva según la finalidad y los objetivos de la misma, lo que implica otro eje que debemos relacionar con los anteriores. En todo caso, la experiencia común nos dice que las entrevistas no estructuradas son habitualmente menos directivas y se corresponden la mayoría de las veces con diseños exploratorios.

Ciertamente, la existencia de diferentes tipos de entrevista permite ejercer la capacidad de elección técnica, pero ésta no puede ser una elección arbitraria, porque debe darse una correspondencia directa entre el tipo de entrevista a emplear y la finalidad de la información que se va a obtener en el desarrollo de la investigación. El criterio básico para seleccionar el tipo de entrevista es, por tanto, el de optimizar la información. En cada tema concreto hay que establecer una relación entre el tipo de entrevista y la relevancia de la información obtenida. Por ejemplo, si queremos saber en qué ocupa su tiempo una persona, utilizaremos un tipo de entrevista cerrada y directiva; si, en cambio, queremos saber qué significados tienen para esta misma persona los diferentes tiempos en los que transcurre su vida, parece más lógico abrir y ser menos directivo. La elección de una u otra forma de entrevista depende por tanto de nuestros objetivos, pero también de las características de la información que nos interesa y de la información previa disponible. Así, en el ejemplo anterior, si previamente sabemos mucho sobre los significados sociales de los tiempos, podremos optar por una entrevista cerrada.

Con carácter general se puede afirmar que las entrevistas, en las cuales la forma de recoger los datos no está estructurada y la dirección no aparece muy definida, es decir, las **entrevistas abiertas**, son útiles en las investigaciones exploratorias, donde se tiene poco conocimiento del fenómeno en estudio. En una investigación exploratoria se pueden construir tanto las preguntas como las respuestas. Si se buscan sólo respuestas, el tipo de entrevista más útil es la focalizada, mientras que si en lugar de respuestas lo que se busca son las preguntas, lo más conveniente es la entrevista no directiva.

Los primeros investigadores sociales fueron rápidamente conscientes de estas diferencias. Así, en la famosa investigación de Elton Mayo en Hawthorne para la WEC en 1927 (Mayo, 1946), ya se vieron obligados a replantear los tipos de entrevista a realizar, puesto que *“al comienzo de la campaña, los investigadores se atuvieron a un esquema de entrevista que comprendía los temas que parecían estar en relación con la moral: la calefacción, la ventilación, los vestuarios, los riesgos de accidentes, las ventajas financieras y las posibilidades de ascenso. Pero en la práctica, se demostró la dificultad que representaba atenerse a este esquema, ya que las personas entrevistadas no se limitaban al tema; los descubrimientos realizados cuando la entrevista se salía del esquema inicial presentaban más interés y sentido que si se hubiera circunscrito a los temas propuestos. Tras numerosas discusiones, se adoptó una nueva técnica de entrevista, llamada “aproximación indirecta”, más conocida actualmente con el nombre de entrevista “no dirigida”. El individuo entrevistado podía elegir las cuestiones de su preferencia, y el entrevistador le dejaba seguir el hilo de sus pensamientos sin intentar controlar la conversación. Este cambio de técnica tuvo como efecto inmediato el alargar la duración de la entrevista de treinta a noventa minutos, y el protocolo, de dos páginas y media a diez páginas”* (Caplow, 1977).

Asimismo, la entrevista no estandarizada y no estructurada permite recoger información sobre cuestiones a las que no se puede acceder habitualmente por las respuestas a las preguntas de un cuestionario formalizado. Por ejemplo, para conocer cuál es la reacción de una persona ante una determinada situación se puede generar un supuesto de la misma para observar, y anotar, la reacción producida. Otra forma de entrevista abierta es la **entrevista en profundidad**, que se refiere a aquel tipo de entrevista de larga duración, en ocasiones en varias sesiones, con un carácter eminentemente exploratorio, que permite, a la vez, construir las preguntas y buscar las respuestas (Vallés, 2002). La entrevista en profundidad se sitúa en la frontera de otra técnica que vamos a describir más adelante: las historias de vida. Lo cual reitera la continuidad de los métodos y técnicas de investigación.

En el caso de las **entrevistas cerradas** la información se recoge de forma estructurada y estandarizada, lo que implica la existencia de un cuestionario previo y muy definido. La entrevista estructurada y estandarizada, que se utiliza básicamente en las encuestas, puede considerarse como parte de un proceso de medición a partir de las diferentes proporciones de respuestas obtenidas por los entrevistadores, por ello la información obtenida debe limitarse estrictamente a las respuestas a las preguntas que se plantean.

En ambos casos, abierta o cerrada, la realización de la entrevista requiere de un adiestramiento técnico que, de forma habitual sólo reciben los psicólogos (Naohum, 1986), aunque se trata de un adiestramiento que, por mantener la relación asimétrica, no considera prioritario el grado de empatía que requiere la evaluación. En todo caso, la psicología ha desarrollado en numerosos manuales los elementos básicos de la entrevista, como un esquema de desarrollo en fases; técnicas para la contactación y el acceso y el correcto trato social a los entrevistados; un determinado orden en las preguntas; una correcta formulación lingüística y conceptual que las haga inteligibles para los entrevistados; capacidad para identificar el contexto social y psicológico del entrevistado; capacidad para eliminar las barreras en la comunicación (sexuales, generacionales, de clase, culturales e ideológicas); atención a los factores ambientales; una identificación aceptable del entrevistador; formas de registro que no se interfieran en la realización de la entrevista y en general un conjunto de reglas de juego aceptables por ambas partes.

### **7.3.- La evaluación y las entrevistas.**

En el ámbito de la evaluación la entrevista suele formar parte de otras técnicas de las que se citan a continuación, aunque en algunas ocasiones puede utilizarse en sí misma como una técnica de recogida de datos, es decir, podemos proponer realizar un número determinado de entrevistas para obtener una serie de informaciones. En este caso se debe determinar quiénes serán los entrevistados, cuáles han sido los criterios de selección y a qué objetivos responden las entrevistas que se van a realizar. Tales elecciones responden a un determinado perfil de representación que también deberá explicitarse.

También deberá fijarse previamente cuál es el tipo de entrevista que se va a realizar (grado de directividad y tipo de estructura), justificándose en todo caso las decisiones. En función del nivel de apertura deberá prepararse un cuestionario o un guión previo y deberán justificarse las preguntas a realizar.

Salvo en caso de entrevistas totalmente cerradas con un cuestionario codificado (lo cual no es nada frecuente, ya que en este caso pasaríamos al territorio de la encuesta), la entrevista debe ser grabada y hay que tener previamente establecidos criterios de selección temática para orientar el registro de los resultados.

Como he mostrado, la técnica de la entrevista impone una ruptura desde el nivel empírico de la pasividad (limitarse a ver y a oír) hasta el nivel empírico de la intencionalidad (preguntar). Sin embargo, una vez producida tal ruptura podemos comenzar a proponer configuraciones de intencionalidad muy diferentes y relacionadas, por una parte, con las vías de aproximación a la realidad social, aunque no con todas, y por la otra, con la constitución de los elementos que definen el quehacer de la investigación social, a saber: primero, **registrar** los hechos y los acontecimientos (registros); segundo, evidenciar el **devenir socio-histórico** (narraciones e historias de vida); tercero, buscar **interpretaciones cualificadas** (informantes clave y Delphi); cuarto, **cuantificar** los hechos y acontecimientos no registrados (encuestas); quinto, volver a la **apertura primordial** (el grupo de discusión) y finalmente intentar lograr la representación integral (total) del conjunto de resultados empíricos como consecuencia de su “complementariedad por deficiencia”.

Podemos aproximarnos a cada uno de estas cinco (+1) configuraciones empíricas desde prácticas muy distintas, en una gran medida porque el análisis de los datos obtenidos en los momentos precedentes (observación, datos secundarios y entrevista) permite una gran variedad de combinaciones. Pero, tradicionalmente, a cada uno de tales niveles se le ha asociado una práctica específica que define y bloquea, cada uno de estos momentos, sobre una técnica específica. Se trata sin duda de una convención útil, siempre que seamos conscientes de su significado provisional.

Comencemos con los registros, ya que, de una parte se trata del procedimiento de recogida de datos que se utiliza con más frecuencia en la evaluación, y de otra parte refleja algunas complejidades metodológicas que no se suelen tener en cuenta. Por este motivo vamos a prestarle una atención singular.

## 8.1. Los sistemas de información y registro.

### 8.1.1.- Una definición o como el todo puede ser la nada.

Los términos “sistema”, “información” y “registro” proceden de sus homónimos latinos “*systema*”, “*informatio/onis*” y “*registum*”, cuya grafía, y no sólo en las lenguas romances, apenas

se ha modificado con los siglos. Para el Diccionario de la Lengua Española de la Real Academia (DRAE) se trata de palabras con múltiples acepciones (hasta 31 para “registro”), no tanto en razón a una posible polisemia, sino todo lo contrario porque se trata de términos que, por su claridad denotativa, por el poder de enunciación y por la univocidad de sus significados, se utilizan continuamente como invariantes en las traducciones entre la mayoría de idiomas.

Se trata de palabras centrales en cualquier idioma, de códigos “*ex opere operatio*”, que proporcionan por sí mismas patente de existencia a los otros términos, como el “registro” de un barco que confirma fehacientemente, con independencia de aquello con lo que podamos tropezamos en la oscura bodega, lo que existe y lo que no existe en una nave determinada. Para los racionalistas tales términos constituyen formas categoriales de primer orden, no transformacionales y no regidas por otras lo que explicaría su alto nivel de condensación y que se utilicen para compensar la polisemia del lenguaje habitual. Aunque de hecho su propia complejidad produzca, como veremos, notables confusiones.

Obviamente los tres términos juntos, sin otra palabra que les acompañe, tratan de denotar la presencia de un orden, un sentido y un significado absolutos, una topología completa en una escala 1/1 de la pura realidad. Lo que en términos de conocimiento supone el ideal de un saber completo y sin fisuras. Cuando en una evaluación se recurre a crear un “registro total y exhaustivo” hay que sospechar, porque este tipo de registros no son ni fáciles, ni comunes, ni en muchas ocasiones posibles. Sin embargo ha sido tradicional que se utilicen como parapetos corporativos, es decir, la tendencia a echar mano de registros propios incluye el mensaje: no te metas que no sabes.

Por ello no es extraño que un sistema de información y registro se refiera, desde la perspectiva particular de la investigación social, a aquella estrategia de recolección de datos que establece un procedimiento mediante el cual se pretende conocer, recopilar y registrar la totalidad de casos o acontecimientos que define una determinada variable, con la finalidad de obtener la información que se precisa para legitimar de una forma contundente el diseño de políticas o intervenciones concretas. Es decir un registro supone a la vez adquirir completamente y utilizar socialmente un “conocimiento total” sobre un tema determinado<sup>22</sup>.

También tenemos que tener en cuenta que los Registros manifiestan otra doble faz, de una parte están concebidos para una determinada finalidad institucional y por ello se completan, como es lógico, con los procedimientos, adecuados y necesarios, de procesamiento, análisis y difusión de la información obtenida, de acuerdo con los objetivos previstos en una praxis definida. Pero por otra parte los datos recopilados pueden servir también, desde las posibilidades que nos ofrece el análisis de datos secundarios, para realizar evaluaciones con finalidades distintas a las previstas por el sistema de información y registro.

En ambos casos debe quedar claro que los Registros no se construyen desde la óptica de la investigación básica y que su metodología no se justifica desde la exclusiva búsqueda del conocimiento sino desde las necesidades y las prácticas institucionales (Arnau, 1978).

Pero a la vez debemos tener en cuenta que, como iremos viendo, en un momento histórico caracterizado por la constante aparición de Registros, los evaluadores no pueden obviar su existencia como fuentes de datos secundarios, pero tampoco pueden limitarse a considerarlos

22 La mejor descripción de la “utilización perversa de los registros y sus consecuencias” se la debemos a Jorge Luis Borges, en la *Historia Universal de la infamia (1954)*, en la cual en un capítulo llamado “Etcétera”, aparece un fragmento titulado “Del rigor de la ciencia”, atribuido a un tal Suárez Miranda y a una obra ficticia titulada *Viajes de Varones Prudentes* [“libro cuarto, cap. XIV, Lérida, 1658”], y que dice lo siguiente: “...En aquel Imperio, el Arte de la Cartografía logró tal Perfección que el mapa de una sola Provincia ocupaba toda una Ciudad, y el mapa del imperio, toda una Provincia. Con el tiempo, esos Mapas Desmesurados no satisficieron y los Colegios de Cartógrafos levantaron un Mapa del Imperio, que tenía el tamaño del Imperio y coincidía puntualmente con él. Menos Adictas al Estudio de la Cartografía, las Generaciones Siguientes entendieron que ese dilatado Mapa era Inútil y no sin Impiedad lo entregaron a las Inclemencias del Sol y de los Inviernos. En los desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas”, (versión más reciente, Barcelona, Destino, 2004). Por cierto, para algunos este podría ser el argumento de “Matrix”.

como tales fuentes de datos, ya que su diseño y administración también puede ser una tarea de los mismos. El hecho de que la evaluación asuma estas tareas posibilita además que los diseños adquieran un carácter menos corporativo y más relevante desde un punto de vista científico.

### **8.1.2.- Sistemas de información y registro, “estadísticas” e indicadores sociales.**

Es muy frecuente utilizar el término “estadísticas” para denominar a los Registros y para diferenciarlos de las técnicas muestrales a las que entonces denominan “encuestas” (Galvez, 1988). A la inversa, la identificación entre “estadísticas” y Registros trata de englobar tanto a los censos como a la muestras porque desde la perspectiva del análisis de datos no hay diferencias entre ambas técnicas que utilizan los mismos procedimientos de medida (Keyfitz, 1968). Desde esta misma óptica el término “registro” resulta redundante y basta con indicar “estadística”, mientras contrariamente en aquellos casos que se utilizan muestras es necesario indicar el tipo de muestreo utilizado, obviamente este es el resultado de sustituir la óptica de la recogida de datos por la óptica del producto de ambas técnicas: la matriz de datos. Un científico social no puede estar de acuerdo con una postura que considera que la forma en que han sido recogidos los datos, es decir los procedimientos empíricos utilizados, no producen ninguna consecuencia metodológica.

Tampoco debemos confundir los Registros con los “indicadores sociales” de los que hablaremos ampliamente en el capítulo 16, en el primer caso estamos ante un procedimiento para la recogida de datos y en el otro ante un modelo de análisis, o de utilización, de datos, el cual incluye la manipulación de Registros así como de otras series estadísticas temporales basadas habitualmente en muestras pero también en otras técnicas de recogida de datos. Los Registros representan, en un esquema racional de la metodología de la investigación social el momento empírico de la generalización intencional de la recogida de datos (Ortí, 1992), mientras que los indicadores sociales responden a un modelo teórico concreto de análisis de la realidad social que utiliza datos secundarios producidos y elaborados por diversas técnicas de recogida de datos (Carmona, 1977).

Ciertamente un Registro casi siempre puede ser utilizado como un indicador y en la práctica esta es la utilización más frecuente, pero ambos se sitúan sobre dos planos conceptuales muy distintos, asimismo si bien casi todos los Registros pueden ser utilizados como indicadores sociales, una parte sustancial de indicadores no son Registros. Como consecuencia, los intensos debates producidos sobre los indicadores sociales a los que nos referiremos en el mencionado capítulo 16, no pueden ser trasladados sin más a los Registros.

## **8.2.- Los sistemas de información y registro en la historia de las ciencias sociales.**

### **8.2.1.- Antecedentes: La modernidad y la lógica de la política democrática.**

La necesidad de disponer de registros, aparece de forma recurrente en todos los imperios antiguos, de tal manera que los censos y otras estrategias forman una parte indeleble de la historia de los procedimientos administrativos<sup>23</sup>. Con el Renacimiento y la posterior emergencia de los Estados Absolutos a finales del siglo XVI, se formaliza y adquiere rango de paradigma la idea de que son necesarios *“procedimientos que permitan describir de forma sistemática y continua el Estado para poder administrar las cosas”*, pero la propia socialización de esta idea en el seno de la naciente sociedad burguesa, la va a convertir rápidamente en una de las banderas con las que la sociedad civil reivindica una gestión democrática de estas mismas “cosas” frente a la discrecionalidad del poder absoluto.

<sup>23</sup> En España existen registros, conservados muy parcialmente o conocidos por referencias desde la época romana (los censos, los itinerarios, los registros de propiedad y de personas e incluso las tablas de mortalidad de Ulpiano), una parte de los registros se mantuvieron en la época visigoda y se ampliaron con diversos censos en la edad media (Sanz, 1956).

El movimiento no ha sido bien estudiado y los antecedentes que se citan son diversos según la nacionalidad de cada autor, ciertos nombres como Tomás Moro, Sansovino, el mismo Felipe II<sup>24</sup> y el periodo 1570-1590, son los que aparecen con mayor frecuencia (Swoboda, 1975), lo cierto es que desde la propia administración de los Estados Absolutos se inicia un proceso histórico cuyos hitos más relevantes van a ser la Ilustración en el siglo XVIII, el nacimiento de la sociología y la estadística positivista en el siglo XIX y finalmente la aparición del canon clásico de la sociología a finales del siglo XIX y principios del XX (Noelle, 1963).

Para la Ilustración el “registro” no era sólo una forma de contabilidad, ya que la “matemática social” formaba un todo con la “aritmética política” tal y como fue formulada por William Petty en 1671 (Lecuyer y Oberschall, 1968) y ambas representaban el ideal científico de la emancipación, por medio, como afirma Condorcet, del “registro de los hechos y su reducción a los datos” (Crepel, 1994), un ideal que incluía, “a) un lenguaje conceptual adecuado, b) métodos técnicos de recogida de datos y c) el sufragismo general igualitario”, (Granger, 1989), en cuyo contexto “el registro imponía la idea de que todos los seres humanos, convenientemente identificados, adquirirían así la plenitud de sus derechos” de tal manera que formar parte del registro “garantizaba la identidad personal y la integridad de los derechos humanos” (Granger, 1989).

La Ilustración impone la referencia cultural a un “ideal algorítmico” (Ovejero, 1987), de un reloj numérico que registra la totalidad de los acontecimientos del reloj universal físico y social (Ovejero, 1987). Un paradigma que irán asumiendo, y desarrollando, primero los fisiócratas con su nacionalismo cuantificador del cual son un buen ejemplo las “enquétes” de Colbert, después la práctica mercantilista que pondrá a punto los instrumentos de recogida de datos y de medida (Viner, 1968), lo que a su vez permitiría la aparición de la demografía, de las academias, de la economía clásica, de la enciclopedia y finalmente la emergencia conjunta de la sociedad civil y la sociología pero ya en el siglo XIX (Lecuyer y Oberschall, 1968).

En el siglo XIX, las condiciones sociales provocadas por la revolución industrial fueron el caldo de cultivo en el que confluyeron ciertas preocupaciones sociales comunes, entre las que conviene destacar las condiciones de vida de los obreros, la situación de la sanidad, de las prisiones y la prostitución, que con nombres como Beatriz Webb, Charles Booth, Morton Edén, John Howard y Parent-Dùchatelet, Quetelet, LePlay y Tónnies representan el auténtico punto de partida de las ciencias sociales (Lecuyer y Oberschall, 1968).

El discurso teórico y las bases conceptuales de estos autores no deben interpretarse desde los actuales parámetros ya que en el centro de sus preocupaciones se situaba la recogida masiva, la acumulación y el registro exacto de datos mediante una serie de técnicas ad hoc, que Quetelet puso a punto entre 1823 y 1832 (Lazarsfeld y Landau, 1968), así como la creación de organizaciones institucionales que gestionaran la recogida de datos y las propuestas de reforma. Era necesario saber todo acerca de los problemas sociales que se consideraban más relevantes para poder actuar sobre los mismos<sup>6</sup>. El mismo Quetelet construye sus conceptos, las famosas tendencias o propensiones, los promedios sociales y las leyes sociales bajo la óptica de un “registro de los hechos sociales que sustituyen a las ocurrencias permanentes de la naturaleza pero para el conocimiento de la realidad social (Lazarsfeld y Landau, 1968).

Por tanto, para todos estos autores, como la clave para el desarrollo de la Ciencia Social se situaba en el registro de aquellas regularidades sociales, que por sí mismas permitirían formular las leyes sociales, de la misma manera que el registro natural de los hechos físicos permitió formular las leyes de la naturaleza. Así se comprende que LePlay, en “*La Méthode Socialé*” (1887)

24 Dos grandes referencias históricas a los registros (publicadas por el INE) son el “Itinerario de Fernando de Colon” comenzado en 1517, así como los censos, los interrogatorios, las preguntas, las topográficas y las relaciones, todas ellas incluyendo estrategias metodológicas y listados de indicadores y variables, realizadas en la última etapa del reinado de Felipe II. Un conjunto de aportaciones que, en las publicaciones internacionales, son conocidas como “la opción filipina”, para referirse al hecho de “realizar registros muy amplios y sistemáticos, con el máximo de indicadores posibles, contando con grandes recursos y desplazar el análisis de las series obtenidas a las necesidades del futuro”.

construya un manual de campo con instrucciones muy precisas de como recoger los datos y como confeccionar los registros, mientras que el análisis de estos datos no se considera desde una perspectiva metodológica sino como parte de un proyecto de reforma social y política (LePlay, 1887), en última instancia el análisis de datos no tiene entidad mientras que su recogida para constituir un registro es el aspecto más relevante.

### 8.2.2.- La influencia de los Registros en la constitución de la sociología clásica.

Sin embargo, como sabemos hoy en día, la persecución y el registro de regularidades asimilables a las del mundo físico, no permite formular leyes sociales, una conclusión que hemos descrito en el apartado 4 del capítulo 2 y a la que llegaron los sociólogos que establecieron el canon clásico de esta disciplina, pero partiendo del marco que les ofrecía el propio paradigma del ideal algorítmico. Así Weber, realiza la primera mención de los “tipos ideales” en 1892 para la “*Verein Fiir Sozialpolitik*” una entidad dedicada a financiar el registro de cualquier tipo de acontecimiento social (Lazarsfeld y Landau, 1968), pero sobre todo debemos tener en cuenta que es Durkheim quién propone y realiza el primer análisis empírico “moderno”, lo que equivale a decir que “*la obra durkheniana, no es tan importante por sí misma [desde las perspectiva de la historia de la filosofía] como por representar la última síntesis, y la más avanzada, del proceso de cuantificación social que se inicia en el último tercio del siglo XIX*” (Ovejero, 1987). ¿Cómo consigue dar Durkheim este paso? Pues gracias a los registros. ¿Pero registros muy complicados? Pues no, sino más bien un tipo de registro muy elemental, que en la actualidad resulta incluso muy deficiente.

#### **SUPUESTO TEÓRICO UNIFICADO DE LOS REGISTROS UTILIZADOS POR DURKHEIM PARA “EL SUICIDIO”.**

CÓDIGO DE IDENTIFICACIÓN DEL CASO:

GÉNERO: [varón/mujer]

EDAD:

REGIÓN DE OCURRENCIA:

FECHA: [Día/mes/año]

DÍA DE LA SEMANA/HORA DEL DÍA:

ESTADO CIVIL:

SECTOR DE ACTIVIDAD: [Primario/secundario/terciario]

CAUSA PROBABLE DEL SUICIDIO: [Dolores físicos/Problemas económicos/Problemas familiares/Problemas sentimentales/Enfermedades mentales /Alcoholismo/ Sentimiento de culpa tras un crimen/Hastío de la vida...]

SUICIDIOS MILITARES: [Cuerpo/Grado/Tipo] [Profesional/voluntario/obligatorio]

Más de un siglo después del “*El suicidio*” no estaría de más recordar que uno de los libros más influyente en la historia de la investigación social se basó exclusivamente en los datos aportados por Registros. Ciertamente Durkheim no trabajó con estos sistemas sino que los usó como datos secundarios por lo que tradicionalmente el trabajo de Durkheim ha sido presentado como un “*análisis de datos secundarios*”, pero ¿de dónde procedían estos datos?, pues de una serie de Registros que Durkheim identificó de forma muy precisa (Durkheim, 1897). Se trataba de unos sistemas muy simples, que recogían muy pocos datos según muestra el supuesto teórico adjunto que acumula los contenidos de todos Registros utilizados en “*El suicidio*” sobre un único cuestionario. Es decir Durkheim trabajó con muy pocos datos y además incompletos, porque procedían de diferentes registros de diversos países europeos.

Puede alegarse que el hecho de que los datos procedieran de Registros es poco relevante, que se trata más bien de una curiosidad histórica, pero no es así, porque “*El suicidio*” es justamente un trabajo empírico posible a partir del hecho de que los datos manejados proceden de Registros. Durkheim trabaja con los datos que estos registros le ofrecen, interpretando que

se trata de “casos reales totales” que proporcionan tasas de suicidio más o menos constantes pero distintas en cada país (Durkheim, 1897), lo que le permite realizar un análisis comparativo con datos procedentes de otros Registros, como los “casos de locura”, los “casos de psicosis alcohólica” en terminología actual, las temperaturas medias diarias o los “casos de homicidios y asesinatos”. Es decir Durkheim se permite comparar cosas distintas porque todas ellas proceden de Registros, lo que no podría hacer si las fuentes fueran muestrales, salvo que se tratara de muestras idénticas aplicadas en las mismas condiciones. Ni tan siquiera habría sido totalmente correcto comparar los datos de suicidios con ciertas variables generales, como tamaño de las familias, predominio racial o religioso, salvo que tales variables también hubieran sido introducidas en la muestra.

En última instancia “*El suicidio*” fue posible porque Durkheim trabajó con Registros y las conclusiones del texto no son, como sugiere Parsons, el producto de un racionalismo cartesiano, sino los resultados de un particular análisis inductivo con los resultados procedentes de diversos Registros y otras fuentes y procedimientos de recogida de datos.

He elegido “*El suicidio*” por su representatividad pero también por honrar su carácter centenario, pero lo mismo podría haber hecho con otros autores clásicos y en todos ellos (desde Weber hasta Merton) se encuentra la misma disyuntiva: hay que crear registros, hay que utilizar los registros, pero por sí mismos nunca son suficientes. Y esto es especialmente cierto para la evaluación.

### **8.2.3.- Las circunstancias de una crisis.**

El proceso histórico de constitución de los Registros condujo a finales del primer cuarto del siglo XX a una situación extraña, de una parte la abundancia de “estadísticas” en las que se había puesto tantas esperanzas de mejora social producía, por falta de usuarios, un “*cementerio de los números*” al que se le veían ya muy pocas utilidades (Swoboda, 1975), pero por otra parte los registros disponibles no parecían capaces de proporcionar elementos para alimentar las necesarias reformas sociales y políticas. Es decir, se tenían muchos datos sobre determinadas poblaciones o circunstancias vitales pero este conocimiento por sí mismo, como acertadamente había señalado Marx, no servía para cambiar el mundo como pretendían los reformadores.

En este contexto es cuando aparecen las técnicas de recogida de datos basadas en muestras que vamos a describir en el capítulo 11. Se puede adelantar en todo caso que la primera versión completa, y justificada, de una muestra la realizó el matemático noruego Kiaer en 1895. A partir de esta fecha se produjeron investigaciones con muestras de manera creciente en los años siguientes, de tal manera que a fines de la década de 1930 el empleo de muestras representativas en sustitución de censos y registros ya se había generalizado en la confección de las estadísticas oficiales norteamericanas (Deming, 1968). Tampoco se puede obviar el impacto político de la encuesta de Roper, Crossley y Gallup en la elecciones presidenciales de 1936, así como las crecientes necesidades de la sociedad de consumo. Lo cierto es que en apenas treinta años podía suponerse que el paradigma de los “registros sociales” había desaparecido difuminado en la noche de la historia, mientras las encuestas por muestras que producían datos representativos de forma casi inmediata ocupaban el centro, cuando no la totalidad, de las técnicas de recogida de datos.

Justamente este será el momento del triunfo del idealismo cuantificador (Martín Serrano, 1978), porque a pesar del papel paradigmático de los registros a lo largo de casi tres siglos, se daban dos cautelas que evitaban la práctica de la idealización de la medida, por un lado el coste y el largo tiempo necesario para la obtención del dato, y por otro lado su finalidad sociopolítica global. En cambio la muestra implica menores costes, pero también inmediatez y por supuesto una gran flexibilidad en cuanto a los objetivos perseguidos, lo que permite la emergencia del idealismo cuantificador.

Este idealismo, que adquiere la forma de un imperialismo cognitivo, consigue su primera víctima en los propios registros, no los elimina como técnica posible, pero si los elimina del mapa

cognitivo y conceptual de los propios científicos sociales, por ejemplo si preguntamos a cualquiera cual es la fuente de datos que utilizó Durkheim para *“El suicidio”* seguramente nos responderá acertadamente que “estadísticas”, pero más difícilmente identificará estas estadísticas como “registros de casos” e imaginará, como ya he explicado, sólo la matriz de datos que los contiene.

En todo caso el éxito en la utilización de muestras fue tal que incluso los censos, que mantienen un sistema de recogida de datos exhaustivo, comienzan a ser explotados por métodos muestrales a partir de 1930 y en general muchas estadísticas vitales como las contenidas en los Movimientos Naturales de la Población son sustituidas, o más frecuentemente mantenidas como sistemas de registro burocrático y completadas por encuestas (Grebenick, 1968). Un ejemplo muy evidente de esta transformación lo constituye la Encuesta Mundial de Fecundidad, que se plantea como un instrumento para compensar la inexistencia (o la baja calidad) de los registros demográficos en muchos países, pero que acaba por proporcionar, a nivel mundial, información no registrada en torno al complejo fenómeno de la fecundidad.

#### **8.2.4.- Los registros en la antropología social.**

De forma paralela desde la antropología social se desarrollaron procedimientos de recogida de datos, como la observación, el registro de observaciones, la observación participante, los censos y genealogías y los testimonios personales, en un periodo de tiempo limitado que se extiende desde la llegada de Boas a Columbia en 1896, hasta la redacción de *“Los argonautas del Pacífico Occidental”* por Malinowski en 1918 (publicado en 1922), es decir un periodo que comienza con la plena vigencia de los Registros y acaba justo antes de la expansión de los métodos muestrales.

Quizás por ello la antropología construyó su propio modelo de técnicas de recogida de datos al margen del resto de las ciencias sociales, y quizás por ello la antropología ha mantenido viva la técnica del “registro intencional”<sup>25</sup> perfectamente asimilable a los Registros, situándola además en el mismo plano que las otras técnicas de recolección de datos, de tal manera que los “censos” que hacen los antropólogos se perciben como una técnica más junto a las encuestas, la observación participante o las historias de vida (Rossi y O’Higgins, 1980). Por regla general se asimilan en Antropología “censo” y “registro”, porque en la mayoría de los casos el único registro utilizado es el de la población, pero en otros se establecen las diferencias y se habla de *“técnicas demográficas”* por una parte y de *“registros de datos”* por la otra (Cresswell y Godelier, 1975).

Pero a la vez, esta es una técnica que también ha ido perdiendo peso en la antropología y ya ni tan siquiera es citada como tal en los manuales académicos más modernos (Kottak, 1993), quizás porque ha ido ocupando, en el propio esquema metodológico de los antropólogos, el lugar de una técnica de sustitución, es decir, se utiliza cuando faltan o hay que completar ciertos datos demográficos, en una gran medida porque es un trabajo “lento y pesado”, aunque también se reconoce su potencial como una estrategia que por una parte permite establecer contactos y vínculos y por otra sirve para recoger sistemáticamente información que las otras técnicas pueden dejar al margen (Navarro, 1983).

Frente al “registro” y cuarenta años más tarde que en la sociología, la utilización de muestreos intencionales no probabilísticos con identificación y selección del caso mediante el procedimiento de “bola de nieve” sin otro criterio de intencionalidad, salvo en algunas ocasiones, cuotas por género y edad, se han hecho relativamente populares especialmente para trabajar con “poblaciones ocultas” (Funes y Romaní, 1985; Diaz et al, 1992; Gamella y Álvarez, 1997; Pallares, 1995), aunque ciertamente tales técnicas se codifican e identifican como “métodos cualitativos”.

25. En realidad no se trata de una técnica sin más sino de todo un estilo científico, cuya pervivencia cabría relacionar con el predominio del método comparativo en antropología: se trata de conocer el mayor número posible de hechos para que las comparaciones resulten lo más ricas posible y para conocer muchos hechos lo mejor es registrarlos. Este es el espíritu que anima la “Guía” de Murdock, la idea de un sistema de registro universal que “clasifique las informaciones sobre la cultura, la conducta y el medio general de cualquier sociedad” (Murdock, 1954), con 88 secciones y 670 categorías de clasificación.

En realidad la confusión entre método y técnica adquiere dos perfiles distintos, en un primer caso la técnica de la muestra sirve para identificar casos a los que después se les va a aplicar otra técnica de recogida y análisis de datos, es decir una metodología concreta como por ejemplo historias de vida, (Funes y Romani, 1985; Díaz et al, 1992; Pallares, 1995), pero en otros casos los muestreos no probabilísticos conducen sólo a la aplicación de un cuestionario, cuyos resultados se van a generalizar, mediante el análisis estadístico descriptivo (Gamella y Álvarez, 1997), es obvio que aunque se utilice el logotipo de “antropología” o “método cualitativo”, no estamos ni ante un registro ni ante ningún tipo de técnica cualitativa, sino sólo frente a una encuesta con los importantes sesgos que los muestreos no probabilísticos producen.

Esta evolución de la Antropología muestra cómo no se puede estar al margen de la legitimidad que cada tipo de sociedad otorga a una u otra técnica de recolección de datos. Por este motivo en poblaciones poco avanzadas o en ciertos ámbitos campesinos, marginales o subculturales, la vieja técnica del registro antropológico es aun muy útil, aunque en otro tipo de poblaciones la opción por las muestras representativas resulta inevitable.

### **8.2.5.- La nueva irrupción de los Registros.**

Pero todo esto ha cambiado en los dos últimos decenios, conformándose una nueva situación de emergencia de los sistema de información y registro, que podemos atribuir a la confluencia de dos tipos de circunstancias, por una parte el desarrollo de las Nuevas Tecnologías de la Información y la Comunicación (las TIC) que incluyen la vertiente del registro y análisis de datos, y por otra parte los cambios sufridos en el papel del Estado y en particular las nuevas misiones asignadas a las administraciones públicas. Ocurre que el Estado se debilita, pierde legitimidad como autoridad pero la mantiene como gestor de la “cosa pública”, lo que obliga a las administraciones públicas a demostrar capacidad no sólo para gestionar de forma neutral los asuntos públicos sino a incrementar de forma continua su eficacia. En última instancia incluso los resultados electorales dependen cada vez más de esta demostración de eficacia, que de las propias ideologías políticas de partida.

Un ejemplo, poco conocido pero relevante (que conozco por mi trabajo) para observar este cambio, lo constituye la transformación operada gracias a los registros informáticos por el sistema judicial y en particular del Departamento de Justicia de los EE.UU., una evolución que comenzó a principios de los años 90 y que lo va alejando de las prácticas, más formales, de los sistemas judiciales europeos (Inciardi, 1993; 1994 y 1996, MacKenzie, 1994).

Como consecuencia de los cambios en los sistemas de registro (y el uso atribuido a los mismos), se ha conseguido que la delincuencia, que el imaginario norte-americano consideraba el “mayor problema” del país desde finales de los años 60, ha pasado a segundo plano en los debates políticos, gracias a una doble estrategia, por una parte el desarrollo de políticas que incluyen medidas alternativas, programas de sustitución, mayor sensibilidad en los criterios de clasificación, endureciendo algunos y flexibilizando otros, así como una evaluación sistemática de todas las actividades.

Pero a la vez los complejos Registros establecidos, que recogen la totalidad de la información, la distribuyen (en términos de gestión del conocimiento y de los casos) entre los diferentes órganos judiciales de los diferentes territorios, a la vez que a la opinión pública, modificando las actitudes, opiniones y percepciones de unos ciudadanos para los que el sistema no funcionaba y que exigían mayoritariamente castigos generalizados (Garland, 2001; Young, 2003). Los registros (y la informática) han contribuido a crear un sistema judicial transparente y la transparencia ha producido resultados magníficos por inesperados.

Se ha producido así y gracias al papel otorgado a los registros (y a la evaluación) una absoluta inversión de valores, de tal manera que una situación que fue una pesadilla política para numerosas administraciones es ahora una baza electoral para cualquier candidato que quiera asumir el “buen funcionamiento judicial”. La idea es que el sistema es eficaz porque “controla

y registra”, porque evita errores (tanto en sentido de “errores judiciales” como en el sentido de “impunidades resultado de la mala gestión administrativa”), y además es esencialmente eficaz porque la combinación entre registro y evaluación lo hace más transparente. Los sistemas judiciales europeos (en los que la presencia de evaluadores externos y de sistemas de registro público resultan inimaginables), tienen aunque que vivir muchas crisis para alcanzar esta situación.

Hemos utilizado este ejemplo radical y crucial para mostrar como los sistemas de registro (en combinación con las TIC) emergen como una alternativa más adecuada a las necesidades de evaluación (y representación) de las políticas.

A modo de conclusión podemos afirmar que parece concluir la etapa histórica de la exclusividad de las técnicas muestrales (que se inició en los años 50 en EE.UU. y en España en los años 70), que apareció asociada a la necesidad de obtener datos fiables de forma rápida, y en especial al desarrollo de las encuestas electorales, es decir un factor sociopolítico como era el proceso de implantación y estabilización de los sistemas democráticos, así como a las necesidades de información del sector publicitario en la sociedad de consumo. Este fin de etapa se corresponde, a partir de la década de los años 90, con la resurrección de los sistemas de Registro, lo cuales parecían haberse olvidado salvo en los manuales que los citaban como una especie de reliquia histórica. Pero como consecuencia del cambio social y el desarrollo de las TIC, comienzan a ofrecer posibilidades (en la gestión y evaluación) que eran inimaginables unos pocos años y la idea de utilizar esta técnica de producción de datos adquiere una singular fuerza, en particular en el campo de la evaluación.

### 8.3.- El diseño de un sistema de información y registro.

¿Es fácil poner en marcha un sistema de información y registro? Tenemos que insistir que hace unos años resultaba muy complicado y requería un compromiso institucional fuerte, tanto por el coste como por el volumen de papeles a manejar y por supuesto a causa de la necesidad de “comprometer” a todos los que debían “recoger” la información. De hecho la mayoría de los viejos registros se sustentan en normas legales porque el “peso de una norma legal” era necesario para que un Registro comenzara a caminar. En la actualidad, en cambio, un registro se puede concebir con una (o varias) personas decididas y un ordenador. Se trata lógicamente de una persona que ocupe un lugar en el programa o la política que le posibilite registrar los datos. Pero esto no es problema porque son muchas las personas que ocupan estos lugares, dispuestas a participar en una evaluación y que de una forma casi artesanal podrían establecer un Registro más completo incluso que los “registros institucionales” de hace apenas unas décadas. Es una cuestión de información y actitud, ya que en la actualidad y con las TIC los medios necesarios son una cuestión secundaria. Claro que en ocasiones las actitudes son lo más difícil de cambiar.

Poner en marcha un Registro es una cuestión sencilla pero requiere algunos conocimientos. Lo primero que debemos saber es que la puesta en marcha de un Registro supone realizar una definición racional de contenidos con una doble finalidad empírica, por una parte recoger los datos para obtener información y por la otra utilizar la información para tomar decisiones. Sin esta doble finalidad el Registro se limita a ser un procedimiento, para Registrar cosas, pero que no sirve para nada.

La clave está en la afirmación: ***“nadie se compromete a llevar bien un Registro si después no le sirve para obtener información relevante y a su vez le ayude a facilitar y mejorar su trabajo”***. De hecho, los profesionales de sectores como el social y el sanitario ya comienzan a estar hartos de registros que no sirven para nada, es decir, que registran pero ni informan ni ayudan a tomar decisiones. Hay que tratar de evitar que en Juventud pase lo mismo.

Estamos, en todo caso, ante un procedimiento que en origen es puramente deductivo, mediante el cual se determinan los datos que se van a recoger, la forma de hacerlo y de concep-

tualizarlos, pero que acaba convertido en un estricto, pero en parte ficticio, procedimiento inductivo de análisis de datos. Una transformación metodológica llena de posibilidades que, por sí misma, merecería la máxima atención por parte de los evaluadores.

En realidad es ahí donde se encuentran las dificultades, es decir no tanto en la recogida de datos, sino en el diseño del sistema. Porque la fase deductiva se inicia con la justificación teórica de las necesidades, así como los componentes principales que van a constituir el Registro, para pasar después a determinar el campo de aplicabilidad, las unidades de conducta a registrar, las categorías o dimensiones del contenido y la delimitación del criterio de exhaustividad (Angera, 1985). Es decir, muchas cosas que requieren su tiempo y un cierto conocimiento teórico y práctico. Es decir, la informática ha resuelto muchos problemas, pero el diseño de los contenidos requiere la presencia de especialistas en la materia y un cierto tiempo para definir cuales son estos contenidos.

Esto significa que desde una perspectiva técnica se requiere una planificación compleja y detallada del registro y su contenido. Tal planificación incluye la realización de ensayos previos, la selección de los recursos y la selección y formación del personal que va a aportar datos y a utilizar el registro. Estas son tareas que sólo puede realizarse con un cierto grado de racionalidad instrumental y desde un lugar con una clara autoridad formal y moral (Rootman, 1983), desde la que se pueda dar respuesta a posibles problemas administrativos, jurídicos, éticos y técnicos.

Por otra parte conviene tener en cuenta que el diseño de un Registro implica a la vez actividades periféricas y centrales. Por actividades periféricas se entiende la parte del trabajo de campo que identifica las unidades o sujetos que producen la información pertinente, así como la recogida sistemática de dicha información sobre un soporte standard para su envío a la unidad central. Las actividades centrales se refieren a la recogida de la información obtenida a partir de la administración física o electrónica de cada uno de los soportes y su acumulación en un registro de los mismos. En todo caso las matrices de datos de cada uno de los registros deben quedar perfectamente identificadas como un registro, nominal o no, pero en todo caso individualizado. Finalmente sobre dicho registro se pueden aplicar las técnicas de análisis, en general técnicas estadísticas, que se consideren oportunas.

En determinados casos, cuando se trata de un sistema local permanente o creado ad hoc para una evaluación concreta, la actividad periférica y la central, pueden ser realizadas incluso por el mismo equipo de evaluación, pero aún en este caso las actividades periféricas y las centrales deben quedar perfectamente separadas.

Asimismo tenemos que tener en cuenta que en términos conceptuales un Registro es *per se* un diseño longitudinal muy preciso, ya que no sólo se aplica a la misma población y con la misma muestra como en los paneles, sino que refleja la totalidad del conjunto de hechos o casos afectados en una población determinada, es decir es un panel completo en el que los datos recogidos poseen un grado de significación muy preciso, porque son, o pretenden ser, todos los datos. Es además una serie temporal que puede acumular los casos por lo que ofrece a la vez resultados diacrónicos y sincrónicos. En todo caso mientras que con los estudios longitudinales o los paneles basado en muestras sabemos más o menos a donde vamos a llegar, con el diseño de los Registros imaginamos una proposición que se mantendrá estable a largo plazo y la hacemos pública, pero deberíamos ser conscientes de que cuando diseñamos un sistema de registro estable *"lanzamos una apuesta contra una realidad cultural imprevisible"* (Goodenough, 1971).

A pesar de la rigidez de su diseño, o quizás a consecuencia de todo ello, los Registros permiten tipos de análisis para los que otras técnicas tienen dificultades o exigen delicadas adaptaciones, por ejemplo cualquier registro contiene automáticamente un diseño por cohorte, porque la edad y la fecha de recogida de la información suelen figurar en casi todos los Registros. Asimismo se acoplan muy bien a los diseños experimentales, ya que se pueden introducir modificaciones, en la realidad registrable o en la forma de registrarla, crear grupos experimentales y grupos control y hasta diseños de doble ciego y el Registro recogerá los efectos a lo largo del

tiempo de todos estos supuestos post-facto. El trabajo del departamento de empleo británico descrito en el prólogo es un buen ejemplo de esta estrategia, ya que a partir de un registro de partida, organiza y distribuye los casos de forma aleatoria a partir de los propios datos del registro, para realizar el experimento. Un ejercicio para el lector: ¿se podría hacer aquí lo mismo con los datos del Registro del INEM?

## 8.4.- Tipos de sistemas de Registros.

### 8.4.1.- Anónimos y confidenciales.

Por un **Registro anónimo** se entiende aquel en el que el registro no contiene la identidad de los sujetos registrados, bien porque, como dicen los epidemiólogos, no establece una relación entre el acontecimiento y el sujeto o bien porque simplemente no se considera de utilidad este dato. Obviamente los Registros anónimos tienen problemas a la hora de evitar duplicidades. Como consecuencia existe una cierta tendencia a crear sistemas que manteniendo el anonimato identifican el caso e impiden las duplicidades. El ejemplo más común lo constituyen aquellos que crean códigos de identidad a partir de datos personales, por ejemplo, la primera letra del nombre, las dos primeras letras del segundo apellido y día y mes de nacimiento con cuatro dígitos, lo que en mi caso personal sería [DAR0705], con una posibilidad teórica de coincidir sólo cada 5.045.760 casos, cifra que se puede aumentar añadiendo cualquier otro elemento por ejemplo una letra que identifique sexo, lo que duplicaría la seguridad.

Por confidenciales se entienden aquellos Registros que recogen la identidad de los sujetos pero que, de acuerdo con las leyes no la difunden. Obviamente los sistemas no son confidenciales por sí mismos sino porque la Ley, la LOARTAD y la LEFED y numerosas normas de diferente rango en el caso español, lo exige, pero a pesar de todo esto plantea muchos problemas ya que, por ejemplo, obliga a regulaciones laborales especiales para los trabajadores (incluidos los científicos) que manejan Registros (Sagardoy de Simón, 1996) y siguen dándose muchas incógnitas sobre el derecho de acceso de los ciudadanos, tanto a consultar sus propios datos como aquellos otros elementos presentes en los registros y que consideran que les pueden afectar (Pomed, 1989).

Por si fuera poco una parte de los Registros están informatizados y son accesibles mediante conexiones telefónicas. Si los piratas informáticos “entran” en los más sofisticados sistemas de seguridad militar, les cuesta mucho menos acceder a los cientos de Registros que recogen información sensible sobre los ciudadanos, aunque esto no ha ocurrido hasta ahora no tanto por la dificultad técnica sino porque el tema no ha sido codificado aún como “interesante” por parte de las representaciones sociales y por tanto no constituyen aún ningún tipo de reto para tales “piratas” (o no lo difunden).

### 8.4.2.- Notifican acontecimientos y notifican casos

Por notificación se entiende, tanto el procedimiento electrónico como el medio físico que transmite la información desde la fuente en que está se produce y se observa, hasta el registro. La ruta de la notificación puede concluir en las actividades periféricas (las cuales mandan entonces sólo las matrices de datos a la central) o recorrer todo el camino desde la fuente hasta el registro central.

También puede ocurrir, y de hecho esta es la estructura que tienen algunos Registros, que ni se notifique ni proporcionen todos los datos al organismo central, manteniéndose el registro completo sólo a nivel local. En estos casos se pueden diseñar proyectos de investigación que recogiendo los datos locales produzcan resultados generales en forma de series estadísticas como un trabajo ad-hoc de acumulación de datos, como ocurre en las estadísticas demográficas básicas.

En ambos casos se pueden notificar acontecimientos, por ejemplo cada vez que cualquier persona acude a presentar una denuncia o a solicitar una prestación se puede notificar el hecho e incluso añadir alguna característica sociodemográfica del denunciante/solicitante, o bien se pueden notificar casos, que en el ejemplo anterior sería la persona que presenta la denuncia, que quedaría identificada como “víctima” con independencia de las denuncias que presentara.

La notificación de acontecimientos es en general anónima pero, aparte de que no evita las duplicidades, contiene información del tipo “respuesta múltiple” por lo que no es muy útil a la hora de ofrecer la proyección social del tema, en cambio mide muy bien el “volumen de trabajo” de las instituciones, en el ejemplo anterior el volumen de denuncias que tramita la policía en vez del volumen real de delitos y su distribución.

Pero también se pueden notificar casos, una tarea más compleja, ya que se debe garantizar la confidencialidad o en el anonimato, pero, como ya he dicho, implica muchos riesgos éticos, en todo caso no mide el volumen de trabajo institucional sino la presencia social de un determinado fenómeno como es la distribución de las víctimas de delitos. Si además el sistema acumula acontecimientos sobre casos se puede obtener una especie de “fotografía natural” del fenómeno. En este último caso debería tratarse de un tema “no íntimo” ya que para acumular acontecimientos el Registro debería ser público para que pudieran retroalimentarlo las actividades periféricas.

#### **8.4.3.- Especializados y no especializados**

Por sistema especializado se entiende aquel que ha sido creado exclusivamente para registrar un hecho particular, como el suicidio, el SIDA, los expedientes académicos o las condenas penales, la mayoría de nuevos Registros son especializados, mientras que entre los antiguos predominan los registros no especializados es decir que recogen varios hechos al mismo tiempo.

Obviamente la condición de especializado o no especializado depende del punto de vista que se maneja, por ejemplo el extinto SURI de Instituciones Penitenciarias era un Registro especializado desde la perspectiva de la situación sanitaria de los internos en prisiones, pero era un Registro no especializado desde la perspectiva de la medición de los niveles de consumo de drogas o posible seropositividad de los presos.

Los sistemas especializados son más fiables pero más costosos, en cambio “colgar” un Registro especializado de otro Registro pre-existente es más barato y rápido aunque entonces se ve inmerso en los avatares del sistema principal. Los sistemas no especializados pueden establecer una relación simbiótica cuando comparten costes, beneficios y toma de decisiones, una relación parasitaria cuando uno de ellos aprovecha la estructura del otro y además le produce costes y problemas adicionales o una relación comensalista cuando aparece un mero aprovechamiento sin costes.

### **8.5.- Dificultades inherentes al marco institucional.**

#### **8.5.1.- Hegemonías corporativas.**

Los antiguos sistemas de registro de base manual recogían, en general, datos relacionados con la demografía, con el registro de las actividades institucionales, con la propiedad y con ciertos problemas sociales que desde la óptica de la revolución industrial se consideraban muy significativos. El Estado, con el apoyo de organizaciones filantrópicas, se ocupaba de gestionar los reg135

istros que, en muchos casos, estaban además regidos por leyes específicas.

En cambio los nuevos sistemas basados en las TIC han sido concebidos desde ámbitos más sectoriales o particulares, pero con distintas referencias hegemónicas en cada país, lo que confirmaría la explicación socio-política de su emergencia.

Así en Estados Unidos el ámbito hegemónico en el desarrollo de los Registros es el Sistema Judicial (Inciardi, 1993; 1994 y 1996; MacKenzie, 1994), en cambio en Suecia es el Sistema de Servicios Sociales (Segraeus, 1994, SIS, 1994) y en España seguramente es el Sistema Sanitario. En los tres casos coincide la presencia de un grupo corporativo de enorme poder en cada uno de los países, con una historia de crisis en su imagen social en el período anterior a la implantación de “sus” Registros, pero que a la vez, quizás para superar la mala imagen pública, se lanza a un proceso de racionalización sectorial con el que además pretende liderar la propia modernización del país. En los tres casos se ocupa además de aquellos problemas que, con independencia de su importancia real, tienen una gran presencia en el imaginario social de cada uno de estos países, es el caso de la delincuencia en EE.UU., de la integración social en Suecia o de la Salud en España.

Esta hegemonía corporativa dota asimismo de determinadas características al proceso de implantación de los nuevos Registros en cada uno de los países, así en España se está adoptando para los Registros el modelo, el lenguaje y los hábitos metodológicos de los epidemiólogos, mientras que los modelos Suecos y Norteamericanos son mucho más sociológicos, ya que por los respectivos temas son sociólogos y trabajadores sociales los responsables de su implantación.

### **8.5.2.- Antiguos y nuevos Registros.**

Podemos hablar también, en relación a sus orígenes históricos, de la actual presencia de dos tipos de Registros, los antiguos que recogen una dilatada trayectoria que en muchas ocasiones enlaza incluso con la “edad de oro” de los registros hace ya un siglo y que se caracterizan por su compleja burocracia, por su manualidad (“estadillos” o “comunicados” que se elaboran sin criterios metodológicos en la periferia y que se acumulan manualmente, sin ningún control de calidad, en la unidad central), por sus dependencias a un cuerpo administrativo concreto y por la falta de fiabilidad de los datos obtenidos que son “rellenados” desde una perspectiva de “evaluación implícita de procesos” a la que se atribuye consecuencias laborales y profesionales, lo que introduce errores, sesgos e incluso falsificación, lo que a la vez produce, en el propio grupo corporativo que lo elabora, una nula credibilidad.

Un modelo, casi tópico y en todo caso extremo, de sistema antiguo lo constituye en España la costosa e inútil “*Estadística de Asuntos de la Fiscalía del Tribunal Supremo*” con la que se elabora el “*Informe de la Fiscalía del Tribunal Supremo*” a cuya presentación acuden todos los años el Jefe del Estado y el Presidente del Gobierno. Ciertamente la presencia de tales autoridades en el acto de presentación de los resultados de un Registro resulta llamativa pero más si consideramos que los datos presentados tienen graves problemas de validez. Al mismo tiempo esta situación representa perfectamente el lugar simbólico que un viejo registro puede ocupar como expresión, no tanto de la labor de una institución concreta y de sus posibilidades como fuente de datos, como de su inserción en el conjunto de poderes del estado.

En cuanto a los nuevos Registros puestos en marcha a partir de los años 90 e identificados con las TIC y los otros componentes señalados en los epígrafes precedentes, se caracterizan además por su completo y explícito diseño metodológico, que incluye las formas de presentación y difusión, así como por su vinculación a las estrategias que demuestran la labor de una corporación o una institución particulares. Los viejos Registros escenifican rituales de autoridad tradicional mientras que los nuevos representan escenas de eficacia que esperan ser apreciadas por un público cada vez más exigente y menos dado a impresionarse por los rituales cuyo contenido no este aparentemente refrendado bien por la ciencia o bien por una ética pública con arreglo a fines.

### 8.5.3.- Ámbitos de aplicación y perspectivas de futuro.

El número de ámbitos sobre los que es posible diseñar los nuevos Registros parece infinito, como infinitos son los posibles tipos de conductas humanas, sin embargo conviene establecer ciertas limitaciones en cuanto al interés de los mismos por parte de los evaluadores y a las consecuencias sociales de mantener un determinado tipo de Registro, por ejemplo, podemos pensar que en un CIJ puede ser interesante el Registro confidencial de los menores que piden información sobre métodos anti-conceptivos, porque hemos puesto en marcha un programa de promoción de la información sexual para adolescentes, pero también tenemos que tener en cuenta el interés “particular” que puede suscitar dicho registro en otros actores sociales movidos por intereses “morbosos” o “morales”. ¿Nos merece la pena correr este riesgo? Una pregunta que no debe responderse en forma abstracta sino que será la práctica de la evaluación concreta la que determinará el **interés real** que pueda tener un determinado Registro.

En todo caso hay numerosos ejemplos de Registros en todos los niveles de la administración (y algunos más en otros países más desarrollados) y el proceso de creación de registros parece imparable. Por este motivo, antes de crear, parece más interesante conocer los que ya existen.

En esta revisión (y a modo de evaluación de la misma) no hay que obviar que la actual emergencia de los Registros responde a una lógica social e institucional que, a la vez, contiene por una parte elementos estructurales como son los cambios producidos en el papel del Estado o la demanda de “información” originada por los altos niveles educativos alcanzados por las últimas generaciones y por otra parte elementos tecnológicos como el desarrollo de las TIC, lo que equivale a afirmar que el futuro de los Registros se juega en un terreno en el que la influencia *per se* de las actitudes que vayan a tomar los científicos, o al menos de los evaluadores, es muy limitada.

Ocurre además que desde la perspectiva del cambio social, los elementos que configuran la creciente presencia de Registros, son aquellos con los que se definen las opciones de futuro más probables, al menos para las sociedades desarrolladas, lo que equivale a afirmar que apenas estamos en el umbral de la “*etapa de los nuevos Registros a partir de las TIC*”. En perspectiva sabemos que la historia de las ciencias sociales, en lo que a las técnicas de recogida de datos se refiere, ha pasado por dos etapas, la de la totalidad explicativa que se prolonga hasta los años 30 de este siglo y la de la urgencia connotativa que, con la encuesta muestral primero y el añadido del grupo de discusión desde los 60, sigue manteniendo la hegemonía. Pero en prospectiva no es difícil imaginar, en parte porque el proceso ya se ha iniciado, una etapa de conocimiento limitado y no sólo en relación a las técnicas de recolección de datos.

La nueva etapa de los Registros se constituye en un momento histórico caracterizado por enormes avances en el conocimiento y crecientes dificultades en sus aplicaciones sociales. Las nuevas potencialidades de las técnicas de la recogida de datos (y en particular los registros) nos desbordan, pero la posibilidad de realizar evaluaciones fuertes y realistas con estos datos se nos antojan complicadas. No es sólo un problema de la evaluación ya los físicos teóricos tienen ya muy clara que no pueden avanzar mucho más lejos, aunque teóricamente podrían, porque chocan con los límites de la capacidad cultural humana y la organización social para interiorizar estos avances, por este motivo están comenzando a replegarse en sí mismos, avanzando al ritmo que marca el desarrollo cultural y social, aunque mantienen enconados debates sobre el sentido y significado de este repliegue.

Desde el ámbito de la evaluación podemos lanzar, a modo de metáfora, una propuesta similar: los nuevos Registros representan posibilidades inimaginables para el conocimiento, pero para nada los vamos a utilizar en la perspectiva que en su día escenificó la ampulosidad totalizante de las expectativas de la ciencia. Desde la evaluación los Registros sólo pueden ser una herramienta para la micro-gestión, para la evaluación y la difusión de las respuestas proporcionadas a problemas sectoriales.

Esto significa que el futuro de los Registros, al depender de los factores estructurales media-

dos que vengo señalando, es una variable independiente de la actitud de los evaluadores, que sin embargo no pueden dejar de percibir a los propios Registros como un componente de todas las nuevas (y más modernas) intervenciones sociales concretas.

## 8.6.- Evaluar a partir de registros: oportunidades y riesgos.

Partiendo de la posición señalada en los últimos párrafos del apartado precedentes nos queda contestar a la pregunta ¿Cómo utilizar los Registros en el ámbito de la evaluación y en particular en la señalada perspectiva de la micro-gestión sectorial?. Como hemos visto las opiniones, las oportunidades y los riesgos son muchos y en ocasiones contradictorios. No cabe duda que, en primer lugar, la irrupción de las oportunidades que nos brindan las nuevas tecnologías debe ser analizarlo como un componente de la propia realidad y del cambio social, en segundo lugar también parece bastante evidente, al margen de sus problemas de validez y pertinencia los cuales no son tampoco más graves que los de otras fuentes de datos, se pueden utilizar los Registros tanto para recoger datos como para utilizarlos como fuente de datos en el análisis de datos secundarios.

Pero no debemos actuar a ciegas, asumir que el dato procedente de un Registro vale por sí mismo, porque es un número, una cifra cargada de magia cientísta, para evitar la fascinación se requiere un conocimiento profundo de la metodología con la que han sido concebidos los Registros en general y construido cada uno de ellos en particular<sup>26</sup>. Un conocimiento que, a la vez, nos conduce hasta una tercera opción: la creación de registros propios, lo más simples y sencillos posibles, pero también lo más fiables y válidos posibles.

La respuesta dependerá, en una gran medida, de la formación previa del evaluador, así el psicólogo estará muy interesado en conseguir que cada dato obtenido se corresponda con alguno de los códigos de protocolo, el sociólogo atenderá más bien al grado de representatividad del emisor, el antropólogo tratará de detectar aquellos datos que no se ajustan a las codificaciones pre-establecidas, el pedagogo y el educador focalizarán su interés sobre aspectos cognitivos del dato y cada profesional se centrará en aquellos aspectos que le hayan enseñado que son los relevantes. Pero todos ellos deberían tener claro que cuando se trata de intervenciones fijas o muy prolongadas o indefinidas, con poblaciones que cambian y evaluaciones periódicas pero continuas, el diseño de un Registro requiere mucha habilidad y abundante práctica metodológica (Barcikowski, 1994). Pero después el manejo es sencillo y además cada profesional podrá optar por la finalidad que le parezca más significativa, según sus intereses y formación.

Asimismo los Registros poseen ventajas para realizar meta-evaluaciones, ya que en la realidad institucional, como ha mostrado el trabajo de Durkheim sobre el suicidio, podemos integrar los datos sin problemas de muestra o de significación poblacional, de calidad o de ponderación de tamaño (Tober, 1994), también se pueden lograr estimaciones de incidencia muy fiables triangulando fuentes, pero este es un método que implica la existencia de tales fuentes y en general las fuentes útiles basadas en técnicas multiplicativas son los propios Registros (Harnoll, 1885), aunque también son posibles otras técnicas como captura/recaptura o la nominación. Es decir tanto si imaginamos una evaluación crítica y de orientación muy cualitativa como si imaginamos una evaluación más administrativa o de gestión, los Registros están ahí para trabajar sobre ellos.

---

26 Las reflexiones realizadas nos permiten pensar que los Registros son "una técnica de recogida de datos con consecuencias metodológicas" y a la vez "una producción social que merecería un especial interés sociológico" (Government Statisticians' Collective, 1993), pero eso no significa que haya que estar de acuerdo con autores como Julien Freund que confunde una técnica de recolección de datos con una teoría sociológica. Freund cita concretamente los modos de recoger los datos en los "intentos clasificadores" de Bacon y de D'Alambert, y lo mismo podría decir de Quetelet, Le Play o Marx, sin percatarse que estos intentos eran justamente en la práctica una teoría social (Freund, 1973), pero que desde la conformación del canon de las ciencias sociales las técnicas y la teorías se sitúan sobre dos planos bien distintos, interrelacionados pero diferentes.

En este sentido podemos establecer una cierta dualidad, entre aquellos que mitifican los registros como una forma muy objetiva de producir datos empíricos y aquellos que los consideran sólo una de las formas (en ocasiones demasiado compleja) de recopilación de datos y que en razón de su propia complejidad (en el diseño metodológico) tenemos que utilizar con precaución. Suponiendo para ambas opciones un criterio de neutralidad valorativa, los primeros se justifican en una gestión no-corporativa, y por tanto socialmente válida, de los instrumentos de evaluación y conocimiento de la realidad, mientras que los segundos se fundamentan sobre la trayectoria histórica de una sociología y una antropología críticas. En el fondo lo mismo<sup>27</sup> y una razón más para apostar por la noción de evaluación fuerte.

---

<sup>27</sup> Se ha dicho que *"el registro que pretende la medida de todas las cosas suscita a la vez fascinación y repulsión"*, pero el evaluador no debe dejarse llevar por ninguno de estos sentimientos sino plantearse dos problemas: el grado de validez de los datos y su utilidad (Boyer, 1994)

### 9.1.- Reconstruyendo con intención.

De forma tradicional el devenir socio-histórico se condensa en la técnica de las **historias de vida**. La técnica de las historias de vida aparece en muchas tradiciones académicas nacionales, con formas un tanto particulares, lo que ha generado una cierta confusión terminológica y una difícil delimitación conceptual, porque en ocasiones un mismo término se emplea para designar cosas muy distintas según las correspondientes tradiciones. Así, por ejemplo en España hay una importante tradición histórica de testimonios y autobiografías, que podemos relacionar con una cierta idea de la “confesión” católica, sin embargo, no ha sido hasta muy recientemente cuando han comenzado a publicarse textos que formalizan esta metodología que, hasta los años 80 y con alguna excepción como la Juan Francisco Marsal, no se vinculaban a los procedimientos de investigación social (Vallés, 1997).

Los enfoques generacionales siempre han priorizado esta técnica, pero hasta fechas muy recientes no se ha puesto en conexión el modelo tradicional de las biografías y la teoría de las generaciones (Comas, 2003).

En este sentido hay que comenzar diferenciando las **historias de vida** en sentido estricto, del uso de documentación o información secundaria sobre la vida y las experiencias de una persona, que pertenecerían al momento empírico de la utilización de los datos secundarios, mientras que las historias de vida suponen y exigen la interacción personal entre el informador y el investigador. Una historia de vida no es algo que se analiza a partir de una determinada documentación, sino algo que se reconstruye entre las dos partes (Sarabia, 1986).

El empleo de documentos personales para reconstruir una peripecia vital tiene una larga historia en la investigación en ciencias sociales, que se remonta al apogeo de la escuela de Chicago. Pero en tales casos no estamos ante verdaderas historias de vida, porque estas obras clásicas se basaron, bien en documentos personales o bien en recopilaciones de narraciones, sin que se produjera una interacción prolongada y mutuamente enriquecedora entre el investigador y el sujeto, cuya vida se iba a reconstruir. De forma simple deberíamos distinguir entre autobiografías, biografías, reconstrucción de casos e historias de vida (cuadro 9.1).

La idea de que una historia de vida no es sólo la biografía del informante, sino el resultado de la interacción entre el informante y el investigador, tiene que ver con la intención particular

Cuadro 9.1. **Clasificación de las narraciones.**

CONCEPTO	CONTENIDO
AUTOBIOGRAFÍA	<i>Visión personal de la propia vida</i>
BIOGRAFÍA	<i>Presentación crítica de una vida por parte de un investigador</i>
HISTORIA O ESTUDIO DE CASO	<i>Descripción, análisis o registro de un evento o proceso social</i>
HISTORIA DE VIDA	<i>Reconstrucción de una vida a partir de entrevistas</i>

de captar el devenir socio-histórico. Ciertamente, una autobiografía elaborada individualmente también puede recoger elementos de este devenir, pero la gran diferencia se refiere a que el momento empírico de la historia de vida capta a la vez la totalidad, la ambigüedad, el cambio y la visión subjetiva de los acontecimientos. De alguna manera desvela el proceso más allá de lo que pueden imaginar sus protagonistas, pero a la vez sólo ellos mismos nos pueden ofrecer esta totalidad, pero sometidos a una determinada dinámica, en la que no nos limitamos a recoger los datos como haría un observador, sino que intentamos rehacer una descripción verosímil de la vida en términos de acontecimientos sociales relevantes (Ruiz de Olabuénaga, 1996). La técnica de las historias de vida se inició en Estados Unidos al calor de varias circunstancias que contribuyeron a su impulso. Quizá la más importante fue la llegada a un país en rápido desarrollo, de masas de emigrantes europeos que debieron adaptarse a unas circunstancias de vida cotidiana y de valores morales y éticos distintos a los recibidos en sus países de origen. Este proceso provocó importantes tensiones sociales que despertaron el interés público y el de los científicos sociales (Sarabia, 1986). Se trata de un proceso muy similar al que actualmente se vive en España.

En Europa, los investigadores sociales no descubren la historia de vida como una posible práctica empírica hasta mucho más tarde. Aunque en este caso el origen es muy diferente, ya que no se trata de investigar determinados problemas sociales, sino de desarrollar los métodos cuantitativos de la sociología. En los años sesenta, Daniel Bertaux se propone, casi de forma clandestina, aplicar la historia de vida a la investigación social (Pujadas, 1992). En el marco de una investigación sobre “*Los determinantes de la movilidad social*”, Bertaux emprende una crítica metodológica de la historia de vida clásica, poniendo a prueba esta técnica de recolección de datos estudiando la supervivencia de la panadería artesanal frente a la panadería industrial.

Finalmente, en los últimos quince años, en el ámbito de los Estudios Culturales, este método de recolección de datos ha conocido un desarrollo espectacular, que podemos atribuir a varios factores. De una parte, la propia temática de los Estudios Culturales, situados en la línea del conflicto de género, étnico y cultural, en la desviación y la marginación; y de otra, la estrategia de dar voz a los investigados y que se conviertan en protagonistas de su propia historia. Se trata en todo caso de una voz relativa, porque el investigador **selecciona** aquello que dicen y lo adecua a sus propios objetivos. No falsifica, se limita a elegir aquello que es verdaderamente relevante al margen de lo que crea el sujeto. En este sentido la voz del sujeto aparece, como en la encuesta, mediada por el investigador.

Las primeras historias de vida producidas en España, y cuya elaboración conocí de primera mano, reflejan muy claramente la idea de “*recoger sólo los acontecimientos relevantes al objeto de estudio*” (Romaní, 1983; Funes y Romaní, 1985). En la primera de ellas, Oriol Romaní mostraba cómo sujetos situados en diferentes puntos del sistema de estratificación social consumían cannabis desde perspectivas culturales antagonistas y percibiendo efectos muy distintos para la misma sustancia. Pero la proyección estructural no era relevante, porque la presencia del can-

nabis en cada uno de los estratos sociales se debía no tanto a diferencias de clase, sino que respondía a la génesis histórica particular, a la forma en que se introdujo y comenzó a utilizarse esta sustancia en cada uno de los estratos o subculturas. En el análisis posterior el investigador podía manejar también hipótesis sobre la influencia de la estructura de estratificación en la producción de estos fenómenos particulares, pero fue la historia de vida la que le permitió captar la **particularidad temporal** de algo que era percibido sólo como subcultural.

En la segunda de las investigaciones ocurrió algo similar, los exadictos a la heroína contaban una historia cuya reconstrucción llevó a determinar la presencia común de un acontecimiento traumático que, combinado con una figura de referencia positiva, explicaba porqué sólo ellos y no otros adictos eran capaces de “abandonar” el consumo de drogas en la España a principios de los años 80. Una vez detectado cuál era este acontecimiento, la historia se centraba en el mismo.

## **9.2.- Produciendo historias de vida.**

Una historia de vida es un relato autobiográfico, obtenido por un investigador mediante entrevistas sucesivas. El principal objetivo es el de obtener el testimonio subjetivo de una persona en la que se recojan tanto los acontecimientos, como las valoraciones que esta misma persona hace del conjunto o de una parte de su trayectoria vital. El método supone que en ocasiones la investigación se realiza con una sola historia y en otras con varias complementarias que se refuerzan o contrastan. En la historia de vida, el investigador no se limita a transcribir lo que le cuentan, sino que induce la narración, la ordena, controla su coherencia interna, su ajuste con los acontecimientos históricos y las dinámicas sociales y culturales, revisa a fondo determinados acontecimientos e incluso los contrasta con fuentes documentales, buscando la comparación y la triangulación (Sartori y Morlino, 1991).

En este sentido la historia de vida es una metodología muy útil para visualizar los cambios sociales y puede tratarse, por tanto, de un tipo de investigación longitudinal (Castilla, 1998). Como consecuencia, la elaboración de una historia de vida requiere previamente realizar un diseño estricto de la investigación que resuelva los problemas de representatividad, validez, y fiabilidad de la investigación. Eso supone al menos que hay que:

1. Elaborar un planteamiento teórico del trabajo que explicita claramente cuáles son las hipótesis de trabajo iniciales.
2. Justificar metodológicamente el porqué de la elección del método biográfico.
3. Delimitar con la mayor precisión posible el universo objeto del análisis (familia, comunidad, grupo profesional, de edad, colectivo inmigrantes...).
4. Explicitar los criterios de selección del o de los informantes cuyas historias se van a recoger, que en general deben responder a dos criterios: el primero, la cantidad de información contenida en la narración, y el segundo, el rol social del narrador en relación al objetivo de la evaluación (Johnson, 1990).

La literatura disponible (Pujadas, 1992; Vallés, 1996), permite sostener que los casos en los que el uso de historias de vida resulta más pertinente se refieren al análisis y descripción de los procesos de desajuste y crisis individual o colectiva, que presuponen modificaciones significativas tanto en el comportamiento, como en los sistemas de valores por parte de los grupos sociales implicados. Quizás por ello los dos temas más frecuentes a los que se aplica en este tipo de investigación son los procesos migratorios y los de desviación o marginalización. El rasgo común que une ambos procesos es que se trata de analizar poblaciones e individuos que se encuentran en tránsito de una categoría social a otra, de un marco social a otro. En tales casos la historia de vida recoge el trayecto entre ambas situaciones, por ejemplo, campo/ciudad o campesino/obrero, al conjugar la perspectiva reconstruida de los protagonistas con los acontecimientos sociales en los que éstos se desenvuelven.

Junto a estos dos grandes temas tradicionales, se puede aplicar esta técnica prácticamente a la evaluación de cualquier programa, como por ejemplo la evolución de las relaciones entre géneros a partir de un programa de igualdad; el cambio en el ritmo de la emancipación a partir de programas de empleo y vivienda; las trayectorias vitales de los minusválidos tras un programa de accesibilidad; el cambio de valores tras un programa de movilidad o la conformación del liderazgo político o asociativo como consecuencia del desarrollo de un programa de apoyo a la participación; así como cualquier otro tema que permita ofrecer una visión diacrónica y holística del mismo.

A pesar de todo, la producción de historias de vida ni es fácil, ni es frecuente, ya que por una parte es una técnica que exige mucho tiempo y dedicación personal y por otro, a pesar de la reciente publicación de diversos manuales que explican con precisión esta técnica, no hay un modelo canónicamente aceptado acerca de cómo trabajar con historias de vida. Ello obliga a formarse leyendo monografías y a utilizar modelos que, como el análisis hermenéutico, puede ser sólo adecuado a un determinado tipo de historia (Riessman, 1993). En cualquier caso, sea cual sea el modelo utilizado, al menos hay que dar los siguientes pasos.

1. Localizar y seleccionar aquellas experiencias que pueden ser utilizables.
2. Atender a los relatores, interactuar con ellos y registrar los relatos.
3. Transformar los relatos en textos.
4. Analizar los textos desde algún tipo de análisis de datos y desde alguna perspectiva teórica concreta.
5. Transformar los resultados obtenidos en representaciones sociales, proyectar las representaciones sobre la realidad social y reconstruir la peripecia vital.

### 9.3.- La evaluación y los relatos.

Desde la perspectiva de la evaluación, la ventaja de la historia de vida, que es también la ventaja de todas las investigaciones que establecen series (a partir de cualquier método) y la de los análisis comparativos entre evaluaciones realizadas en diferentes periodos, se refiere a que la visión diacrónica de los acontecimientos proporciona un tipo de información a la que no se puede acceder o no se puede percibir cuando los datos se refieren a un periodo muy delimitado.

Esta ventaja se convierte en un inconveniente si consideramos que la mayor parte de las intervenciones sociales (e incluso las políticas) poseen un horizonte temporal muy limitado. En tal caso la narración histórica resulta imposible, porque no hablamos de una larga serie de acontecimiento situados a lo largo del tiempo, sino de un acontecimiento único (o unos pocos) situados sobre un periodo muy limitado.

Para percibir el marco de relaciones entre la evaluación y las historias de vida podemos utilizar el contenido del texto sobre “Las Políticas de Juventud en la España democrática”, que se ha publicado en fechas recientes (Comas, 2007). En dicho libro señalábamos la necesidad de realizar algunas **investigaciones** históricas sobre las Políticas de Juventud de la dictadura, pero más adelante indicábamos también que se deberían realizar **evaluaciones** sobre “las políticas de participación y el papel de los agentes sociales en las políticas de juventud de la democracia”. ¿Cómo podemos hacer esto? Pues reuniendo datos a partir de estadísticas y realizando un análisis comparado de los diversos estudios sobre participación realizados en el periodo democrático. Pero una de las maneras de hacer una buena evaluación sobre estas políticas de participación sería reunir un número suficiente de historias de vida de protagonistas del asociacionismo juvenil desde los primeros ayuntamientos democráticos hasta aquellos que por cumplir treinta años los abandonaron en el cambio de siglo. Sería una evaluación que se referiría a un cuarto de siglo, pero que nos ayudaría mucho a mejorar las actuales políticas de participación.

### **10.1.- La compleja cuestión de la mediación social.**

La cuestión de la mediación social contiene crecientes niveles de confusión conceptual, en una gran medida porque desde distintas disciplinas y áreas de conocimiento y desde diferentes teorías se está utilizando el mismo término para referirse a cosas muy diferentes. Por este motivo conviene clarificar algunos aspectos del concepto, porque la cuestión de los informantes clave y los informadores cualificados está íntimamente ligada a la cuestión de la mediación social. Como técnica de recogida de datos para la evaluación el uso de informantes e informadores tiene mucho futuro si aprendemos a relacionarla de una forma conveniente con la mediación.

En cuanto a la confusión terminológica, de una parte, por “mediación” se entiende una estrategia de “resolución de conflictos” de la que se derivan muy variadas ramas, desde la muy formalizada “mediación familiar” hasta la llamada “mediación en conflictos políticos” con dos propuestas bien diferentes, la que responden a la llamada “mediación psico-cultural” de Marc Howard Ross y la que responde a las diferentes “estrategias de negociación” en Ciencias Políticas. Entre ambas formas de entender lo que es la mediación se sitúan otras prácticas con una orientación psico-social (laboral, de género e igualdad, étnica y la muy reciente “mediación ecológica”) con grados de desarrollo e implantación muy diferentes.

Por su parte Manuel Martín Serrano ha popularizado (y ampliado) en España la llamada “Teoría de la Mediación Social” que se aplica al territorio de la información y la comunicación y que trata de explicar como se difunden y se imponen los diferentes modelos culturales, como se internalizan en los sujetos y como conforman finalmente los valores sociales (Martín Serrano, 1978, Martín Serrano, 1986). En esta perspectiva el Medio de Comunicación es el mediador social por excelencia ya que produce potentes efectos de persuasión sobre la conciencia de los sujetos. Muy ligada a esta teoría aparecen las diversas visiones críticas que vinculan la mediación a la publicidad y el consumo y que he reseñado en un reciente texto (Comas, 2008).

Por si esto fuera poco la irrupción de la obra de Lev Vygotsky ha desvelado, en el campo de la psicología, lo que ha venido en llamarse también “mediación social” para referirse a los procesos de aprendizaje en los que intervienen unos “instrumentos” psicológicos llamados “mediadores” (signos y herramientas), resultado de la acumulación cultural, la cual supone un “medio” (en el sentido de que media con las características biológicas de los sujetos). El “constructivismo” de Vygotsky tiene muchos elementos en común con la teoría del constructivismo social y desarrolla contenidos muy similares a los llamados Estudios Culturales y muy especialmente en el terreno de los Estudios y las teorías de Género, en todos ellos está impulsando la utilización de esta noción de “mediación social”.

Es cierto que toda esta literatura tiene en común el reconocimiento de la importancia de lo social para configurar los procesos individuales y en este sentido está influyendo de manera notable en los programas y en las políticas de intervención social. Como consecuencia en muchos programas sociales, y en el caso del área de juventud tanto las llamadas “políticas afirmativas”, como las “políticas participativas”, se va adoptando cada vez más, con diferentes matices, la perspectiva de la mediación social, para trabajar con los jóvenes, para ayudarles en su proceso de maduración, hacerles más activos y ayudarles a tomar decisiones bien informadas. Se trata de objetivos, todos ellos, difíciles de evaluar.

Por tanto debemos pensar en tipos de la evaluación que nos permitan construir indicadores y recoger datos en torno al impacto de los programas basados en “acciones de mediación”.

Para poder hacerlo parece conveniente volver al concepto original de mediación (y mediador) social tal y como fue formulado por Paul Lazarsfeld, Bernard Berelson y Hazle Gaudet, al estudiar las elecciones norteamericanas de 1940 (es uno de los “*experimentos cruciales*” del siglo XX) y publicado unos años después bajo el título de “*The People Choice*” (Lazarsfeld, 1944). En dicho trabajo se demuestra que en la decisión de votar a uno o a otro partido candidato, indicadores como la categoría económica, la filiación religiosa y el lugar de residencia desempeñan un papel muy importante, que puede utilizarse para ofrecer un *primer índice de predisposición política* (IPP) en el que cabía clasificar a la gente, ya que estas características sociales marcan la probabilidad de que votaran por uno u otro partido

Pero esta previsibilidad política, que se traduce en una “teoría del efecto nulo” y establece un cierto grado de determinismo electoral, también incluye la noción de “doble escalón”. Así mientras la teoría del efecto nulo demuestra que la propaganda política no influye sobre la voluntad de los individuos, es decir, no la modela ni la construye, porque ya está pre-establecida y la persuasión política, encuentra su límite en estas disposiciones previas que constituyen una verdadera “mediación social” (en el sentido que ha popularizado la obra de Vygotsky), una mediación biográfica, social y cultural que determina su identidad ideológica. La paradoja es que se trata de una mediación que les determina, pero a la vez los convierte en sujetos autónomos a la hora de tomar sus decisiones, porque la publicidad electoral no les influye fácilmente.

Pero la noción de “doble escalón” nos dice que aunque la mayor parte de los sujetos no toman decisiones de voto persuadidos por la campaña electoral (y otras iniciativas o intervenciones sociales), algunos (muy pocos) pueden cambiar su decisión como consecuencia de la misma. Ocurre que los que “escapan” de la mediación social son los únicos “mediadores” posibles, porque son los únicos que “traducen”, con razonamientos aceptables para el discurso pre-establecido por la mediación social, la explicación del cambio. Estos mediadores son los que influyen en los demás por la credibilidad previa de su ajuste al discurso mediado. Es decir, convierten los contenidos de la mediación social previa en los argumentos del cambio de orientación. Estos sujetos son “mediadores” porque se les reconoce su capacidad para modificar las opiniones de los demás y forman un necesario “doble escalón” para acceder a la persuasión. Como consecuencia una elección política (y cualquier otro acto social), se construye en esta relación entre lo social y lo individual. Sabemos la decisión que va a tomar el sujeto en función de ciertas variables (que forman parte de su proceso de socialización), pero a la vez si varían las

creencias o los valores de sujetos muy ajustados a la condición social más estándar, estos influyen en muchos otros sujetos. Por esto “The People Choice” se orienta de una parte hacia la identificación de los factores de mediación social productores de decisiones políticas, pero de otra parte muestra el interés por localizar a los “mediadores” que siendo los más ajustados a estos factores, son a la vez los únicos que, de ser persuadidos, pueden modificar la actitud de otros ciudadanos. Otra cuestión clave que detectaron los autores del famoso estudio se refiere al hecho de que la conversión del comportamiento de los electores pasa, de forma inesperada, por los votantes con fuertes y definidas identidades políticas. Estos electores altamente identificados con organizaciones políticas son los que tienen orígenes sociales más ajustados a cada una de las propuestas políticas. Cambiar su decisión no es fácil ni frecuente. Pero un número reducido de cambios de opinión entre ellos puede producir un efecto dominó muy importante sobre el resto de los votantes.

Esta es la lógica que subyace bajo los programas basados en las metodologías de Bolas de Nieve (Arenas, 2006; ONU-SIDA, 2000; García, G., Gutiérrez, E. y Morante, I. (2004), García, G., 1998; García, G. 2002). Se trata de captar a Agentes con un intenso reconocimiento social entre sus pares, lo que suele coincidir con sujetos muy ajustados a sus “mediaciones”.

En estas metodologías se recurre, en los ejemplos citados, a algunos presos con prestigio y un profundo conocimiento de la cultura carcelaria para convertirlos en agentes de salud muy efectivos, o en otro de los ejemplos citados, se recurre a adictos a drogas con una intensa trayectoria de consumo, es decir, sujetos muy mediados por su entorno “próximo” (una palabra de Vygotsky) para desarrollar acciones de “reducción de daños”. Captarlos supone facilitarles algunas informaciones y otros aprendizajes, congruentes con su discurso previo y que les parezcan creíbles y útiles. Son después ellos los que, de manera natural, difunden estas informaciones y modifican los hábitos de otros sujetos mediados por esta cultura próxima.

He elegido ejemplos un tanto extremos (o al menos que no suelen aparecer en los ámbitos de juventud) por razones didácticas, pero de hecho es el mismo trabajo de los “mediadores juveniles” cuando, por ejemplo, consiguen convocar a jóvenes para que se afilien en asociaciones juveniles, participen en determinadas actividades, acudan o eventos o apoyen causas. Un buen mediador juvenil no es nunca un joven muy militante, muy activo o muy preparado para hacer de “mediador”, sino un chico o una chica, “normal”, muy tópico y típico, aunque bien valorado en su entorno, que no tiene especial predisposición hacia una tarea, pero que de pronto dice a sus pares: ¿Y si nos organizamos para hacer X?.

La pregunta es ¿Cómo podemos evaluar, en la perspectiva de la Evaluación Fuerte, programas basados en mediadores sociales? Pues de diferentes maneras moviéndonos con agilidad en la tabla redonda de los métodos y las técnicas de recogida y de producción de datos. Aunque las técnicas basadas en Informantes e informadores parecen las más adecuadas.

## **10.2.- Los informantes clave.**

La historia de la técnica de recogida de datos mediante informantes clave es muy sencilla porque se trata de un procedimiento metodológico con una única complicación técnica: los criterios de selección de los informantes. Y esta complicación apenas se ha presentado a lo largo de la historia de dicha metodología porque en los casos en los que se ha aplicado el medio ha determinado una respuesta automática: Los informadores clave dispuestos a ejercer este rol era pocos, su motivación difícil y por tanto la selección imposible.

La metodología se inventó, al mismo tiempo que la Observación Participante, y sin ningún antecedente previo, gracias al conjunto de casualidades que llevaron a Bronislaw Malinowski a permanecer aislado en las islas Trobriand. Los Informantes Clave eran aquellos nativos que le hablaban y que le contaban cosas sobre aquella cultura. No se trataba de los jefes, ni de la gen-

te especialmente preparada, sino de sujetos observadores, curiosos, empáticos y dispuestos a compartir sus conocimientos. En todo caso conviene tener en cuenta que la idea de contar con el conocimiento y las informaciones proporcionadas por “este tipo de gente” se ha desarrollado en una cierta tensión entre la noción de “gente corriente pero interesa en transmitir información” y “gente que, por su rol social, accede a información privilegiada”. La tensión entre ambas opciones casi siempre se ha resuelto por la vía de las posibilidades, es decir, los estudios basados en Informantes Clave son siempre investigaciones ubicadas en lugares sociales muy periféricos, en los que otras metodologías son difíciles de aplicar y en los cuales tampoco abundan los informadores con muchas destrezas y que quieran colaborar. La historia de los trabajos con Informantes Clave es la historia de soluciones imaginativas a dificultades para el trabajo de campo (Cresswell y Godelier, 1975; Foster, 1974; Bastide, 1971).

En los últimos años, por la vía de la llamada “antropología urbana” el tema de los Informantes Clave se ha dirigido hacia la cuestión de la “selección de buenos Informantes” en un medio social urbano y “normalizado”. Una opción que, en mi opinión, ha acabado por subsumirse con mayor eficacia en otras técnicas como la “entrevista en profundidad”, la “historia de vida” y el método Delphi que se explica a continuación.

Se puede, sin embargo, utilizar esta metodología en los ambientes sociales y culturales de difícil penetración, desde minorías étnicas hasta grupos juveniles con subculturas muy poco convencionales, en las que, para lograr algún tipo de efecto se requiere la presencia de mediadores sociales.

Evaluar un programa o una acción en estos ambientes puede resultar complicado y utilizar Informantes Clave la única opción posible y por tanto la más recomendable. Lo que no parece muy lógico es recurrir por “razones de preferencia conceptual” o por determinadas creencias, a una metodología que tiene opciones alternativas más prácticas en poblaciones normalizadas.

En todo caso para usar esta metodología (que no está muy bien descrita en la mayoría de manuales) hay que tener en cuenta algunos hechos.

1. La condición de Informante Clave requiere un cierto compromiso por parte del mismo y del evaluador. Ambos deben asumir que la suya es una relación larga y compleja en la que lo personal y lo institucional pueden superponerse con facilidad. Es decir, el vínculo entre el informante clave y el evaluador no es el de una entrevista, ni una serie de entrevistas, sino el de una “relación” más o menos prolongada.
2. Por este motivo a los Informantes Clave no se les “entrevista” sino que “se habla” con ellos, aunque se puedan establecer Guiones sobre la información deseable y registrar las informaciones que proporcionan.
3. La relación con los Informantes es un proceso abierto en el que ambas partes aportan información y obtienen beneficios. El Informante realiza su tarea, una tarea que implica una cierta dedicación, porque recibe algo a cambio, por ejemplo información sobre el programa o la acción que esta siendo evaluada. En un momento determinado el Informante puede “participar” en el proceso del programa. Conviene señalar que se ha producido mucha discusión sobre la cuestión de “pagar” a los Informantes Clave y no es un asunto resuelto.
4. El mejor informante es aquel que cumple las condiciones señaladas por Lazarsfeld para ser un “mediador social”, es decir un sujeto profundamente mediado por su entorno, bien informado pero a la vez abierto a modificar sus opiniones. De hecho ser Informante Clave implica traicionar fuertes creencias para defender opciones que mejoran la vida del colectivo.

### 10.3.- El método Delphi.

El método Delphi representa, en la actualidad, la tercera metodología en volumen, tras las encuestas y los grupos de discusión, para la industria de “investigaciones y estudios”. Hace

apenas 20 años sólo lo utilizaban en España los sociólogos que hacían “prospectiva de futuro”, más tarde se comenzó a utilizar en los “estudios de mercado” con la idea de hacer “prospectivas de mercado y tomar decisiones”, posteriormente el sector financiero lo adoptó como una metodología para “evitar la incertidumbre” y finalmente ha llegado al área sanitaria que lo utiliza como “técnica de consenso” para establecer protocolos, planificar acciones y servicios e incluso elegir itinerarios terapéuticos.

El nombre de método Delphi se tomó del Oráculo de Delphos, que se suponía predecía el futuro de aquellos que preguntaban sobre el mismo. Sin embargo, el método fue puesto a punto por la Rand Corporation en Estados Unidos en 1959. El argumento de partida fue que “cuando no es posible disponer de conocimientos científicos sobre una materia el juicio de los expertos es la mejor manera de aproximarse a la realidad”. Por este motivo algunos le llaman panel de expertos. La mayor parte de los Delphi se dedican a hacer previsores de futuro. De hecho el primero de los que realizó la Rand Corporación trataba sobre “las consecuencias de la Guerra Nuclear”. Pero también se puede utilizar para algo que ocurrió en el pasado (u ocurre en el presente) y en este sentido es útil a la evaluación.

En síntesis el método Delphi consiste en someter a un grupo de expertos o especialistas a que respondan dos o más formularios sucesivos destinados a recolectar sus opiniones y visiones sobre el tema en cuestión. Cada vez que se rellena el formulario se denomina ronda. Uno de los objetivos básicos de un Delphi consiste en el logro de pronósticos que sean el producto de consensos lo más sólidos posibles. La manera para alcanzar este consenso supone que, en las rondas posteriores a la primera ronda, cada experto que ha realizado pronósticos que se desvían en algún grado de los pronósticos más consensuados tenga la posibilidad de modificar su respuesta. Para ello se le presentan los resultados de la pregunta específica en que ha existido tal desviación, y se le propone que la reconsidere si encuentra argumentos razonables para ello. De lo contrario, debe ratificar su respuesta anterior.

El resultado final del Delphi es un conjunto de pronósticos más o menos consensuados sobre los diferentes temas y aspectos presentes en el cuestionario, a lo que se añaden las innovaciones producto de las respuestas a preguntas abiertas. Dicha información, materializada en una base de datos susceptible de análisis, permite la construcción de un escenario muy probable sobre lo que ha ocurrido o puede ocurrir (Landeta, 1999, Ruiz Olabuénaga e Ispizua, 1989).

En la segunda o sucesivas ronda del Delphi, aquellos expertos que en al menos una de las preguntas se hubiesen desviado de las respuestas más consensuadas podrán replantearse las mismas, proponer una nueva respuesta o insistir en la anterior. En este caso suele exigírseles una argumentación en torno a su posición. En algunos Delphi la información es absolutamente confidencial, mientras que en otros casos se trata de abrir un debate público o entre los expertos. En ambos casos la condición (público o confidencial) se establece de forma previa con claridad.

Los expertos participantes lo hacen por dos razones, bien por su reconocimiento como expertos o bien porque reciben una compensación económica por un trabajo que en ocasiones es muy pesado. En los últimos años al popularizarse el método, algunos “expertos” reciben frecuentes invitaciones a participar en Delphi (en mi caso y en el año 2007 han sido cinco), la mayoría me ha reportado información sobre por donde van los consensos (en otro he discrepado radicalmente de un grupo de conclusiones por lo que vamos a realizar un debate público sobre las mismas), pero en algún caso me ha resultado pesado, reiterativo y poco motivador. En este último caso he contestado forzando mi grado de consenso para evitarme trabajo posterior.

Algunos detalles complementarios sobre el método.

1. La selección de los expertos es algo complejo a lo que se presta poca atención, porque suelen ser las figuras más conocidas de un tema o un área de conocimiento. Una búsqueda más sistemática de “expertos” siempre será más difícil y no siempre garantiza mejores resultados. Existe la tendencia natural a “pasar” de los nombres conocidos y utilizar

personas que se supone conocen sobre el tema pero sus conocimientos no están tan reconocidos. Esto en ocasiones produce sonoros fracasos porque un informador experto es siempre un “mediador reconocido” y la única forma de establecer este reconocimiento es justamente el “reconocimiento público de su mérito”. También es cierto que los que poseen un estatus de “expertos” y son reconocidos como tales (y se conocen entre ellos) tienden a establecer consensos corporativos o interesados.

2. Se requiere establecer, una vez realizada la selección, el compromiso previo de colaboración del informador experto, para hacerlo se expone de una forma muy detallada cual va a ser todo el proceso (y sus condiciones) y el informador debe aceptar formalmente su participación en el mismo. A pesar de ello los abandonos son frecuentes.
3. Se requiere la presencia de un “moderador” (o equipo de moderación), que es quien prepara los cuestionarios, recibe las respuestas y establece los resultados de consenso para las sucesivas rondas. El control administrativo de un Delphi es en ocasiones algo pesado, aunque ahora las TIC facilitan la tarea (es tan fácil como crear un grupo de correo y enviar los cuestionarios como archivos sobre programas que permitan re-escribir). Han comenzado las experiencias con Foros e incluso con Messenger, aunque en este caso se corre el riesgo de dejar fuera a los expertos que no dominen estas tecnologías.
4. El cuestionario puede contener preguntas abiertas, que habilitan un espacio para contestar libremente la respuesta que se juzgue más adecuada, y las preguntas cerradas, pre-codificadas en ítems definidos. Algunos cuestionarios se construyen a partir de escalas. Cuanto más cerrado es el cuestionario más fácil es el trabajo de establecer los consensos, aunque esto obliga a un mayor trabajo previo y la información puede perder matices.
5. Se puede establecer una prueba del cuestionario con un grupo experimental reducido.
6. En el informe final los investigadores (o en este caso los evaluadores) no deben incluir sus propias opiniones y el análisis se limita a facilitar un relato de los consensos establecidos y si acaso a formular o describir los desacuerdos y sus respectivas argumentaciones.

#### **10.4.- Evaluación, mediación e influencia.**

El territorio metodológico de los informantes y los informadores es a la vez especialmente útil a la noción de evaluación fuerte, pero también especialmente peligroso. Porque se puede conseguir información muy válida (y que no se podría conseguir por otros medios) pero también puede ser profundamente manipulado.

Un informante y un informador es una persona (o un grupo organizado de personas en el caso del método Delphi) situado sobre un estrecho filo en el que no es fácil mantenerse. Su propia condición de mediador/mediado (en el sentido de Lazarsfeld) le puede conducir hacia el lado del evaluador o hacia el lado de sujeto que ha sido objeto de la intervención. En el caso del Informante Clave puede parecer que aporta información al evaluador pero quizá esté tratando de engañarle en función de los intereses de su grupo. Así en los análisis diagnósticos sobre grupos marginales es muy frecuente que los informantes exageren los problemas del colectivo para conseguir que el evaluador se posicione a favor de reclamar más recursos públicos para el mismo. A la inversa un Informante puede realizar denuncias muy comprometidas sobre corrupción en el manejo interno de estos mismos recursos por razones muy personales. En ambos casos estamos ante un buen Informante que “controla lo que pasa”.

En mi experiencia con métodos Delphi (tanto como investigador como, últimamente, como informador), no puedo dejar de reconocer que los consensos siempre se alcanzan en torno a ciertos intereses corporativos o de clase social. Por ejemplo si preguntamos a un grupo profesional sobre un problema que no les interesa demasiado o sobre el que no tienen respuesta, es fácil que el consenso se alcance afirmando que el problema no existe o afecta a muy pocos, en

cambio si el problema interesa (o permite mejorar los ingresos salariales) el consenso se alcanza reflejando la “gravedad” del mismo.

Todos estos peligros pueden evitarse (o al menos reducirse) reconociendo su existencia y colocándolos en el contexto de la evaluación fuerte que venimos describiendo: los informantes y los informadores nos proporcionan muchos datos y también pueden manipularnos. Sabemos que es así. Pero a la evaluación fuerte no le importa, porque ella evalúa la realidad de los programas sociales y en un momento determinado las diversas actitudes de los informantes e informadores pueden constituir una evidencia muy relevante para mejorar el programa. En todo caso hay que tener mucho cuidado para no dejarse influenciar por intereses particulares que acaben por redundar negativamente sobre la intervención.



## La encuesta: la organización industrial y tecnológica

### 11.1.- La legitimidad social de la encuesta.

Avanzamos hacia aquel nivel de aproximación empírica a la realidad con una mayor presencia pública, la cual podemos, en parte, explicar por su formato constitutivo, es decir, la **representatividad** expresada por medio de la precisión estadística. En realidad, la encuesta no ocupa este lugar por sus cualidades intrínsecas, de la misma forma que la terapia individual, el periodismo de investigación, la cirugía, las tradiciones artísticas y el diseño arquitectónico, tampoco ocupan los lugares relevantes dentro de sus disciplinas particulares exclusivamente por sus propias cualidades. Están ahí porque la sociedad legitima estas prácticas, al tiempo que todas ellas comparten un cierto rechazo público (recordemos el éxito de la conocida frase-leyenda sobre las estadísticas y las mentiras recogida en el capítulo 6), o al menos una actitud de burla por parte de esta misma sociedad.

Por tanto, la encuesta ocupa un lugar socialmente relevante, pero en términos metodológicos es sólo una técnica de recogida de datos equivalente a las demás. En cuanto al lugar social que ocupa la encuesta en la sociedad, éste tiene que ver con la expectativa de la búsqueda de un reflejo en un espejo creíble para esta misma sociedad<sup>28</sup>. Un espejo que devuelve una imagen a la sociedad, que en ocasiones se reconoce en la misma (y entonces la encuesta es fiable), pero en otras ocasiones devuelve una imagen poco agradecida que produce o bien rechazo o bien la mirada oblicua del olvido. Cada mañana la sociedad desayuna con una encuesta para, antes de salir a la calle, saber si todo está en orden. De tal manera que en el año 1975 una mayoría de españoles consideraba que la democracia no era un sistema político demasiado adecuado para España y dos años después, en 1977, las virtudes incomparables de la democracia gozaban de una favorable unanimidad (Ortí, 1986). La encuesta es para la sociedad el espejo del deber ser, pero a la vez, para los investigadores sociales y en particular los evaluadores, es ésta una técnica de la que no podemos prescindir porque es la única

<sup>28</sup> Las otras técnicas de recogida y producción de datos que se exponen en esta parte del Manual no gozan de este estatus de credibilidad (y de rechazo) público, quizá porque "sólo" son propias del ámbito de las ciencias sociales (como si todas ellas, incluida la encuesta, no fueran exclusivas de dicho ámbito). Pero la encuesta recibe el parabién de otros actores sociales que le han otorgado su actual estatus, en particular del resto de los científicos, que la asumen en ocasiones de una forma muy poco crítica y muy inocente; de los Medios de Comunicación, que le pueden otorgar a la vez el estatus de fiable e incierta según les convenga; del ámbito político, que le ha otorgado un estatus, a la vez, de fuente de información e instrumento electoral; y del ámbito empresarial, que planifica sus acciones a partir de las mismas, y a las que puede atribuir las culpas de los fracasos.

que nos garantiza tanto la generalización de los datos y resultados, como el reconocimiento de la condición de espejo verdadero<sup>29</sup>.

En el caso de la evaluación la encuesta es una técnica imprescindible, no sólo por la credibilidad social de la misma (y el riesgo del rechazo cuando los datos obtenidos no resultan aceptables), sino por el problema de la representatividad cuando las poblaciones a interrogar tienen un tamaño demasiado grande para utilizar otros métodos. También es cierto que la representatividad no es una característica exclusiva del método, ya que se pueden aplicar criterios muestrales a otras técnicas de recogida de datos, pero, hoy por hoy, la mayor garantía de representatividad se liga a la encuesta.

Pero para entender mejor tanto el lugar social que ocupa la encuesta como la imposibilidad de prescindir de la misma, en la lógica de la legitimidad y la legitimación, voy a referirme a cuatro cuestiones. Primero sus orígenes ideológicos, desde los que se puede entender el poderoso marchamo con el que ha imprimado a la sociedad; segundo su alto grado de formalización, que permite una producción industrial; tercero su plasticidad tecnológica, expresada en la capacidad para adoptar estrategias de entrevistas muy distintas entre sí; y cuarto su utilidad al margen de los problemas de validez que pueda presentar.

## 11.2.- Los orígenes de la encuesta representativa.

El deseo de conocer las tendencias de la opinión pública ha sido una de las obsesiones permanentes de los dirigentes políticos y religiosos a lo largo de la historia. Este es un deseo que se manifiesta en numerosos ejemplos, aunque la mayoría reflejen un deseo insatisfecho (Stoetzel y Girard, 1973). Sólo en el mundo contemporáneo, tanto en los países democráticos como en las dictaduras siempre que el tipo de sociedad fuera lo suficientemente avanzada, este deseo se ha convertido en exigencia y la exigencia en una realidad tecnológica que se ha incorporado a la propia acción política, social y económica.

La técnica de la encuesta no fue teóricamente desarrollada con propósitos científicos, sino por razones administrativas y compasivas. Los orígenes administrativos de la encuesta se solapan con los orígenes de la estadística, con nombres como Quételet y Le Play, pero los manuales suelen citar diferentes encuestas como las pioneras, como por ejemplo la que realizó Engels en 1845 en Manchester sobre las condiciones de trabajo de los obreros de una fábrica. En todo caso, las supuestas encuestas pioneras no son tales, porque realmente no se produjeron auténticas encuestas hasta que no apareció una determinada cultura pública, que no comenzó a manifestarse hasta finalizar la Primera Guerra Mundial en la Europa más desarrollada y en Estados Unidos, aunque después en apenas unos decenios se generalizó al mundo entero. Los trabajos previos a la utilización de las muestras fueron sobre todo de tipo estadístico y demográfico, se limitaba a la recolección de datos sobre las poblaciones de un modo similar a los censos.

Así, el *movimiento de la encuesta social* en Gran Bretaña pretendía comprender las consecuencias de la revolución industrial, para intentar soslayar los efectos negativos mediante reformas adecuadas en ámbitos como las penitenciarias, la marina, las fábricas, las minas o los orfanatos. El movimiento tiene que ver con la encuesta, porque *para reformar* se suponía que previamente había que *cuantificar* la realidad, lo que condujo al movimiento a la utilización sistemática de las técnicas estadísticas, por parte de reformistas y filántropos. Algunos trabajos del movimiento

---

<sup>29</sup> La creciente legitimación social de las técnicas cualitativas, en especial el grupo de discusión, le está también conduciendo hacia este estatus público. Los grupos ya comienzan a ser, también, un espejo en el que la sociedad se mira, aunque en este caso el gran agente provocador ha sido la publicidad y el reconocimiento del estatus de espejo se limita a los sectores sociales más ilustrados. Un ejemplo muy claro de este proceso ha sido la epidemiología y la planificación sanitaria: hasta principios de los años 80 sólo valían los registros institucionales, de forma progresiva las encuestas se convirtieron también en contenedoras de datos "verdaderos" y en el presente siglo, al menos en el ámbito anglosajón, la "verdadera verdad" la aportan los grupos de discusión. El problema es que las técnicas y métodos de recogida y producción de datos no son la Verdad, sino meros e insuficientes instrumentos de aproximarnos a la realidad.

de la encuesta social, como los de Bowley, que empleó la muestra aleatoria en sus encuestas realizadas entre 1912-1914, marcaron un hito en el desarrollo de esta tecnología, aunque sólo sea porque introdujeron la técnica del muestreo. Hasta aquel momento las “encuestas” habían sido realizadas por personas muy ocupadas que no podían dedicarle mucho tiempo y de tal manera que el trabajo de campo lo llevaban a cabo voluntarios sin ninguna preparación y escasa organización del campo. Bowley fue el primero que utilizó encuestadores pagados y reprodujo, al final de su libro, una descripción de esta técnica de recogida de datos como tal.

En EE.UU. ocurrió algo similar, sólo que allí los reformistas procedían de ambientes religiosos. La Asociación Americana de las Ciencias Sociales fue fundada por personas preocupadas por la pobreza. El primer presidente de la asociación fue Lester Ward, quien afirmó que *“el verdadero objetivo de la ciencia es beneficiar al hombre. Una ciencia que fracasa en este objetivo, a pesar de lo agradable que sea estudiarla, está muerta”*. En este contexto el trabajo de Booth fue replicado en la encuesta de Pittsburgh de Paul Kellogg y sus colaboradores. Estos investigadores escenifican la transformación de la encuesta desde el modelo reformista primigenio que utilizaba la encuesta, para reflejar la realidad de las condiciones de vida, lo cual, a su vez, servía no sólo para denunciarlas, sino para cambiarlas, hasta dotarla de un rango académico que le condujo hacia un desarrollo espectacular sobre los dos campos que mejor definen la sociedad moderna: el mercado y la democracia política.

En el primer caso, las encuestas modernas evolucionaron partiendo de los primeros estudios de marketing que se realizaron después de la Primera Guerra Mundial, según un modelo de diseño experimental, el cual elaboró una amplia batería de test para medir las preferencias del consumidor. Tales test comenzaron a partir de los años 30 a utilizarse en encuestas aplicadas a la población general. Esto proporcionó la base para que los medios de comunicación encargaran las primeras encuestas sobre preferencias temáticas y más tarde sobre conocimientos y preferencias artísticas o literarias, con cuya práctica, especialmente en las revistas populares, cuajó la idea de que las condiciones sociales podían medirse y luego contarse. Ciertamente los cuestionarios eran muy primitivos, las entrevistas eran breves y ningún tema fue explorado en profundidad, pero la impresión social de que todo podía ser medido para facilitar el acceso de los ciudadanos a los bienes de consumo se expandió rápidamente (González del Río, 1996).

En el segundo caso, es decir, por lo que se refiere a las encuestas electorales y de opinión, el año 1936, como consecuencia de las elecciones presidenciales en EE UU, marca el antes y el después de esta técnica. En aquel año, como todos los sociólogos saben, con ocasión de la primera reelección del presidente Roosevelt, los observadores políticos concordaban en que el fracaso del presidente era más que probable. Una gran revista, la *Literary Digest*, publicaba, como de costumbre, los resultados de una votación particular después de haber examinado las respuestas de 2.400.000 personas y anunciaba su derrota. En el mismo momento Archibald Crossley, Elmo Roper y George Gallup, con una encuesta basada en una pequeña muestra de 4.400 personas acertaron el resultado de la elección. No sólo habían nacido las encuestas de opinión, sino que aparecían como un elemento básico de la acción política democrática (García Ferrando, 1986).

En aquel mismo periodo surgieron las primeras encuestas de uso y distribución del tiempo (EUT), aunque en este caso el protagonismo correspondió al sistema socialista instaurado en Rusia tras la revolución de 1917. La primera EUT la realizó Strumilin, en 1924, sobre los trabajadores de Moscú, dando lugar a una tradición de medida de los tiempos que se prolongó hasta 1989. Se trata de una tradición distinta a la de Occidente, porque sólo se pretendía obtener datos para la planificación económica y social, lo que, al disponer de abundantes recursos, planteaba notables mejoras metodológicas (Aguinaga y Comas, 1997).

Como consecuencia de todo ello, hacia el final de la década de los treinta la investigación por encuesta se generalizó en todo el mundo desarrollado, empezó a introducirse en la universidad, e incluso los demógrafos, en colaboración con estadísticos aplicados, empezaron a desarrollar métodos de muestreo para satisfacer las demandas estadísticas, abandonando otros sistemas

de registro. En general, la encuesta adquirió status de prestigio compuesta a partes iguales de ciencia, reformismo social, triunfo de la sociedad de consumo y ser uno de los componentes que garantizaba la propia estabilidad democrática.

Se trata de un proceso que, de forma similar, fueron viviendo diversos países al ritmo del desarrollo político y económico y que en España se produjo entre los años 60 y 70, siendo el INJUVE el principal protagonista de este proceso (Comas, 2007), al que más tarde se añadirían el Instituto de Opinión Pública y, al final, las empresas de Estudios de Mercado. En este momento la encuesta representa, en el imaginario social, uno de los elementos que a la vez definen e integran la sociedad de consumo, la democracia política e incluso el estado de bienestar.

### **11.3.- El proceso para la elaboración de encuestas.**

Elaborar y producir una encuesta es un procedimiento estándar, utilizado internacionalmente y tan elaborado que ya casi lo sabemos todo sobre él. Es un procedimiento complejo, que sin embargo ha producido una literatura técnica casi inabarcable, aunque sus componentes mínimos pueden ser aprendidos con rapidez por diferentes profesionales (Manzano, 1996).

La encuesta implica preparar un cuestionario, bien inédito o bien uno ya utilizado con éxito (Azofra, 2001, García Ferrando, 1986); supone realizar un diseño muestral probabilístico, disponer de una red de entrevistadores profesionales que seleccionen a los entrevistados de acuerdo con algún criterio (Rodríguez Osuna, 1991), una estructura de supervisión del campo, capacidad para tabular, codificar y grabar los datos, y un programa estadístico de los que desde finales de los años 80 ya pueden utilizarse en los ordenadores personales (Pons, 1993).

La técnica es la misma en Valencia que Tokio, en Los Ángeles que en Buenos Aires, aunque lógicamente en cada uno de estos lugares hay que adaptar el lenguaje, los conceptos y en una gran medida el estilo de contactación para realizar la entrevista. En todo caso, la aplicación de cuestionarios similares, desde la encuesta internacional de fecundidad hasta las encuestas europeas de valores, ha posibilitado la realización de interesantes comparaciones internacionales.

En resumidas cuentas, la encuesta es una industria que mantiene un sector de actividad económica no estrictamente científico, pero que produce información científica. En este sentido la encuesta es un método de conocimiento muy similar a la definición de evaluación que hemos proporcionado en el capítulo 1: una práctica social que se basa en los criterios de la ciencia. En ese sentido, las mediciones que se efectúen en una encuesta forman parte integrante del método y acontecer científico, de la misma manera que la evaluación de impacto de un programa de empleo para jóvenes es también una operación científica.

Pero, al mismo tiempo, una encuesta es un instrumento que se puede aplicar a diferentes situaciones. De este modo, la mayoría de datos que se identifican como “estadísticas” son en realidad encuestas, por ejemplo la EPA, la EPF, los análisis y previsiones de coyuntura económica y hasta los resultados censales se obtienen mediante encuestas. Las audiencias de los MCM se siguen midiendo por encuestas, incluso el estado de salud de la población se mide mediante éstas. La única amenaza para la hegemonía de la encuesta la constituye el reciente crecimiento de los sistemas de registro que la aplicación de las técnicas muestrales parecía haber eliminado durante el primer tercio de este siglo.

Una de las razones que explican la hegemonía universal de la encuesta, aparte de las relacionadas con su papel social, podría ser la flexibilidad con la que se puede adaptar el modelo estándar antes presentado a diversas situaciones. Para captar esta plasticidad podemos tomar como ejemplo las estrategias para la entrevista.

#### 11.4.- Estrategias para la entrevista en encuestas.

Son varias las posibles estrategias que se ofrecen al investigador para el diseño de las entrevistas en una encuesta. Dependiendo del tipo de interacción necesaria, de la duración y del contenido de la misma, podrá elegir entre la entrevista personal, la telefónica o la respuesta a un cuestionario remitido por correo.

Es evidente que cada una de estas posibles estrategias implica ventajas o debilidades, al igual que condiciones óptimas de aplicación. Uno de los aspectos a tener en consideración al diseñar una investigación es, primero, el grado en que cada estrategia de recolección de datos será comprendida por los entrevistados y, segundo, en qué modo será viable su realización.

Comenzando por la **entrevista personal** o “cara a cara”, hay que decir que se trata de un procedimiento de recolección de datos que implica una comunicación verbal entre el entrevistador y el entrevistado, entre cuyas ventajas destaca, en primer lugar, la obtención de una mayor proporción de respuestas que las obtenidas mediante la aplicación por correo de cuestionarios autoadministrados. En segundo lugar, la entrevista personal ofrece la posibilidad de obtener información de personas incapaces de rellenar un cuestionario si no son motivadas o ayudadas; en tercer lugar, la interacción verbal permite tanto clarificar preguntas al entrevistado como obtener aclaraciones de éste. Por último, en el caso de las entrevistas personales, cabe la opción de contrastar algún tipo de respuestas con observaciones directas.

La entrevista personal puede realizarse en un domicilio particular (utilizándose entonces el muestreo probabilístico), en instituciones o aglomeraciones (para los muestreos intencionales o de casos) o en la calle (para los muestreos por cuotas). Cada uno de estos ambientes posee sus propias complejidades, sus ventajas y sus inconvenientes.

De manera global, entre las desventajas de la entrevista personal aparece el coste del trabajo de campo; en segundo lugar, que la relación entre entrevistado y entrevistador afecte a la validez y fiabilidad de las respuestas, por la influencia de las diferencias sociales mencionadas cuando hablábamos de “la entrevista”. En este sentido, la habilidad del entrevistador es muy importante en este tipo de entrevista, pero a la vez es un arma de doble filo porque puede “alinearse” con el entrevistado para manipular al investigador. Asimismo, los entrevistadores pueden, consciente o inconscientemente, cambiar sutilmente el significado de una frase, convirtiéndola en tendenciosa, simplemente poniendo énfasis en determinadas palabras, cambiando la redacción o amenazando involuntariamente al entrevistado, presionando para obtener una respuesta o la aclaración de su significado.

Tales problemas inducen a la búsqueda de fórmulas para intentar evitar o minimizar tales efectos. Entre ellas se encuentra el prever las características de la población en estudio, intentando reducir los sesgos relacionados con el género, la edad, las diferencias culturales o los rasgos de carácter, enfatizando la formación de los entrevistadores.

La segunda estrategia realiza encuestas a través del **teléfono** y, aunque no es algo nuevo, en los últimos años esta técnica de recolección de datos ha experimentado un fuerte apogeo en los países más desarrollados en los que todos los investigadores parecen detectar una amplia aceptación social al hecho de ser entrevistado por teléfono (Bosch y Torrente, 1993). Eso mismo es más dudoso en los países con un menor nivel de desarrollo, o con sistemas democráticos inexistentes o poco consolidados, lo que parece indicar que es una técnica cuya legitimidad camina en paralelo al desarrollo social, económico y político. En realidad ha sido influido en el desarrollo tecnológico de los medios de comunicación, las técnicas de investigación social y el marketing, en especial la investigación de mercados, el banco de pruebas para desarrollar la encuesta telefónica, que se ha desarrollado básicamente en EE.UU., debido a los problemas de coste y accesibilidad que plantea en este país la entrevista personalizada con selección de sujetos por rutas aleatorias. Además, EE.UU. fue el primer país en el que casi la totalidad de hogares tuvieron teléfono.

En los últimos veinte años se ha generalizado el uso del sistema de recogida de datos CATI (Computer Assisted Telephone Interviewing), esto es, entrevista telefónica asistida por ordenador, que actualmente viene a representar alrededor del 40 por ciento del total de entrevistas que se hacen en los países de Europa y más del 70 por ciento de las que se hacen en EE.UU.

Pero lo cierto es que la gente es muy reacia a contestar entrevistas por teléfono, porque lo ven como una intromisión en su vida privada y porque además no se fían. De hecho, existe en algunos sectores bastante recelo a la utilización de esta técnica. Wert afirma que *“...la hipótesis de que las respuestas obtenidas por este conducto estarían más afectadas, que las recabadas directamente, por sesgos de insinceridad son, o bien directamente fruto de prejuicios sin más fundamento, o bien responden a hipótesis insuficientemente documentadas”* (Wert 1996). Si existe algún problema en el uso de las entrevistas telefónicas es la total dependencia del procedimiento a una cobertura telefónica total por parte de la población que se desee estudiar. El teléfono es un medio de comunicación “relativamente reciente” si se considera como medio de comunicación masivo. En España la generalización del teléfono es un hecho que se produce en los últimos 15 años, lo que ha implicado que hasta muy recientemente no haya sido posible su utilización para investigaciones con garantía de representatividad.

También es cierto que en los últimos años la introducción masiva del teléfono móvil esta produciendo muchos problemas para las técnicas basadas en CATI, en parte porque inmigrantes y jóvenes no tienen o no utilizan teléfono fijo y los móviles carecen de “guías telefónicas” desde las que establecer muestras territoriales.

En todo caso, resulta muy difícil, por no decir imposible, establecer muestras estratificadas para encuestas a partir de los datos telefónicos, por lo que hay que acudir siempre al muestreo aleatorio simple, a los que se puede añadir a partir de la declaración del entrevistado, algún tipo de cuota, aunque esto conduce a la eliminación a posteriori de algunas entrevistas. Otros problemas se refieren a la necesidad de una menor duración en la entrevista (el límite estaría en los 10'-15', aunque entre los jóvenes puede ser un poco más amplio) y al hecho de que los entrevistados pueden mentir con mayor facilidad (Wert, 1996).

Sin embargo, tales desventajas se compensan por la facilidad y la rapidez con la que se puede medir el impacto de un evento concreto y el óptimo de control y de supervisión, lo que en parte explica su crecimiento en los últimos años, especialmente para medir el impacto de acciones, acontecimientos o intervenciones políticas.

Para vencer las reticencias a la hora de participar en las encuestas telefónicas, se han desarrollado nuevos métodos de estimulación, a la vez que se ponen en práctica otros para estimar su validez. No basta solamente con conseguir que el entrevistado acepte contestar la entrevista, es preciso además que no aumenten los prejuicios a la hora de dar sus respuestas. La mayoría de estos métodos se basan en la doble llamada, la primera para adquirir un compromiso y la segunda para contestar a la entrevista.

La tercera estrategia es la **encuesta por correo** (Bosch y Torrente, 1993), que consiste en enviar un cuestionario a la persona seleccionada para facilitar la información, conjuntamente con una serie de instrucciones para contestarlo y un sobre preparado para su reenvío de vuelta al investigador. Las ventajas de las encuestas por correo residen en que son mucho más baratas, que garantizan mejor la confidencialidad y que permiten muestras más amplias y dispersas, ya que *“los cuestionarios pueden ser enviados por correo, los entrevistadores no”* (Selltiz 1975).

Pero las encuestas por correo tienen por ahora numerosos inconvenientes, el principal de los cuales sigue siendo las altas tasas de “no respuesta”. De hecho, la historia de la encuesta por correo es la lucha por poner en marcha estrategias que permitan aumentar las tasas de respuesta. Por ejemplo, el TDM (Total Design Method), que plantea un conjunto de pasos que interactúan favoreciendo la respuesta. El método se orienta al crear en el entrevistado una percepción de aumento de recompensas y disminución de costes al responder al cuestionario. Se basa en elementos como la comunicación anticipada, con carta de presentación o llamada telefónica, la

correspondencia personalizada, un cuestionario simple y atractivo y un seguimiento sistemático. Tiene el inconveniente de tener un coste relativamente alto, y ser rígido de manera que cuando no se sigue en su totalidad se producen peores resultados (Bosch y Torrente 1993).

La aplicación de cuestionarios por correo ahorra todos los costes de entrevistadores y de supervisión, restando sólo los de planificación, diseño, muestreo, copias y correo. Una segunda ventaja es la de eliminar los posibles efectos generados por la interacción entre entrevistador y entrevistado.

En la práctica, la administración del cuestionario por correo es especialmente útil cuando se busca una respuesta que exige la consulta de algún dato o documento por parte del entrevistado. También se considera que puede tener una mayor utilidad que las entrevistas cuando se trata de hacer preguntas que pudieran ser especialmente embarazosas de contestar. En estos casos, la aplicación de cuestionarios por correo tiende a proporcionar un volumen mayor de respuestas que en el caso del empleo de entrevistas personales (Nachimias 1976). La técnica mixta entrevista personal y respuesta por correo produce sin embargo excelentes resultados. Por ejemplo, en el ámbito de los estudios de juventud, en una encuesta realizada en el año 2002 por este mismo autor, las preguntas sobre sexualidad se auto-rellenaban y se incluían en un sobre cerrado. El resultado ha sido que los datos obtenidos son muy distintos a los del resto de las encuestas españolas sobre sexualidad. Hasta el punto que por primera vez las respuestas, necesariamente equivalentes de mujeres y varones, resultan concordantes (Comas, 2002).

La situación general, sin embargo, es que las encuestas administradas por correo suelen obtener un bajo índice de respuestas si se las compara con las alcanzadas mediante entrevistas personales. Además, aparecen otra serie de problemas que se tienden a olvidar.

Primero, que esta técnica es sólo factible cuando el porcentaje de la población que sabe leer y escribir se aproxima al 100%, lo cual, por regla general, no es un problema entre los jóvenes españoles. Asimismo, las direcciones de los entrevistados deben ser accesibles y se requiere que el servicio postal del país sea razonablemente eficaz. Por último, a semejanza de las entrevistas telefónicas, los individuos deben estar dispuestos a contestar los cuestionarios personalmente y devolverlos por correo, lo que en algunos casos puede suponer un riesgo para la confidencialidad de los datos.

La novedad más importante de los últimos años es la sustitución del envío postal convencional por el envío a través de correo electrónico, para muestras de población que dispone de estos sistemas de comunicación. Los resultados parecen un tanto sorprendentes, porque el volumen de respuestas es mayor que con el correo convencional. Pero si consideramos este resultado a la luz de lo que vengo diciendo, el mayor nivel de respuesta se relaciona lógicamente con la imagen social de la encuesta y el tipo de actitud, en relación a la información y el bien común, aunque dependiendo del tipo de cuestiones que se les planteen, que mantienen una mayoría de usuarios de las nuevas tecnologías (Tsagarousianou, 1988).

## **11.5.- La validez y la utilidad de la encuesta.**

Se sabe que la encuesta arrastra, como todas las técnicas de recolección de datos, numerosos problemas de validez, entre los que podría destacarse:

- Errores de no observación: equivocación en el muestreo, no respuestas y falta de cobertura adecuada.
- Errores en la recogida de datos: del instrumento, del método, del entrevistador y del entrevistado.
- Errores en el tratamiento de datos: en el procesamiento y en el análisis.
- Y por supuesto el resto de errores estocásticos y probabilísticos, cuyo conocimiento requiere un cierto grado de especialización.

Pero, a la vez, esta técnica ha desarrollado formas más completas y complejas de corregir, tanto los errores de validez como de fiabilidad, por lo que éste no es el problema de su utilidad. Los problemas de validez de la encuesta en realidad tienen que ver con el lugar social que ocupa como parte del sistema de producción y reproducción social. Un hecho que invalida, justamente porque elimina la validez del propio instrumento, la utilidad científica de las encuestas obtenidas indiscriminadamente por procedimientos industriales. Es decir, la propia eficacia de la tecnológica que ha conducido a la preeminencia de la encuesta, acaba por ser la causa de su crisis, porque a mayor implantación del instrumento mayor banalidad de sus aportaciones científicas.

No se trata por tanto de un problema de falta de validez técnica, sino de un problema de validez del instrumento en ciertas condiciones de aplicación y con relación a determinados temas. Es el caso, evidentemente bien conocido por mi parte, de las encuestas sobre drogas: la inclusión de preguntas sobre drogas en encuestas generales concebidas como procedimientos industriales arroja datos muy distintos que en encuestas diseñadas específicamente y aplicadas, a veces por la misma red de campo e idéntico muestreo, como “encuestas sobre drogas”, sin que las diferencias puedan atribuirse a errores sistemáticos o aleatorios. Algo parecido ocurre con la EPA y las preguntas sobre actividad de otras encuestas, y lo mismo podemos decir de las frecuentes encuestas sobre consumo y ocio: las actividades se confunden en los términos de las preguntas según el tema anunciado, con la consecuencia de que a diferentes encuestas diferentes resultados.

Un ejemplo claro de lo que vengo diciendo lo constituyen las encuestas sobre la participación social, voluntariado y altruismo, que producen cifras muy dispares y que dependen del contexto institucional explícito que plantea la investigación y la visión del mismo que tienen los ciudadanos de la misma (Comas, 2002).

## 11.6.- Las encuestas en la evaluación.

Con todo lo dicho resulta evidente que las encuestas son un instrumento adecuado para la recogida y producción de datos en las evaluaciones. Es además la técnica que goza de un mayor reconocimiento institucional y de una mayor aceptación por parte de la opinión pública. Aunque también suscita los mayores rechazos si la imagen que proporciona no resulta aceptable.

En una gran medida las ventajas de las encuestas se relacionan con su carácter industrial, es decir, es una tecnología muy avanzada que utilizan las empresas de estudios de mercado con un grado tal de eficacia, fiabilidad y validez, que contrasta abiertamente con toda práctica artesanal de las mismas. Y esto es muy importante en el terreno de la evaluación, donde es frecuente encontrar encuestas, con cuestionarios diseñados específicamente para esta evaluación concreta, con trabajos de campo realizados por una especie de “voluntariado” y con grabaciones de datos aún más artesanales. Esto no es un problema para otras técnicas donde aún, desgraciadamente, predominan condiciones de “espontaneidad”, pero en el caso de las encuestas resulta inaceptable.

Toda evaluación mediante encuesta requiere recurrir a procedimientos industriales y a las empresas especializadas, lo demás es ilusorio. A modo de ejemplo podemos resolver el problema de la distribución de agua y el desagüe de la misma en nuestra vivienda recurriendo a un fontanero y a los materiales idóneos, aunque también podemos recurrir a nuestro sentido común y a una compleja red de mangueras. No deja de ser original, pero tardaremos más, será más caro (al menos en términos de horas de trabajo), el servicio nunca será del todo eficiente y con el tiempo tendremos que ir sustituyendo mangueras con las consiguientes inundaciones. Lo mismo pasa con las encuestas.

Obviamente, recurrir a estos procedimientos industriales, a pesar de que sus costes están siempre muy ajustados, es en términos presupuestarios relativamente caro, mientras que la encuesta espontánea es “más barata”, aunque en términos de eficacia la evaluación será peor.

## El grupo de discusión: la interacción y la administración de los silencios

En la práctica habitual el Grupo de Discusión representa a la vez tres opciones alternativas. En primer lugar, es un procedimiento tecnológico cuya creciente formalización facilita la tarea del investigador; en segundo lugar, es también un dispositivo conversacional en el que puede practicarse la observación; y por último, se trata de un supuesto “paradigma metodológico” que se constituye por oposición al resto de las técnicas.

Pero en el terreno de la práctica de la evaluación el Grupo de Discusión es una técnica de recogida de datos, en la cual el investigador provoca **un diálogo que obliga a la participación**, en el que el investigador consigue que el mundo se manifieste de forma natural, pero alrededor de un tema y una situación controlada. En este sentido el Grupo de Discusión supone **un diseño cuasiexperimental** con el que el investigador puede observar aquello que ha provocado. A la vez, el Grupo de Discusión funciona como un escenario donde los participantes se sienten protagonistas, reconocidos y representativos, lo que les lleva adquirir conciencia de su papel social y a expresar tanto sus sentimientos personales como los de colectivo con el que se supone que se identifican.

La forma de proceder para utilizar grupos de discusión en evaluación ha sido descrito en otro Manual publicado por el INJUVE (Comas, 2001). En este voy a resumir parte de aquella propuesta con el fin de presentar de forma sintética dicho método.

### 12.1.- El paso final y la apertura primordial.

El grupo de discusión constituye el último nivel de aproximación empírica a la realidad. Se supone que es aquella práctica que, siguiendo la elipse propuesta, nos conduce hacia una contrastación integral, condensada y realista de la información que hemos conseguido acumular mediante otros procedimientos de recogida de datos y que refleja el momento empírico que cierra y agota los procedimientos para la recolección de los mismos. Pero esto no es cierto del todo, porque a la vez que cierra operativamente, reabre la elipse hacia otros planos topológicos, hacia otras concepciones, percepciones y realidades en los que la elipse puede reproducirse indefinidamente. El grupo de discusión es necesario para cerrar el sistema integral de los métodos y técnicas de investigación y a la vez es necesario para reabrir estos mismos métodos y técnicas.

En este sentido no se trata, en todo caso, de un cierre que “concluye” la investigación, sino sólo del cierre categórico del esquema de las técnicas, ya que la técnica del grupo de discusión representa esencialmente el momento de la apertura en el camino inductivo que hemos iniciado con la observación.

El grupo de discusión es el continente más amplio para la información acumulada que a la vez se proyecta reflexivamente para realinear a las otras prácticas, desde una perspectiva en la cual la evaluación:

- a) nunca se cierra,
- b) avanza y retrocede a impulsos y
- c) es inductivo-deductiva y por tanto debe circular permanentemente entre lo teórico y lo empírico.

Esto supone, al menos en mi experiencia y en contra de algunas interpretaciones que piensan que se trata simplemente de recoger “otros datos”, que el grupo de discusión supone un momento empírico singular en el que, aparte de los datos obtenidos, condensa un cierto grado de reflexión y de participación social. Una reflexión y una participación que posibilitan, a la vez, el cierre y la apertura.

Ciertamente, también las otras técnicas de recolección de datos contienen efectos reflexivos, incluso he descrito la encuesta como el más bruñido de los espejos sociales, pero el grupo de discusión, aunque se refleje en un espejo más tosco, contiene una doble mirada que se proyecta de la imagen del espejo al sujeto y del sujeto otra vez sobre el espejo. Es como la conciencia de clase de Lukacs (o según se dice las meigas), existe, pero hay que mostrarla para que pueda ser.

Esto significa que en la práctica todas las evaluaciones deberían incluir grupos de discusión y de hecho así lo hacen, aunque este momento empírico se subsuma, de manera inconsciente en los debates, seminarios, presentaciones, reseñas y ponencias posteriores que casi todas las intervenciones producen.

Por este motivo no hay que confundir el grupo de discusión con la “entrevista de grupo” y tampoco con la “reunión de grupo”, porque tales denominaciones se refieren al grupo básico, a la esperanza y al deseo de estar reunidos y expresarse, lo que pertenece a otros momentos empíricos distintos durante los cuales el investigador se limita a escuchar (observación) o a preguntar (entrevista). En cambio, el grupo de discusión se refiere al hecho real de tener que interactuar sólo para discutir el tema que propone el investigador, que orienta el contenido de una discusión, en la que el referente no es el investigador, sino lo que dicen los demás participantes en el grupo. El grupo se convoca para que discuta y se le permite estar reunido sólo el tiempo en que discute, pero durante este tiempo es una microsociedad en la que el investigador no está presente.

Ciertamente, el grupo de discusión es una técnica de investigación social que, al igual que las otras técnicas, también se basa en la palabra, pero, *“en el grupo de discusión lo que se dice - lo que alguien dice en determinadas condiciones de enunciación- , se asume como punto crítico en el que lo social se reproduce y cambia, como el objeto, en suma, de las ciencias sociales. En toda habla se articula el orden social y la subjetividad”* (Canales y Peinado, 1995).

En este sentido el grupo de discusión representa una microsociedad pero, a la vez, no tiene ninguna semejanza con los grupos que funcionan en la vida cotidiana, es un grupo artificial sin equivalente en la vida real, que, además, debe seguir siendo artificial durante todo su desarrollo. Es como afirma Ibáñez, un grupo simulado, imaginario, provisional que hemos construido para una manipulación muy concreta y delimitada, para que elabore un discurso, siendo este el motivo por el que hay que evitar que los participantes se conozcan, que no hayan mantenido ninguna relación previa para que su identificación con un grupo natural no interfiera en la dinámica del grupo de discusión. El grupo es una técnica de producción de datos sobre un esquema inductivo, pero a la vez es la puerta de salida del inductivismo hacia la reflexión: el grupo se ha diseñado para obtener información (selección de personas, elección de tema, control del

espacio y el tiempo,...), pero al mismo tiempo, en el momento en el que el grupo empieza a actuar, su naturaleza artificial deviene en una interacción social natural.

Sin embargo, como ya he mencionado al referirme al conjunto de aquellos que se denominan “métodos cualitativos”, aún manteniendo estos principios básicos el concepto y la técnica del grupo de discusión es distinto según las diferentes tradiciones nacionales, aunque en ocasiones no es tanto una cuestión de concepto como de nombre. Así, el grupo de discusión tal y como se ha practicado en España hasta ahora, siguiendo el modelo formalizado en los escritos del grupo de sociólogos encabezados por Jesús Ibáñez, Ángel de Lucas y Alfonso Ortí, creo que encaja perfectamente entre las técnicas descritas en la última versión del “Focus Group Kit” de Morgan y Krueger para la SAGE, en la que ya rompen definitivamente con la tradición Mertoniana. Esto vendría a confirmar que el grupo de discusión se va asimilando, en los años 90, a una práctica estándar, aceptable por parte de los medios profesionales, el público, el mercado y las instituciones, como le ocurrió en los años 50 a la encuesta (Bercher y Maple, 1996).

## 12.2.- La técnica del Grupo de Discusión.

Centrándonos en el caso concreto de España, la logística del grupo de discusión ha pasado de ser un tema menor, apenas veintitrés páginas sobre casi cuatrocientas en el texto clásico de Jesús Ibáñez (Ibáñez, 1979), a ocupar todo el desarrollo de la exposición (Canales y Peinado, 1995). La razón es clara, en los años 70, era una técnica a justificar y a solidificar teóricamente sobre bases científicas aceptables, como podía ser en aquel periodo el psicoanálisis; mientras que en los años 90 se trata de una práctica que debemos acometer con rigor, al tiempo que las coartadas teóricas se sustituyen por criterios de validez relacionados con el rigor logístico, es decir, con criterios de control y selección muestral (Martín, 1997). Esta evolución natural no supone que debamos prescindir de los componentes teóricos ni en el diseño ni en el análisis, sino simplemente se supone que ya no es necesario justificar permanentemente la técnica y basta con aplicarla correctamente.

Lo paradójico del caso es que el triunfo de la técnica se ha visto acompañado de las primeras llamadas de atención sobre “**la crisis de la representación**”. Muchos autores consideran que en las sociedades más desarrolladas, o bien los sujetos tienen una imposibilidad radical de representarse la realidad, o bien se la representan desde la condición de Debord, del mercado como espectáculo recreado por la publicidad (Schnaith, 1999). En ambos casos los grupos habrían dejado de ser productivos porque se limitarían a “repetir” la narración institucional o publicitaria. Creo que en una gran medida esto es cierto: la mejora de la logística tecnológica del grupo de discusión ha corrido en paralelo a su legitimidad social y ésta al perfeccionamiento de su condición de espejo. La encuesta y el grupo de discusión confluirían así en una cierta banalidad, al tiempo que adquieren respetabilidad.

Pero a la vez ambos representan momentos empíricos particulares, es decir, momentos que no podemos ignorar en la elipse de las técnicas de investigación social, lo que implica que debemos tratar de recuperarlos. Y la forma más adecuada de hacerlo es considerar que su creciente banalidad es una paradoja que nos remite a la totalidad de la mesa redonda de los métodos y técnicas de recogida y producción de datos que hemos expuesto en el capítulo 4.

Aunque sea un tema complicado que exige ciertos conocimientos epistemológicos, veamos como se está produciendo esta crisis que dará que hablar, al menos entre los profesionales más conscientes los próximos años: Claude Leford sostenía que los intentos de racionalizar lo irracional, de objetivar lo subjetivo, de explicar lo que sólo podía ser comprendido, no suponían errores metodológicos, sino paradojas vitales. Pero paradojas en el sentido de Winnicott, es decir, paradojas como aquellas que sostienen nuestros impulsos primarios (Winnicott, 1971), lo que implica que las paradojas metodológicas, si no se deben a errores, son una fuente de crea-

tividad. El dato banal es un objeto transicional, que quiere ser el verdadero objeto en un proceso invertido de simbolización, pero que sólo es un estadio en el proceso de investigación.

La paradoja es que lo cualitativo (y el grupo de discusión en particular) se ha constituido como crítica al operacionalismo de la encuesta, pero una vez constituido establece su propia logística e incluso establece registros. Luego sus mismas críticas a la encuesta se le vuelven en contra. Pero ¿son válidas? Sí como elemento constitutivo de la propia técnica, no en el contexto de la mesa redonda de los métodos y las técnicas de producción y recogida de datos. En este sentido, los Grupos de Discusión han condensado los rasgos de la creatividad de los últimos decenios, han sido objetos transicionales, pero al final el objeto real no ha aparecido, porque el objeto no existe en relación con cada una de las técnicas, sino sólo en el conjunto de todas ellas.

Por ello, hace unos años todos pensábamos que la logística del grupo de discusión, estas normas de aplicación estándar, no podían concebirse como tales, porque el grupo de discusión se construía a partir de la experiencia del investigador, el cual decidía cómo realizar la selección de los participantes, la forma de lanzar el tema, el esquema de actuación y, por supuesto, la interpretación del discurso durante la reunión y su análisis posterior.

Pero en los últimos años la bibliografía sobre esta tecnología no sólo ha crecido, sino que ya muestra un perfil básico más o menos común, al menos en ciertas cosas. Primero que **para seleccionar a los participantes** hay que utilizar una cualquiera de las muestras estructurales no probabilísticas, con criterios de pertenencia a los que atribuímos posibilidades de producir un discurso con relaciones relevantes, en función de su competencia lingüística y de la saturación del grupo. También sabemos que cuanto más definida esté la selección, más específica será la información que se obtenga, pero que aun en el peor de los casos, una selección muy poco acertada siempre nos proporciona alguna información.

Los grupos deben reflejar oposiciones sociales (por ejemplo a un grupo de jóvenes estudiantes se le opone un grupo de jóvenes trabajadores), pero no pueden contener cierto tipo de oposiciones internas. Ello implica que hay que excluir cierto tipo de relaciones entre la que se pueden señalar las diferentes clase social, edad y género, las relaciones ecológicas y espaciales, como ciudad/campo o centro/periferia, y las oposiciones conflictivas por sí misma, como patronos/obreros o estudiantes/profesores. También podemos lanzar hipótesis en torno a nuevas oposiciones sociales (por ejemplo, jóvenes integrados frente a jóvenes anti-globalización). Finalmente para que el grupo funcione es preciso una composición heterogénea que permita el intercambio, porque un grupo homogéneo producirá un discurso totalmente redundante, pero tan heterogéneo que no se pueda manifestar ningún tipo de consenso (Ibáñez, 1986).

En segundo lugar, para que el grupo de discusión funcione correctamente el **número de participantes** debe estar situado entre cinco y diez. Dos personas no constituyen un grupo, sino una relación especular. Con tres actuantes tendríamos un grupo triangular que implica otro tipo de técnica. Si los componentes son cinco, los canales de comunicación entre sus miembros superan ya el número de participantes, con lo que la relación grupal se hace posible. Pero más allá de nueve, los posibles canales de interrelación son tantos que el grupo tiende a disgregarse en conjuntos de menor tamaño, con lo que se vuelve inmanejable (Ibáñez 1979).

En tercer lugar, no hay demasiados criterios sobre el **número de grupos** que son precisos para realizar una investigación. Sabemos que no tiene por qué ser muy grande y que las buenas investigaciones se han hecho con muy pocos grupos de discusión, aunque bien diseñados, controlados, registrados y analizados. En todo caso la cifra mínima son tres, porque con dos grupos no podemos abarcar más que las diferencias lineales en los discursos. En todo caso, todo dependerá de la variedad de los grupos, ya que grupos con una composición similar saturan el discurso de una forma monótona y redundante, lo que en última instancia viene a decirnos que no importa tanto el número de grupos como el tipo de estructura que establezcamos con los mismos (Comas, 2001).

En este sentido, la **composición interna** del grupo no es indiferente. Así, cuando el grupo se divide en dos secciones, una con mayor tendencia a callar y otra con más tendencia a hablar, si lo podemos prever, habrá que reforzar la sección que tiene mayores dificultades para hablar o “empujarla” durante la reunión.

En cuarto lugar, la **convocatoria de los participantes** puede hacerse utilizando las redes sociales de comunicación, tanto la red privada compuesta de familiares, amigos, vecinos, compañeros de trabajo o conocidos, como la red pública que utiliza personas formalmente identificadas por alguna característica. En ambos casos es muy importante comprobar la identidad social atribuida por aquella persona que ha realizado el contacto, que suele ser un profesional un tanto maleado por montar grupos, en los que además en los últimos años suelen aparecer “profesionales de la participación en grupos de discusión”, por lo que la tendencia más reciente supone seleccionar a los participantes directamente mediante un muestreo que combine varias “bolas de nieve”.

En quinto lugar, debe **registrarse el texto** producido por el grupo de discusión en cinta magnetofónica y/o en vídeo, de lo que se tiene que informar a los participantes, así como del uso que se va a hacer del mismo (ver capítulo 13). El magnetófono registra el componente lingüístico del discurso y el vídeo añade los componentes no verbales. En ambos casos la mención al registro, aparte de la función técnica necesaria, cumple una función denotativa, confirmando la dimensión laboral del grupo, pues en el registro se va a depositar el producto del trabajo.

En sexto lugar, el grupo exige una **dirección**, aunque el papel de este director sea aparentemente pasivo. Ocurre que cuando un grupo funciona efectivamente los miembros adoptan posiciones simétricas a veces reiterativas y en otras excluyentes, pero se trata de posiciones reversibles y transitorias, que no se manifiestan si el director no provoca al grupo con nuevas propuestas temáticas, con reformulaciones o con interpretaciones aparentemente inocentes. En todo caso, esta es una tarea puntual y si el grupo sólo funciona basado en preguntas y respuestas resulta inhábil. La tarea del director supone además una permanente interpretación del transcurrir del grupo, captando el deseo, manifiesto y latente, del propio grupo, ya que el **análisis**, en esta técnica, no es sólo un procedimiento posterior a la recogida de datos, sino que se inicia en el mismo momento en el que el Grupo de Discusión se reúne.

Finalmente, en cuanto a la **duración** es necesario que se asigne un límite máximo, pero lo suficientemente flexible como para poder recoger el discurso del grupo cuando está maduro y evitando prolongaciones innecesarias. Este es un tiempo variable que depende de la dinámica particular de cada grupo y del tema a tratar, lo que equivale a decir que depende del tipo de investigación y del grado de cristalización del discurso. En general, la duración de un grupo de discusión oscila entre los sesenta minutos y las dos horas.

### 12.3.- La utilidad de los Grupos de Discusión en la evaluación.

Los Grupos de Discusión suponen una estrategia metodológica y una técnica de recogida de datos especialmente adecuada a la evaluación, en una gran medida porque aportan información muy pertinente a un coste discreto, es decir, son muy eficientes. En algunos casos las encuestas pueden ser más eficaces, pero su coste las hace inaccesibles. Otras técnicas son más baratas, pero su grado de eficiencia es inferior a los Grupos de Discusión.

Pero la utilización de los Grupos de Discusión en la evaluación de políticas, programas y actividades exige algunos comentarios particulares. En una publicación anterior ya se daban cuenta de ciertas dificultades de los Grupos en relación a la evaluación (Comas, 2002), pero al mostrar ahora el conjunto de métodos y técnicas de recogida de datos podemos sistematizarlas.

La cuestión es que, como hemos visto, los Grupos de Discusión son una técnica muy poco directiva, lo que implica que no resulta fácil fijar su atención sobre los objetivos concretos de la evaluación. Para lograrlo se suele recurrir a guiones más o menos cerrados y el “director” va planeando los temas a modo de preguntas para recoger las diferentes posiciones en torno a cada

uno de estos temas. Pero esto no es un Grupo de Discusión, sino una entrevista en grupo. Que tiene su utilidad, pero es un método distinto.

Porque el Grupo de Discusión según hemos tenido ocasión de explicar supone una recreación libre, espontánea y participativa de las condiciones sociales reales. Una recreación espontánea que sólo se limita por el eslogan de la convocatoria (por ejemplo: “valorar el programa de jóvenes creadores”), el perfil de los distintos grupos y polaridades que se establecen entre los mismos (por ejemplo, en el caso anterior dos grupos de participantes en el programa y dos grupos de creadores no participantes, cada uno de ellos con un perfil demográfico, social o ideológico antagónico) y la capacidad del “director” para evitar que “la conversación se desvíe hacia otros temas”.

En realidad este no es un problema exclusivo de la evaluación, sino que aparece ligado al creciente prestigio de “lo cualitativo” y a la expansión de una metodológica que nació en la Sociología (en concreto en el ámbito de los Estudios de Mercado), que ya es hegemónica en esta área de conocimiento y que se está expandiendo a gran velocidad hacia la pedagogía, psicología, filosofía y lenguaje,... para acabar haciéndose notar en ámbitos tan ajenos como las nuevas tecnologías y ciertos procedimientos de las ingenierías (y por supuesto en el ámbito de la epidemiología). Una expansión que ha supuesto una inevitable banalización (que incluye la confusión entre Grupo de Discusión y entrevista en grupo), lo cual no es preocupante, pero sí lo es la transformación de “lo cualitativo” en una estrategia para saltarse las reglas formales del método científico.

En este proceso de canalización se está produciendo una creciente confusión entre la metodología tradicional de los Grupos de Discusión, que servía para “privilegiar a los actores sociales, preservando su identidad, respetando sus categorías y facilitando su propio lenguaje”, para instrumentalizarse como una fórmula de legitimar perspectivas ideológicas. Demasiada gente realizando Grupos de Discusión, sin ningún tipo de formación y entrenamiento, para establecer discursos corporativos a partir de guiones cerrados y exhaustivos. Una práctica que equivale a realizar una encuesta, cuyo cuestionario se completa en una entrevista colectiva, en la cual “los entrevistados” son unas pocas docenas y además sin ningún criterio muestral. Los Grupos de Discusión que siempre fueron un procedimiento muy emic de dar la palabra al otro, se convierten así en un procedimiento etic de proyectar las creencias de los investigadores a través de una metodología cuyo logotipo acumula, al menos hasta ahora, mucho prestigio.

Este riesgo resulta especialmente relevante en el campo de la evaluación, ya que resulta fácil instrumentalizar Grupos para establecer conclusiones predefinidas y favorables a determinadas creencias. La manera de evitarlo, aparte de exigir rigor en la aplicación de esta metodología, pasa por la consideración del papel del “evaluador ilustrado” del que hemos hablado en el capítulo 3, que en un contexto de complementariedad por deficiencia del conjunto de los métodos debe comprometerse para evitar utilidades sesgadas de los diferentes métodos de investigación. Los estándares de evaluación que aparecen en los anexos pueden ser de mucha utilidad para evitar este tipo de problemas.

### PARTE III

#### LOS PROCEDIMIENTOS EN LA EVALUACIÓN

*“Los cuatro preceptos: 1) no admitir jamás como verdadera cosa alguna sin conocer con evidencia que lo sea, 2) dividir cada una de las dificultades que voy a examinar en tantas partes como sea posible y se requiera para su mejor solución, 3) conducir ordenadamente mis pensamientos, comenzando por los objetos más simples y más fáciles de conocer, para ir ascendiendo poco a poco, como por grados, hasta el conocimiento de los más complejos, 4) hacer de todo enumeraciones tan completas y revisiones tan generales que estuviera seguro de no omitir nada”*

René DESCARTES (1637), **Discurso del método**, Madrid, Tecnos, 1990.



### 13.1.- Una tarea para los artesanos abiertos a la participación.

El análisis de los datos obtenidos por las técnicas de recogida de datos descritas en el capítulo anterior es un momento esencial de la evaluación. A la vez es uno de los momentos más complejos de la misma, en el cual un exceso de espontaneidad o de confianza puede generar más errores. En el diseño de la evaluación y por supuesto en el desarrollo de las técnicas de recogida de datos existen procedimientos más o menos estándar, protocolos incluso, que han sido descritos en los dos capítulos anteriores y que un profesional avisado y con independencia de su titulación puede “aprender” con relativa facilidad sin llegar a ser un “experto” en los mismos.

Sin embargo, el análisis de datos siempre se ha considerado una tarea “artesanal” que se aprende con el tiempo y con la experiencia. De hecho no hay manuales (y muy pocos textos) sobre el análisis de datos y los que se titulan así se limitan a hablar de las técnicas y métodos de recogida de datos con incursiones más o menos afortunadas a los procedimientos estadísticos. Pero ni lo uno ni lo otro supone un verdadero análisis de datos.

Así puedo tener transcritas docenas de entrevistas que opinan que “la formación de los mediadores fue insuficiente”, pero el análisis no puede limitarse a asumir sin más dicha opinión, sino que debe interpretarla. Debe, por ejemplo, entender que ésta es una opinión que tiene que ver con el hecho de que los que acudieron a la formación no eran verdaderos mediadores sociales sino profesionales a los que les sobraba formación pero que iban en busca de otro certificado. También puedo disponer de una muestra suficiente de respuestas para inferir pruebas estadísticas en una tabla de contingencia (por ejemplo la de “chi cuadrado”), pero en sí misma tal prueba se limita a dejar constancia de la relevancia estadística de los resultados, pero tal relevancia puede ser positiva o negativa según se formulara la pregunta y esto es algo que el analista debe

saber interpretar. El mundo científico, incluso el ámbito especializado de la ciencia básica, está lleno de ejemplos de supuestas pruebas estadísticas que se suponía que aportaban correlaciones y que, sin embargo, por el contexto significaban exactamente lo contrario.

En los últimos años se han producido intentos de sistematizar el análisis de datos y apartarlo de su vínculo artesanal. Por ejemplo, en el ámbito industrial de los estudios de mercado (y la publicidad) se ha desarrollado el modelo de “data mining” (o “minería de datos”, como se ha traducido), pero su propia complejidad lo aleja de las posibilidades que ofrecen estas páginas. Además no estoy convencido de que pueda sustituir la tarea de un analista experimentado.

En todo caso, si esto ocurre en los dominios más especializados, qué no puede ocurrir en una evaluación muy socializada. De ahí que el papel del evaluador cobre una especial relevancia en este punto. Pero atención, este también es un momento del proceso en el cual el evaluador no puede “actuar por su cuenta”, porque el análisis de los datos de la evaluación es justo una de las coyunturas que exige una mayor participación, porque el análisis de datos del “experto” requiere el contraste de los agentes sociales involucrados que le ayudan en la interpretación.

Todas estas consideraciones previas nos permiten entender la estrategia que vamos a seguir en este capítulo. Vamos a recuperar alguno de los tipos de evaluación que hemos descrito en el capítulo 3 y vamos a comentar algunos aspectos particulares respecto a la configuración y el análisis de los datos en cada uno de ellos.

Esta configuración del Manual, en la que aparece una tercera parte destinada a presentar el análisis de datos, tras una serie de capítulos en los que se describen los métodos y técnicas de recogida de datos, responde finalmente a una clara intención: debemos tratar de evitar como sea la confusión entre ambos. Ni recoger datos es analizarlos, ni analizar los datos supone limitarnos a coleccionarlos. Se trata de tareas distintas. A la elipse de la mesa redonda de las técnicas, a la idea de la complementariedad por deficiencia, le corresponden la idea de las diversas finalidades del análisis. En el caso de la evaluación estas finalidades se corresponden con los diferentes tipos de evaluación.

Ciertamente, hay métodos y técnicas más adecuadas para un tipo de análisis o de evaluación y a ello nos vamos a referir en los próximos apartados, pero las posibles combinaciones son casi infinitas y pueden llegar a lo más lejos que nos deje nuestra imaginación.

### **13.2.- El análisis de la realidad y el diagnóstico de la situación.**

El análisis de la realidad y/o el diagnóstico de la situación suponen condiciones previas para emprender una evaluación (y por supuesto la planificación del plan o el programa), porque sin el conocimiento del “antes” resulta imposible la valoración del “después”. Ambas son maneras, terminologías para referirse a lo mismo. La primera es una expresión más típica de las ciencias sociales, la segunda de la psicología y del área de trabajo social, en parte porque mientras análisis de la realidad parece referirse a “los sujetos”, el diagnóstico de situación supone tener en cuenta a los recursos institucionales. Pero esto no es así para todos los autores. En los últimos años la terminología se ha enrevesado con otras expresiones, como por ejemplo “ecomapa”, que en algún caso aportan ideas interesante, como que lo institucional retroalimenta lo social y viceversa.

Con independencia de los criterios corporativos y las terminologías utilizadas hay varias formas o estrategias para enfocar el análisis de la realidad.

## LA ESTRATEGIA DE LOS PROBLEMAS<sup>30</sup>.

DESCRIPCIÓN. La única realidad que nos interesa es la de los problemas que se supone que afectan a los jóvenes. Tales problemas se identifican a partir de un cierto grado de consenso social en torno a su existencia. Una vez definido el problema se mide en términos objetivos, interesándonos especialmente por su tamaño y las posibles causas que lo producen. Por ejemplo, un problema puede ser “el fracaso escolar”, lo podemos medir a partir de un determinado indicador, observar su distribución y hacernos preguntas para determinar las razones que lo producen.

VENTAJAS. La facilidad para llevar a cabo la investigación, porque una vez delimitado el problema se puede investigar de forma sistemática. Otra ventaja se refiere a la posibilidad de realizar propuestas concretas y fácilmente valorables. Finalmente, la estrategia de los problemas tiene la ventaja del apoyo social, porque los ciudadanos entienden que el papel de las instituciones es “solucionar problemas”.

INCONVENIENTES. Los problemas verdaderos no son necesariamente los que constituyen la agenda social. Existen problemas invisibles que suelen ser más graves. Además, se confunden problemas y reivindicaciones, lo que implica que “a más soluciones más reivindicaciones”, de tal manera que cuantos más problemas resolvamos más problemas tendremos.

## LA ESTRATEGIA EVOLUTIVA Y COMPARATIVA.

DESCRIPCIÓN. Se supone que se pueden elegir determinadas cuestiones, o incluso “problemas”, y se estudia su evolución a lo largo del tiempo.

VENTAJAS. La comparación histórica ha sido siempre un método muy utilizado en las ciencias sociales. La idea de crear series para seguir la evolución de un fenómeno y establecer “cuál es la situación actual”, resulta práctica e inteligible. Es, además, muy rica en las conclusiones que ofrece el análisis.

INCONVENIENTES. El posible rechazo por parte de los jóvenes para los cuales la comparación con el pasado no resulta pertinente en lo que consideran “las características del presente”. En parte, esto ocurre porque esta estrategia ha sido muy utilizada para afirmar que “antes todo era mejor” en términos puramente ideológicos. En realidad, tal inconveniente desaparece si el análisis evolutivo se realiza de una forma adecuada.

## LA ESTRATEGIA ESTRUCTURAL O CONCEPTUAL<sup>31</sup>.

DESCRIPCIÓN. Se parte de una teoría, un modelo o una construcción política o ideológica y se jerarquizan explícitamente algunos temas o indicadores. Por ejemplo, el modelo marxista basado en la noción de “clase social” suponía analizar todos los datos desde la perspectiva del conflicto de clases. En la actualidad el modelo liberal (el individualismo

---

30 Para algunos autores en vez de “problemas” prefieren usar “necesidades” y el análisis de la realidad es entonces una “evaluación de necesidades” o incluso un “diagnóstico de necesidades”. En castellano, tanto en el DRAE como en el María Moliner, “tener una necesidad” es “tener un problema”, pero además “necesidad” equivale a “impulso irresistible que hace que las causas obren infaliblemente en cierto sentido” “aquello a lo cual es imposible sustraerse, faltar o resistir”, “carencia de las cosas que son menester para la conservación de la vida”, “falta continuada de alimento que hace desfallecer”, “especial riesgo o peligro que se padece, y en que se necesita pronto auxilio”. En este sentido “necesidad” es un término fuerte equivalente a un “problema para la supervivencia” y el análisis de necesidades, sería el análisis de aquellos problemas que amenazan la vida. Se trata de una análisis que hay que hacer en relación a determinadas situaciones y colectivos, pero que hay que reservar para estas intervenciones y no generalizarlo asimilando “necesidad” a cualquier “problema” y en ocasiones incluso a un servicio que poco tiene que ver con las necesidades vitales. Conviene también señalar que otros autores identifican el “análisis de necesidades” con la identificación de los recursos necesarios para desarrollar el programa (diferenciándolo, o no, del “análisis de necesidades sociales”), lo que añade confusión a esta cuestión.

31 Esta estrategia debe ponerse en relación a la discusión sobre el papel del evaluador tal y como se ha descrito en el capítulo 3.

metodológico) supone analizar todos los datos desde la perspectiva de “sujetos libres que toman decisiones (básicamente de consumo) a partir de actitudes propias o inferidas”. Lo mismo podríamos decir de la opción bien por las políticas de juventud afirmativas o bien las políticas de juventud de transición. Un ejemplo reciente realizado por CJE (Vidal y otros, 2006) nos abre interesantes perspectivas en relación a esta estrategia: la “construcción” se refiere a un objetivo (el aprendizaje de la participación) y se desarrolla en un ámbito (las organizaciones juveniles), el análisis diagnóstico se refiere a lo que ha ocurrido con el objetivo en dicho ámbito y a partir del mismo se realizan sugerencias para la posterior planificación de programas y a la vez se establecen los requisitos para su evaluación.

**VENTAJAS.** Trabajar desde un marco conceptual y teórico de referencia aporta notables facilidades para el analista. Dispone además de una “escuela” con otros analistas que comparten su punto de vista y que aportan conceptos y resultados de análisis que sirven para reforzar la propia posición. Además muchos modelos son compartidos por una parte del público y los profesionales, lo que permite que tales análisis sean más aceptables e inteligibles.

**INCONVENIENTES.** Todos los análisis de carácter estructural y conceptual se mueven sobre un “tempo histórico” determinado por una serie de condiciones que los convierten en referencia. Por ejemplo, con seguridad la actual oleada en favor del constructivismo y el postmodernismo no es una casualidad, sino que tiene que ver con la radicalización de las posiciones liberal-conservadoras y la hegemonía de las clases medias. Esto implica dos graves inconvenientes: no es fácil realizar análisis al margen de estas oleadas conceptuales y cuando la ola concluye (como paso con el marxismo), no suele quedar mucho de los avances logrados porque la ola “teórica” siguiente lo arrastra todo. En todo caso, un evaluador sensato tratará de evitar que las olas lo arrastren, sin oponerse del todo a las mismas para evitar que ahoguen.

## **LA ESTRATEGIA EMIC.**

**DESCRIPCIÓN.** Se articula a partir de la pregunta “qué quieren los jóvenes” en relación a cualquier tema elegido, para después formular tal pregunta y sistematizar las respuestas. En la estrategia emic sólo caben las opiniones y las actitudes de los propios jóvenes, lo que implica, aunque esto no es fácil, evitar cualquier planteamiento conceptual, teórico o político previo.

**VENTAJAS.** Se trata de una estrategia muy aceptable por parte de los jóvenes (aunque en general la aceptación explícita se limita a los jóvenes más organizados y participativos), a la vez se conjuga con el mandato constitucional (en España) de promocionar la participación, y en este sentido parece una ampliación de la democracia y una prolongación de la ciudadanía.

**INCONVENIENTES.** La variedad de las demandas, ya que de una parte un adolescente tiene intereses muy distintos a los de un adulto joven que trata de emanciparse. Aparecen además otros factores que fragmentan el colectivo en múltiples intereses. Se supone que esto se resuelve, en la fase del análisis, segmentando al colectivo juvenil. Pero tal segmentación no evita que los “acostumbrados” a expresarse impongan sus demandas, mientras que aquellos que tienen mayores dificultades para elaborar o transmitir estas demandas (y que quizá tengan más necesidades), no saben para nada expresarlas.

## **LA ESTRATEGIA DE LA INVESTIGACIÓN GENERAL.**

**DESCRIPCIÓN.** Se constituye a partir de una visión evolutiva que combina las demás estrategias, es decir, evolución de los “problemas”, evolución de las “demandas” y consideración de las diversas alternativas teóricas y conceptuales.

VENTAJAS. Ofrece una información completa, abriendo la evaluación a todas las respuestas posibles. Crea sinergias entre los tipos de análisis, goza de mucha credibilidad profesional y resulta, en términos científicos, la más adecuada.

INCONVENIENTES. Su notable complejidad a la hora de aplicarla a políticas o programas concretos. Además no es fácil sintetizarla, lo que produce un “efecto río de informaciones” que llega a aburrir a casi todos salvo los que han realizado el análisis.

### UN EJEMPLO: ¿Qué tiempo hace?

Supongamos una hipotética ciudad española en la cual la media histórica de la temperatura máxima de la segunda quincena de junio es de 43 grados centígrados. En un determinado año, en la última de estas dos semanas la media se sitúa en 40 grados con una subida repentina, ya que la semana anterior la media de las máximas solo alcanzó los 34 grados. El comentarista de la televisión local señala que se está viviendo una “insoportable ola de calor”. Tal afirmación divide a la opinión pública en la que aparecen posiciones muy divergentes

Para algunos, tal ola de calor no existe, porque normalmente para las fechas indicadas el calor es mayor. En realidad dirán que se está viviendo un verano bastante tolerable. Los que mantienen tal opinión manejan una imagen comparativa y diacrónica de la realidad.

Otros, en cambio, opinarán que cuarenta grados resulta en cualquier caso una temperatura insoportable y mientras dure habrá que calificarla de “ola de calor”. Los que mantienen tal opinión manejan una imagen presentista y sincrónica de la realidad.

Si realizáramos un estudio empírico en torno a estas opiniones, seguramente comprobaríamos que los que están trabajando, que pueden pensar en otras circunstancias de ambas maneras, se sitúan, sin embargo, de una forma mayoritaria en la imagen insoportable del presentismo; pero por otra parte, los que están de vacaciones y se pasan el día en la piscina o en la playa recurren más frecuentemente a la imagen amable de la comparación con otros años.

De la misma manera, otros, como los dueños de las piscinas, los que tienen locales con aire acondicionado o los vendedores de helados, que en otras condiciones también pueden opinar bien de una manera o bien de otra, en las circunstancias descritas afirmarán que “no es nada” e incluso los más atrevidos reclamarán que aparezcan “calores como los de otros años”.

Aparte, algunos ciudadanos muy sensibilizados por el cambio climático alegarán que se trata de “síntomas preocupantes” y propondrán que se tomen medidas urgentes para reducir las emisiones de CO<sub>2</sub>.

La pregunta es ¿el comentarista ha hecho bien en calificarla de insoportable ola de calor? Pues como vemos depende y su prestigio se vería realizado si realizara una afirmación como la siguiente: “estamos viviendo una temperatura inferior a otros años pero que, debido a nuestro clima, son bastante insoportables para aquellos que están trabajando. Esto no es así en otros lugares debido, posiblemente, a los efectos del cambio climático”. En este caso, sólo el grupo residual de los vendedores de helados no se sentiría incluido en la afirmación.

¿Cómo hay que realizar los análisis de la realidad?, pues tratando como ha hecho el comentarista del anterior ejemplo de englobar las diferentes perspectivas, aunque siempre aparecen interpretaciones que responden a modelos particulares que no estarán de acuerdo con nuestra síntesis, aunque la suya aún sería menos representativa

-----

En cualquier caso, el análisis de la realidad en la evaluación previa para el diseño de un programa o un plan de juventud en España no parte nunca del “bolsillo vacío”, porque ya hay mucha investigación disponible. Tanto en el nivel nacional como en el autonómico e incluso en muchos municipios. Olvidar todo este trabajo acumulado es comenzar muy mal la intervención (y la eva-

luación). Existen cientos de Estudios de Juventud (Comas, 2007), con referencias territoriales distintas, pero que podemos utilizar para realizar este diagnóstico.

A modo de resumen final podríamos decir que una política de juventud en España, cualquier política, plan o intervención, no puede formularse sin tener en cuenta el análisis de la realidad que aportan los estudios generales de ámbito nacional realizados por el INJUVE (aparte están los monográficos y los de otras organizaciones como la Fundación SM), cuyo conocimiento parece necesario para realizar un buen diagnóstico para planificar una política de juventud ¿hasta que punto los técnicos de juventud poseen una visión global de todos estos informes que les permita inferir los elementos sustanciales del análisis de la realidad? Pues si no la tienen difícilmente pueden planificar y evaluar sus programas.

### 13.3.- El análisis del diseño.

El análisis del diseño, que aparece como uno de los elementos sustanciales de la evaluación en cualquier manual, resulta, sin embargo, uno de los aspectos menos trabajado de la evaluación. De hecho, la experiencia española en esta materia es casi nula, lo que quizás explique el hecho de que los trabajos sobre calidad sean un sustituto de la evaluación de diseño. En todo caso, la función de la evaluación del diseño permite retroalimentar la gestión del programa.

La mayor parte de las evaluaciones de diseño tienen un origen institucional y son por tanto muy formales. Una evaluación de diseño típica consta de un número cerrado de preguntas que se clasifican en temas. Es muy frecuente que las respuestas deban ser de tipo binario (Sí o No). En todo caso aún las respuestas binarias deberán fundamentarse con evidencias bien argumentadas y documentadas.

Por ejemplo, una pregunta podría ser: ¿Cuenta el programa con evidencia de estudios o investigaciones nacionales y/o internacionales que muestren que el tipo de servicios o productos que brinda el programa es adecuado para la consecución de sus objetivos? Y la respuesta podría ser (Sí o No), pero después es preciso añadir la explicación adecuada a dicha respuesta.

#### VEAMOS UN EJEMPLO DE ESQUEMA DE EVALUACIÓN DE DISEÑO.

##### A.- Características del programa

Se trata de una breve descripción del programa que incluye unas Características Generales del Programa, que contiene la información básica y relevante del mismo.

##### B.- Análisis de la contribución del programa a los objetivos globales de las políticas de juventud. (Preguntas a contestar)

1. ¿El problema o necesidad prioritaria al que va dirigido el programa está correctamente identificado y claramente definido?
2. ¿Existe un diagnóstico actualizado y adecuado, elaborado por el propio organismo de juventud para sustentar la razón de ser del programa?
3. ¿El Fin y el Propósito del programa están claramente definidos?
4. ¿El Fin y el Propósito corresponden a la solución del problema?
5. ¿El programa cuenta con evidencia de estudios o investigaciones nacionales y/o internacionales que muestren que el tipo de servicios o productos que brinda el programa es adecuado para la consecución del Propósito y Fin que persigue el programa? De no ser así, el evaluador deberá investigar y, de existir, presentar dicha evidencia.
6. En base a los objetivos estratégicos de la dependencia y/o entidad que coordina el programa, ¿a qué objetivo u objetivos estratégicos está vinculado o contribuye el programa?

7. De lo anterior, analizar y evaluar si existe una relación lógica del programa con los objetivos de los Planes de Juventud de la UE o del Estado (o autonómica si es local).

#### C.- Presentación de la matriz de evaluación

##### *Presentación vertical de la matriz.*

8. ¿Las Actividades del programa son suficientes y necesarias para producir cada uno de los Componentes?
9. ¿Los Componentes son necesarios y suficientes para el logro del Propósito?
10. ¿Es claro y lógico que el logro del Propósito contribuye al logro del Fin?
11. Considerando el análisis y la evaluación realizados en este punto, ¿la lógica vertical de la matriz de indicadores del programa es clara y se valida en su totalidad? Es decir, ¿la lógica interna del programa es clara?
12. Si no es así, proponer los cambios que deberían hacerse en el diseño del programa y en su lógica interna. Estos cambios deberían reflejarse en la matriz de indicadores definitiva del programa.

##### *Presentación horizontal de la matriz.*

13. En términos de diseño, ¿existen indicadores para medir el desempeño del programa a nivel de Fin, Propósito, Componentes y Actividades e insu-  
mos?
14. ¿Todos los indicadores son claros, relevantes, económicos, adecuados y pueden arrojar medidas?
15. De no ser el caso, la institución evaluadora, en coordinación con el programa, deberá proponer los indicadores faltantes y necesarios para cada ámbito de acción o las modificaciones a los indicadores existentes que sean necesarias.
16. ¿Los indicadores incluidos en la matriz de indicadores tienen identificada su línea de base y temporalidad en la medición?
17. ¿El programa ha identificado los medios de verificación para obtener cada uno de los indicadores?
18. Para aquellos medios de verificación que corresponda (por ejemplo encuestas), ¿el programa ha identificado el tamaño de muestra óptimo necesario para la medición del indicador, especificando sus características estadísticas como el nivel de significancia y el error máximo de estimación?
19. ¿De qué manera el programa valida la veracidad de la información obtenida a través de los medios de verificación?
20. ¿Se consideran válidos los supuestos del programa tal como figuran en la matriz de indicadores?
21. Considerando el análisis y evaluación realizado en este punto, ¿la lógica horizontal de la matriz de indicadores se valida en su totalidad?
22. Si no es así, proponer los cambios que deberían hacerse a la lógica horizontal de la matriz de indicadores (indicadores, medios de verificación y supuestos).

#### D.- Población potencial y objetivo

La población potencial corresponde a la población total que presenta la necesidad y/o problema que justifica el programa y, por ende, pudiera ser elegible para su atención. *Por ejemplo, los jóvenes mayores de 18 años que no han obtenido un título de la ESO.* La población objetivo será aquella que el programa tiene planeado o programado atender en un periodo dado de tiempo, pudiendo corresponder

a la totalidad de la población potencial o a una parte de ella. *Por ejemplo, los jóvenes mayores de 18 años que no han obtenido un título de la ESO, y que se encuentran en situación de desempleo en los últimos 6 meses.*

Tanto la población potencial como la población objetivo pueden estar definidas en regiones, municipios, localidades, hogares y/o individuos, en su caso. La población potencial también puede estar constituida por agentes sociales organizados de naturales muy diferentes.

23. ¿La población que presenta el problema y/o necesidad (población potencial), así como la población objetivo están claramente definidas?
24. ¿El programa ha cuantificado y caracterizado ambas poblaciones, según los atributos que considere pertinentes? (En el caso de individuos, en términos de edad, sexo, nivel socio-económico u otros atributos que sean pertinentes).
25. ¿Cuál es la justificación que sustenta que los beneficios que otorga el programa se dirijan específicamente a dicha población potencial y objetivo?
26. ¿La justificación es la adecuada?
27. ¿Los criterios y mecanismos que utiliza el programa para determinar las unidades de atención son los adecuados? (Señalar principales mecanismos).
28. ¿Existe información sistematizada y actualizada que permita conocer quiénes reciben los apoyos del programa (padrón de beneficiarios), cuáles son las características socio-económicas de la población incluida en el padrón de beneficiarios y con qué frecuencia se recoge la información?

E.- Análisis de la vinculación entre los procedimientos establecidos y los objetivos del programa

29. ¿El diseño del programa se encuentra correctamente expresado en procedimientos y normas de trabajo?
30. ¿Existe congruencia entre tales procedimientos, los objetivos del programa y su lógica interna?

F.- Posibles coincidencias, complementariedades o duplicidades de acciones con otros programas aplicados en el territorio.

31. Como resultado de la evaluación de diseño del programa, ¿el diseño del programa es el adecuado para alcanzar el Propósito antes definido y para atender a la población objetivo?
32. ¿Con qué otros programas sociales podría darse algún tipo de complementariedad y/o sinergia?
33. ¿Con qué programas de juventud o sociales generales podría existir duplicidad?
34. ¿El programa cuenta con información en la que se hayan detectado dichas complementariedades y/o posibles duplicidades?

G.- Adecuación del equipo.

Completar por cada miembro la correspondiente fila del cuadro de la siguiente página, para analizar posteriormente el grado de adecuación entre los procedimientos, las normas de trabajo, los objetivos del programa y las responsabilidades y compromisos asumidos y la posibilidad de desempeño. También se valora la suficiencia del equipo.

H.- Fortalezas, Retos y Recomendaciones

El evaluador externo deberá identificar al menos (X) fortalezas y también (X) oportunidades, debilidades y amenazas, deberá establecer en forma concisa cada fortaleza y oportunidad, o en su caso debilidad y amenaza, incluyendo la evidencia que la sustenta e indicando la fuente del análisis. Cada fortaleza y oportunidad,

o en su caso debilidad y amenaza, deberán estar ordenadas de acuerdo a su relevancia, donde la primera será de mayor importancia con respecto a la segunda y así sucesivamente.

Por cada debilidad o amenaza se deberá establecer al menos una recomendación precisa que ayude a solventar el reto o problema que fue identificado. Es importante que las recomendaciones estén cuidadosamente analizadas para garantizar su viabilidad y efectividad.

Si se incorpora una recomendación que trasciende las atribuciones del programa, se debe señalar claramente si su implementación corresponde a la dependencia y/o entidad responsable del programa o a otras instituciones públicas. Asimismo, se debe indicar si la implementación requiere de cambios normativos y/o legales o recursos adicionales.

Es necesario mencionar que todas las recomendaciones deberán estar ligadas al menos a una debilidad o amenaza, es decir, se pretende no generar recomendaciones generales, sino específicas para cada reto o problema (debilidad o amenaza) detectado en la evaluación.

En todo caso y aún en este ejemplo, hay que tener en cuenta “la paradoja de la evaluación de diseño”, que expusieron hace ya muchos años Michel Crozier y Erhard Friedberg en un libro que tradujo sin publicar y difundió el propio Instituto Nacional de Administración Pública. Ambos autores, sin referirse explícitamente a la evaluación de diseño, pero describiéndola como un “análisis previo”, mostraban numerosos ejemplos de esta paradoja: (1) desde una perspectiva institucional muchas evaluaciones de diseño pueden concluir que los programas y las intervenciones analizados resultan incompletos, contienen contradicciones y ambivalencias; (2) sin embargo, aún en estas condiciones suelen aplicarse, porque ya se han puesto en marcha y han sido presupuestados; (3) pero al concluir han producido efectos y han adquirido algún sentido inesperado y positivo (Crozier y Friedberg, 1977).

### **13.4.- El análisis de la implementación y de los procesos.**

El análisis de aquellos aspectos de la evaluación que se refieren al “durante” es el más habitual y se relaciona esencialmente con el registro de los mismos, de sus actividades y de las valoraciones que realizan los clientes o usuarios del mismo. También se pueden utilizar otras técnicas, en particular la Observación Participante, pero lo habitual es trabajar a partir de registros.

Los objetivos de la evaluación de implementación, de evaluación de procesos así como de la evaluación de procedimientos y personal, son diferentes y corresponden a la definición que hemos dado en el cuadro 3.1. Pero en este apartado (el Manual ya es casi demasiado extenso) los agruparemos bajo una única categoría y un único ejemplo.

Se trata además de un tipo de análisis que tiene mucho que ver con la cuestión de la “calidad” y sus relaciones con la evaluación, tal y como se ha descrito en el capítulo 2 (apartado 2.3). De hecho, especialmente en el ámbito industrial y empresarial, son muchos los que evalúan procedimientos limitándose a aplicar la ISO-9000 o algún estándar de calidad similar y no conciben la evaluación de otra manera. En este caso vamos a alejarnos de este estándar para aproximarnos más hacia las evaluaciones de procesos tal y como se diseñan en el ámbito educativo.

Existe la costumbre de evaluar los procesos en una perspectiva interna que se limita a preguntar: ¿Hemos hecho lo que queríamos hacer? Pero esto no es correcto del todo porque hay que tener en cuenta, aparte de los objetivos y de los procedimientos definidos en la planificación, algunos aspectos, relacionados tanto con la implementación como con el proceso, entre los que podemos destacar el contexto, el ámbito temático y el territorial, los aspectos político/administrativos e incluso el perfil de los profesionales involucrados. Es decir, al menos para los ámbitos de juventud, propongo que las evaluaciones de procesos tengan muy en cuenta todos estos

aspectos y no se limiten a aplicar simples protocolos de doble respuesta (o escalas) con los que “medir” si las cosas se han hecho o no. Obviamente para evaluar estos aspectos se requiere utilizar técnicas de recogida de datos que puedan recoger otros datos u opiniones sobre los mismos.

Después conviene definir cual es el objeto de la evaluación. En este caso podemos pensar en el programa en cuanto a tal (es lo más habitual), pero también en el tipo de política pública que lo sostiene, en el grado de participación de los Agentes Sociales previstos o incluso en el logro de los objetivos del programa al margen de los resultados o los impactos que hayan producido. Un ejemplo de este último caso puede ser una evaluación de procesos en un programa de información o de prevención que determiné que los mensajes han llegado efectivamente a la población diana. Como puede observarse, los tipos de evaluación se superponen pero dan lugar a informaciones diferentes según el momento y la finalidad de la evaluación.

Finalmente la propia evaluación de procesos requiere un **esquema secuencial** de constataciones determinado por la planificación del programa y el papel otorgado a los diferentes actores sociales, profesionales e institucionales en el mismo. Cada uno de los componentes del programa debe ser colocado en este esquema secuencial ya que algunos de ellos “filtran” a los siguientes. Por ejemplo, si no se ha preparado un folleto descriptivo difícilmente se puede evaluar si se ha realizado la distribución del mismo. Aunque en ocasiones las demandas institucionales son peculiares, así en la evaluación de un reciente programa no se preparo (ni se difundió, claro) un folleto informativo que facilitaba un teléfono de consulta para información sobre sexualidad y ETS. Pues bien la administración insistió en que se debía añadir a la evaluación una valoración del funcionamiento del teléfono ya que estaba planificado y presupuestado. Me limité a decir en tono irónico “que no había recibido ni una sola llamada”, lo que provocó una columna de opinión en el periódico local en la que se afirmaba “que como nuestros jóvenes creen que lo saben todo sobre sexualidad, ya no preguntan”, lo que como cualquier técnico de juventud sabe es más que incierto. De ahí la importancia de incorporar la lógica secuencial a cualquier evaluación de procesos (y en parte de resultados).

¿Cuáles son los criterios (los aspectos) a evaluar en esta lógica secuencial? Por ejemplo en un programa de promoción de la participación apoyado en las entidades juveniles.

El primer criterio tendrá que ver con la claridad en los objetivos, no tanto en la perspectiva de su formulación (esto es más bien evaluación de diseño), sino en que se han traducido en acciones efectivas y reales (lo que a su vez implica cobertura).

Un segundo criterio se relacionaría con los aspectos organizativos, por ejemplo si la convocatoria se realizo en tiempo y forma, si los recursos llegaron a tiempo, si las entidades recibieron bien la información, si no tuvieron dificultades en presentar sus proyectos...

El tercer criterio puede referirse al grado de participación de las entidades (cantidad, diversidad y representatividad), es decir, cuantas se presentaron, cuales son y las razones por las que algunas no lo hicieron. En este punto sería conveniente establecer cual ha sido el grado de aceptación por parte de las mismas. Con este ejemplo podemos además observar que si en la planificación hubiéramos formulado algún objetivo relacionado con “el establecimiento previo de un cierto consenso con las entidades juveniles en torno al contenido del proyecto”, la evaluación de este criterio del proceso (el grado de aceptación) sería, casi de forma automática, positiva. Lo que refleja como se pueden retroalimentar, de una forma fácil, las diferentes fases, momentos y tipos de la planificación y la evaluación.

Un cuarto criterio puede ser la valoración técnica del proceso por parte de los profesionales que lo han llevado a la práctica. Si además en el programa aparecían objetivos transversales, por ejemplo relacionados con participación ciudadana o promoción de asociaciones de estudiantes, hay que tratar de establecer si estas áreas se han implicado de manera adecuada.

Un quinto criterio, muy particular del ejemplo de un programa de participación juvenil, puede referirse a la emergencia de nuevas iniciativas y liderazgos, siempre que el programa hubiera pre-

visto “abrir nuevas vías y ámbitos de participación y tratar de que nuevos colectivos sociales se incorporen a la mismas”.

Un sexto criterio podía ser y de haberse previsto un órgano de seguimiento establecer si había funcionado de manera adecuada, si había recibido toda la información pertinente y si había tomado decisiones sobre el funcionamiento del programa.

Estos seis criterios no son exhaustivos salvo para el ejemplo mencionado, porque en realidad el análisis de procesos, se basa en una simple suposición: existe un plan y hay que constar si se ha cumplido, y esto significa que los criterios de la evaluación los determina el plan. Obviamente si no hay plan no es posible evaluar procesos.

### **13.5.- El análisis de los resultados y los impactos.**

#### **13.5.1.- La función del análisis de resultados.**

Se supone que el análisis de los resultados y los impactos es el “verdadero” objetivo de toda evaluación. Pero en la realidad este tipo de análisis es muy poco frecuente debido a su dificultad. No se trata de una dificultad metodológica intrínseca al propio procedimiento de evaluación, sino a lo “delicado” de este tipo de análisis. Porque un análisis de la realidad o de diseño tiene un público objetivo limitado (para el que estas evaluaciones son muy importantes) formado por aquellos que intervienen en la evaluación. También es importante para que aquellos que son “clientes” en dicha intervención, especialmente cuando el programa “trata de resolver problemas”, aunque en este caso su interés se limite a que “sean atendidos lo mejor posible”.

Sin embargo en el análisis de impactos o resultados la evaluación se convierte en una información para el público en general con una fuerte carga de connotaciones políticas. Una evaluación de procesos, por ejemplo, nos ayuda a mejorar el programa, pero si nos equivocamos en la misma, las consecuencias se limitan al propio programa y sólo suelen ser parcialmente negativas (por ejemplo pérdida de eficacia), aunque también pueden influir en el resultado final. Sin embargo cuando hablamos de evaluación de resultados e impacto las consecuencias pueden ser muy importantes. Pueden, por ejemplo, llevarnos a revisar una política o un plan y dejar de lado la atención a un determinado colectivo, pueden cambiar las estrategias de intervención y pueden tener efectos electorales de importancia.

Esto nos obliga a ser especialmente rigurosos en la evaluación de resultados e impacto, con exigencias como las mencionadas en los párrafos siguientes, pero también hay que tener en cuenta que la evaluación de resultados requiere ciertas cautelas (y constataciones acumuladas) antes de convertirse en certezas. De hecho una evaluación de resultados requiere traspasar el nivel de las tablas de contingencia o de la recopilación simple de verbatimines, para pasar al nivel de las ecuaciones estructurales, del análisis de contenido o de los procedimientos semánticos del análisis estructural. Y todo esto requiere profesionales preparados.

#### **13.5.2.- La selección de estándares, dimensiones e indicadores.**

El primer problema que si podemos resolver de una forma relativamente fácil, se refiere a la selección de los estándares, de la dimensiones y de los indicadores de la evaluación de resultados e impacto.

Se resuelve fácil si se lee este Manual, porque de hecho, desde el inicio, desde la propia definición de evaluación y de la noción de evaluación fuerte, hemos estado dándole vueltas a la cuestión. Hemos hablado de estándares en el capítulo 3 y se han incluido dos sistemas de

estándares como anexos, ya se ha hablado de matrices e indicadores y vamos a desarrollar todo el tema de las dimensiones y los indicadores a lo largo del capítulo 16.

¿Por qué se ha incluido entonces este epígrafe en el apartado de evaluación de resultados? Pues en parte para recordar que la selección de estándares, dimensiones e indicadores en una evaluación de resultados, no es un asunto baladí. Pero en parte para explicar que la selección de estándares, dimensiones e indicadores no es neutral. Elegir uno u otro estándar o indicador tiene consecuencias políticas y sociales y es absolutamente falso que tanto estándares como indicadores puedan ser apolíticos. No existe un conjunto de estándares o indicadores “científico”, cuya selección no tenga consecuencias políticas. De hecho incluso cuando se está realizando esta afirmación de neutralidad se efectúa una elección política. Una elección muy política diría.

Por este motivo he elegido el sistema de estándares de evaluación del sistema de Naciones Unidas, porque supone, en términos políticos, una apuesta clara por los derechos humanos y la democracia, porque trata de crear una comunidad internacional fuerte y competente, siendo además el resultado de un trabajo multidisciplinar y multisectorial. No todos los países (en especial algunas teocracias religiosas) están de acuerdo con este sistema, ni lo están todos los colectivos sociales y profesionales. Pero sus propuestas que apelan bien a la tradición cultural bien a alguna versión particular de ciencia o pseudo-ciencia, no tienen la garantía que ofrecen Naciones Unidas.

También se ha incluido como anexo el sistema del JCSEE, porque es el resultado de un trabajo muy sistemático desarrollado por un grupo de evaluadores (la mayoría del ámbito educativo), que supuso un antes y un después en la cuestión de los estándares de evaluación.

Con esto podemos entender la importancia de seleccionar de estándares, dimensiones e indicadores que respondan a criterios de normalidad democrática y técnica, pero a también a criterios éticos.

### **Un caso famoso y aleccionador.**

*En el año 1990, en EE.UU., se nominó, tras conocerse los resultados de una evaluación de resultados, a la Comisión para la Igualdad de Oportunidades en el Empleo, al Premio Presidencial de la Calidad. El dato que dio lugar a la propuesta se refería al continuo y sensible aumento del “porcentaje de éxitos” en los casos de discriminación llevados a los tribunales por parte de La Comisión, en el periodo 1983-1989, es decir, justamente el periodo que llevaban las administraciones republicanas en el poder (con Reagan y Bush padre).*

*Sin embargo la comunicación de la nominación generó una fuerte oleada de protestas entre los “clientes” de la Comisión (organizaciones de minusválidos, minorías étnicas, mujeres,...), porque si bien la cifra era cierta también lo era que el número de casos llevados a los tribunales por la Comisión había descendido hasta un 32% en aquellos siete años. Lo que implicaba que el número real de éxitos era, en términos absolutos, menor e incluso se apreciaba un ligero descenso del volumen anual de casos de éxito.*

*Los grupos que habían propuesto la nominación pasaron al contraataque y señalaron que este era el resultado de una actitud más correcta por parte de las empresas que eran menos “discriminatorias con las minorías” que en el pasado. Tras unos días de desconcierto ante este argumento los “clientes” reunieron datos y pudieron mostrar como los casos de denuncia no atendidos por la Comisión también habían aumentado en aquellos siete años en un 22%. (Player, 1990).*

*Llegados a este punto los analistas se preguntaron la razón por la cual la administración republicana había cometido el “error” de presentar tal nominación y*

*descubrieron que la Comisión tenía instrucciones precisas para reducir el impacto de las demandas sobre las empresas, para no “interferir en la vida económica y no poner trabas al crecimiento”. La idea de las administraciones republicanas era que las Leyes anti-discriminación, que en ningún caso iban a derogar, no podían interferir en el desarrollo económico y para ello no se debía presionar demasiado a las empresas, salvo en casos de incumplimiento flagrante o grave escándalo público. La Comisión había seguido con gran éxito estas instrucciones, que se convertían así en un estándar y en una dimensión de la evaluación y había elegido un indicador acorde con las mismas, por tanto se merecía el premio.*

*Claro que sus clientes, en particular el Movimiento de los Derechos Civiles, tenían como misión y objetivo reducir la discriminación, lo que implicaba que sus estándares y dimensiones eran otros y por tanto el indicador tenía que ser distinto.*

*Llegados a este punto los analistas se formularon una pregunta interesante: ¿Era lícito establecer estándares invisibles, que además no tenían en cuenta las consideraciones éticas y democráticas que justificaban la existencia de las Leyes anti-discriminación? ¿Era lícito priorizar el interés en el crecimiento económico y en la inversión empresarial obviando el contenido y sentido de las Leyes? La respuesta de este Manual es que no, porque hemos adoptado un sistema de estándares y dimensiones (basada en el bienestar de las personas y los derechos humanos) que convierten esta pretensión en inaceptable. Aunque también hay que entender que una parte de los ciudadanos ligan el bienestar sólo al crecimiento económico.*

*En todo caso este ejemplo nos permite entender que la evaluación de resultados y de impacto, depende de la perspectiva (y los intereses) de los diversos agentes sociales. Para algunos mientras que un impacto concreto puede ser positivo, para otros es negativo. Esta divergencia de opiniones no refleja un problema metodológico sino ético, que se resuelve en la elección de los estándares.*

### **13.5.3.- El control de variables y efectos espurios.**

Otro de los problemas importantes a los que se enfrenta la evaluación de resultados e impacto se refiere al llamado “error aleatorio” relacionado con los efectos no considerados de otras variables no tenidas en cuenta. Un tipo de error que en parte puede atribuirse a la técnica (le hemos llamado “error estocástico” en la parte II), pero que en general tiene que ver con variables inesperadas, desconocidas y externas que no se pueden controlar mediante procedimientos simples de “validez interna”.

Con un ejemplo se entenderá mejor: supongamos que queremos determinar cuales han sido los resultados de un programa de promoción de la tolerancia intercultural en el colectivo de jóvenes de una Comunidad Autónoma. Supongamos que además tenemos claros cuales son las dimensiones y los indicadores para medir este resultado (cambio de actitudes mediante un cuestionario que han sido validado en otras investigaciones o evaluaciones) y que mantenemos la hipótesis que una mayor exposición a las actividades del programa (intervenciones en MCM, conciertos y otros actos inter-étnicos, intervenciones de sensibilización en centros escolares, apoyo a actividades del movimiento asociativo juvenil,...).

De entrada podemos suponer que, además del grado de exposición, ciertas variables como edad, valores previos, género, nivel educativo y tipo de enseñanza, tipo de actividad de ocio preferente o clase social, influirán en el resultado e incluso nos ayudaran a explicarlo. Por este motivo podemos introducirlas en el diseño, algunas de manera fácil, como la edad y el género, y otras de una forma más complicada, como por ejemplo la cuestión de los valores previos. En todo caso al hacerlo estas variables dejan de ser espurias y se convierten en variables de control que nos ayudan a explicar los resultados.

Pero si no sabemos cuales son estas variables o no podemos introducirlas en el diseño entonces se nos plantea otro problema. A la vez pueden aparecer variables espurias que no tengan nada que ver con lo que tratamos de medir, pero con las que se establecen una clara relación. Así uno de mis profesores de estadística copiaba en la pizarra la evolución del volumen de nupcialidad en España entre 1955 y 1970 (con base 100 el primer año), con la evolución de la construcción de toneladas brutas de barcos en Japón en el mismo periodo y nos pedía que calculáramos un índice de correlación. Salía clavado, lo que le llevaba a concluir con ironía: “vaya, si pretenden ustedes casarse tendrán que estar muy atentos a la evolución de la construcción naval en Japón”.

Para este problema sólo hay una solución: poder conocer cuales son estas variables e introducirlas, o eliminarlas según sea el caso, en el diseño. Y este conocimiento sólo puede proceder de una estrategia de acumulación de evaluaciones, es decir cada día podemos saber más sobre las variables espurias si hacemos más evaluación y la difundimos (y registramos) de manera más precisa (ver capítulo siguiente). En todo caso esto no evitará que los continuos procesos de cambio social y de mejora de los programas produzcan la emergencia de nuevas variables espurias.

Pero además existen otros procedimientos de control de variables, uno de ellos es la repetición de las medidas con otras audiencias, muestras o sujetos (algo relativamente fácil si utilizamos grupos de discusión o entrevistas), otro procedimiento tienen que ver con la complementariedad de métodos que hemos explicado en la parte II, un tercer procedimiento es la “triangulación” que explicamos en el capítulo 15. Pero el procedimiento más usual se refiere al uso de “grupos control”. Veamos en que consiste esta metodología.

#### **13.5.4.- Los grupos control y los diseños cuasi-experimentales.**

La utilización de grupos control es una metodología derivada de los procedimientos experimentales de las ciencias de la naturaleza. Se trata de realizar un diseño que incluye dos grupos equivalentes y comparar los resultados obtenidos. Por ejemplo, si ofrecemos en un laboratorio a una docena de ratas durante un mes, para comer, queso con guindillas u otro producto desagradable, además de galletas (grupo experimental), mientras que a otra docena de ratas se les da queso sin ningún aditivo y además galletas (grupo control) y, trascurrido el mes, se facilita a las primeras ratas queso sin guindillas pero ya sólo comen galletas, mientras que las segundas siguen tan aficionadas al mismo como siempre, podemos concluir que las ratas no discriminan las guindillas en la mezcla de queso con guindillas.

El ejemplo de las ratas y el queso con guindillas nos permite establecer tres posibles tipos de diseños de análisis de resultados con grupos control:

1. El más sencillo es el diseño pretest-postest, es decir se hace una medición antes de realizar la intervención y otra después para medir los cambios. En el caso de las ratas se trataría de determinar antes de la intervención el orden de preferencia de los tres productos: queso simple, queso con guindillas y galletas, y contrastar en el post como han variado sus preferencias. No hay que insistir mucho en esta técnica ya que es de conocimiento general en especial en los cursos de formación. En todo caso son muchos los autores que consideran que este no es un diseño experimental. De hecho no lo es. Pero su generalización en la evaluación de programas (especialmente en el ámbito educativo) nos lleva a considerarlo una modalidad cuasi-experimental.
2. El segundo es el diseño con grupo control neutral.
3. El más complejo es ya un experimento en sentido estricto e incluye grupos con diversas intervenciones alternativas.

Cuando el grupo experimental es el que recibe la intervención y el grupo control simplemente es el que no recibe la nada (neutral), el diseño es relativamente fácil, así en el ejemplo anterior del programa de promoción de la tolerancia intercultural podemos tratar de evaluar los resultados de las sesiones de sensibilización escolar midiendo el cambio de actitudes en las aulas y cursos

que han recibido la intervención y comparándolas con la misma medida en otras aulas y cursos que no la hayan recibido. El único problema importante es, en este caso, que los grupos sean verdaderamente equivalentes en términos muestrales. Lograda esta equivalencia tendríamos un diseño cuasi-experimental porque el grupo no forma parte del experimento, ya que sólo es un grupo supuestamente "neutral".

Porque los verdaderos diseños experimentales requieren además no tanto comparar los grupos experimentales con los grupos neutrales de control, sino contrastar diferencias entre varios tipos de intervención. Por ejemplo en un programa de vivienda un verdadero experimento sería establecer cuatro alternativas a un grupo idéntico (edad, actividad, ingresos, apoyos familiares,...) de jóvenes perfectamente seleccionados de forma aleatoria para formar parte de un grupo u otro:

- Alquiler subvencionado.
- VPO.
- Propiedad hipotecada.

Con un coste anual idéntico durante un largo periodo de años, midiendo las características de la respuesta y la continuidad de la misma en cada grupo.

Aparte de las evidentes dificultades metodológicas y materiales para realizar este experimento, aparece el problema de que esto no es legal, porque no pueden ofrecerse alternativas diferentes a jóvenes que tienen los mismos derechos.

Asimismo el verdadero diseño experimental en un ámbito social real exige además la práctica del "doble ciego" para evitar las proyecciones, tanto del investigador como de los investigados que son conscientes, de forma inevitable, de su condición. Un doble ciego se conforma a partir de cuatro grupos equivalentes y que respectivamente:

- Son sujetos experimentales y lo saben.
- Son sujetos experimentales y creen ser del grupo control
- Son sujetos del grupo control y lo saben.
- Son sujetos del grupo control y creen que son del experimental.

Además, en un diseño completo se debería engañar a los investigadores de tal manera que creen que la mitad de los miembros de cada grupo pertenecen al mismo pero la otra mitad al grupo antagonico. Además en un experimento de múltiples factores debería haber tantos grupos como factores a investigar. Hacer esto y conseguir las adecuadas equivalencias de representatividad, controlando además las variables espurias es hoy por hoy casi imposible. Aunque posiblemente por falta de recursos en el área de las políticas sociales. Por este motivo en la evaluación de resultados debemos limitarnos a los diseños cuasi-experimentales mencionados más arriba.

### **13.6.- Otros análisis más particulares.**

Existen otros tipos más particulares de evaluación con funciones más especializadas, que citamos a continuación, ya que pueden ser utilizados en alguna ocasión. Debemos aclarar que su condición marginal responde a la prioridad política y la centralidad de otras formas de evaluación, en particular la evaluación de resultados y de impacto. De hecho, se trata de tipos de análisis que ofrecen una información de alto valor, pero de difícil encaje sobre una lógica administrativa tradicional. En este sentido se consideraran formas de evaluación innovadoras (Roselló, 2004).

#### **13.6.1.- Análisis de la equidad.**

Trata de medir la distribución de los resultados obtenidos entre la población beneficiaria del proyecto. Se trata de una medida de justicia y de accesibilidad, es decir, si todos los beneficiarios

teóricos han sido beneficiarios reales y cuáles han sido las causas que explican una mala distribución de los mismos.

En términos cuantitativos el análisis de equidad trata de medir las desviaciones en la distribución sobre la “curva normal” (una “U” invertida). Éste es un concepto estadístico que establece que la mayoría de los casos se sitúa sobre la media teórica y su entorno, siendo nulos los valores extremos (por ejemplo los que “han obtenido muchos beneficios” o los que “no ha obtenido ningún beneficio”). Cuanto más se ajuste la distribución a la curva normal más equitativo habrá sido el programa.

En cambio, en términos cualitativos se trata de determinar la percepción entre los beneficiarios (y quizá algunos observadores) de que la distribución ha sido correcta y los componentes que producen desajustes han sido controlados. La perspectiva cuantitativa nos ofrece un dato objetivo (que podemos utilizar como contra-argumento), mientras que el relato cualitativo nos indica la percepción social del programa con independencia de cuales hayan sido sus efectos.

Conviene destacar, en todo caso, que podríamos haber obtenido la percepción social de forma cuantitativa (por ejemplo con una encuesta de satisfacción) y la distribución objetiva mediante un método cuantitativo (por ejemplo un Delphi).

La evaluación de la equidad es el contrapeso democrático de los tipos de evaluación más centrales (procesos y resultados) que interesan de forma prioritaria a los profesionales. Como no es fácil encontrar financiación para realizar evaluaciones de equidad, ni los profesionales se muestran muy entusiastas ante las evaluaciones de equidad, la tendencia general es que “la equidad”, como hemos explicado en el capítulo 2, se convierta en un componente transversal de todos los procedimientos y procesos de evaluación y que los evaluadores deben garantizar.

En el ámbito de las políticas sociales la cuestión de la falta de equidad ha sido muy aludida aunque poco estudiada. Se le conoce con el nombre de “efecto Mateo”, porque así lo formuló Robert K. Merton en alusión a una expresión del Evangelio de San Mateo “*Aquellos que más tienen más se les dará*”. En el ámbito de los Servicios Sociales los profesionales conocen bien el “efecto Mateo” y se supone que están orientados y preparados para eludirlo. Pero la existencia de evaluaciones de equidad facilitaría, con información más precisa, esta tarea. En el ámbito de Juventud la cuestión apenas ha sido tratada, en una gran medida porque el “efecto Mateo” es más que evidente y parece difícil evitarlo en un contexto de prestaciones y servicios de carácter universal.

### **13.6.2.- Análisis de la sensibilidad.**

Se trata de un tipo de evaluación realizada desde la perspectiva de los beneficiarios. Se trata de determinar, más allá del análisis de resultados e impacto, si el proyecto ha satisfecho las expectativas y demanda de los usuarios concretos. Es una medida muy individual (aunque el resultado puede presentarse en términos colectivos) y en general sólo se ha utilizado, para este tipo de evaluaciones, las encuestas de satisfacción.

Puede imaginarse en el futuro diferentes tipos de análisis de sensibilidad, en parte para realizar, de una forma más completa, análisis de impacto y en parte para facilitar el control democrático y administrativo de las intervenciones.

### **13.6.3.- Análisis de sostenibilidad.**

La noción de evaluación de la sostenibilidad se ha construido en los últimos años en respuesta a la aparición de un mercado de “programas evaluados y/o de éxito”, con una duración limitada y una supuesta garantía de resultados. Sobre los errores asociados a esta noción hemos hablado de forma amplia en el capítulo 2. Pero a la vez hay que reconocer que el mercado de este tipo de programas se ha consolidado, en parte porque vienen certificados por universidades norteamericanas con el supuesto prestigio de un diseño cuasi-experimental, en parte porque su coste refleja un “presupuesto cerrado y temporalmente limitado”, lo cual se ajusta a la dinámica

de las legislaturas políticas y los presupuestos anuales. Pero además, porque el “paquete” incluye, al menos en apariencia, todos los flecos de la intervención, lo que permite a las administraciones “olvidarse del peso de la complejidad y la concreción” a la que suelen aludir algunos profesionales y que pone nerviosos a algunos administradores y gestores.

La sostenibilidad es una noción tomada del campo de la economía y que se ha proyectado sobre la ecología (y los movimientos en torno a la globalización) y que se podría expresar como “la capacidad del proyecto para mantenerse y mantener sus resultados a lo largo del tiempo”. Una vez establecida esta definición se ha puesto en marcha una cierta reflexión en torno al hecho de que el logro de objetivos “para una sola ocasión” y el logro de objetivos de forma permanente quizá exigen diseños diferentes.

En el ámbito profesional la noción de sostenibilidad tiene muy buena prensa en una gran medida porque se asocia a estabilidad laboral, pero también a los sistemas de prestaciones sociales de mayor prestigio. También aparece habitualmente asociada, aunque no hay ninguna evidencia en este sentido, a “un mejor control de los gastos” y una reducción de los costes. A título muy personal considero que la evaluación de sostenibilidad responde, por encima otras formas de evaluación, a las exigencias éticas de la evaluación.

## **BIBLIOGRAFÍA Y DOCUMENTACIÓN COMPLEMENTARIA.**

- COHEN, E. y FRANCO, R, (1993), **Evaluación de Proyectos Sociales**, Madrid, Siglo XXI.
- VEDUNG, E. (1997), **Evaluación de políticas públicas y programas**, Madrid, MTAS.
- BALLART, X. (1992), **¿Cómo evaluar los programas y servicios públicos? Aproximación sistemática y estudio de caso**, Madrid, MAP.



## El informe de evaluación y la difusión de los resultados

Una evaluación sin su correspondiente informe es una evaluación inexistente. El simple conocimiento de los resultados de la evaluación por parte del equipo que realiza la intervención (o los responsables político-administrativos) implica una paradoja radical ya que niega la existencia de la propia evaluación. Una evaluación que no difunde sus resultados carece de entidad real, no tanto porque es desconocida para terceros, sino porque “no se ha completado”. Debemos recordar que Kant explicó y demostró que una evaluación que no ha sido difundida, que no ha sido contrastada por la comunidad de profesionales, investigadores y ciudadanos interesados, y que no ha sido codificada como tal, no se ha concluido y por tanto es una evaluación fallida.

En el anexo II, atendiendo a los estándares de evaluación del sistema de Naciones Unidas, la cuestión del Informe de evaluación ocupa una cuarta parte de texto, con 18 estándares y 36 ítems. En Naciones Unidas se le da mucha importancia al informe justamente porque el objetivo de la difusión es algo central para este organismo (como lo debería ser para todos). Los estándares descritos son tan completos y están tan bien estructurados que nos podríamos sino limitarnos a transcribirlos.

En este sentido tales estándares deberían ser obligatorios para cualquier evaluación, y una vez hecha esta afirmación no tendríamos que decir nada, pero con el fin de completarlos vamos a añadir algunas interpretaciones.

### 14.1.- El informe de evaluación: contenidos.

El informe de evaluación (o informe final) se realiza una vez concluida la fase de análisis. Se trata de un informe único que una vez elaborado puede segregarse en varios formatos. Por ejemplo, un resumen ejecutivo, un resumen para los usuarios del programa o una nota de prensa para los MCM. También se pueden preparar artículos científicos para difundirlo entre los profesionales y adquirir reputación.

El informe final de la evaluación tendrá siempre un destinatario formal principal, que en general será una unidad administrativa, que en ningún caso será diferente de aquella que haya realizado, promocionado o financiado el programa, el plan o la intervención. En el caso de los programas y políticas de juventud serán los organismos públicos de juventud. No es correcto que

el destinatario formal sea otro organismo (salvo el poder legislativo cuando así se establezca en el propio Plan o programa), aunque se podrán entregar con posterioridad copias o resúmenes a otros organismos o agentes sociales. La mala costumbre de entregar las evaluaciones originales a organismos diferentes al que ha realizado el programa (y con el que se supone mantienen una relación jerárquica) debe ser erradicada.

El contenido del informe final debería seguir las instrucciones de los estándares de evaluación del sistema de Naciones Unidas que aparecen en el anexo II. Pero para concretarlo describo el informe final de una evaluación realizada en paralelo a este Manual para una Fundación que desarrolla un programa educativo y que en este preciso momento está sobre mi mesa pendiente de entregar.

El contenido es el siguiente:

1. Ha sido escrito en “Tahoma”, con un tamaño de fuente de 12’ y a 1 espacio.
2. Incluye un Resumen Ejecutivo de tres páginas.
3. El informe contiene (1) un índice; (2) una introducción de 14 páginas en la que se describe el contexto y las características del programa y sus objetivos; (3) un capítulo de 22 páginas en el que se describe el planteamiento y los procedimientos seguidos en la evaluación; (4) un segundo capítulo de 34 páginas en el que se describen los resultados obtenidos; (5) otro capítulo de 7 páginas en el que aparecen, bien justificadas, 5 fortalezas y/o oportunidades, 5 debilidades y/o amenazas y 5 recomendaciones; (6) un capítulo de cinco páginas con las conclusiones (diferente del resumen ejecutivo, aunque en ocasiones puede ser lo mismo); (7) bibliografía y (8) los anexos, que en este caso son muy amplios, ya que incluyen el programa completo, un cuestionario y la explotación de los marginales.

Se trata de un “borrador” que se va a presentar impreso en papel y en archivo electrónico (concretamente en Word y con un archivo distinto para el resumen ejecutivo, otro para los anexos y otro para el Informe en sí).

Dicho borrador va a ser discutido con la entidad (y en este caso con los profesionales que han desarrollado el programa). La discusión no modificará la esencia del informe, aunque se pueden introducir cambios que impliquen mejoras en la información obtenida o en el análisis realizado. En el caso improbable de que se produjeran desacuerdos básicos o los profesionales rechazaran como equivocadas las conclusiones, la evaluación habría fracasado y debería volver a reconsiderarse. Esta opción no impide que en ocasiones, especialmente cuando se detectan problemas éticos, el evaluador no pueda mantener una posición firme y crítica con el programa. En cualquier caso, sólo se llega a este extremo cuando de la evaluación se derivan consecuencias legales para aquellos que realizan la intervención.

En todo este proceso nos hemos orientado buscando la máxima repercusión de la evaluación, en un doble sentido de utilidad para la Fundación y los profesionales del equipo de trabajo y de proyección de las tareas de la Fundación hacia otros agentes sociales. La noción de repercusión no puede ser ni obviada ni enmascarada en la realización del informe (Mondragón, 2004).

Una vez establecido el texto definitivo, se preparan los procedimientos de difusión de dicho Informe.

## **14.2.- Publicación, distribución, registros y bases de datos.**

El informe puede ser publicado formalmente (es decir con ISBN) de forma completa o parcial. También puede publicarse una parte del mismo, el resumen ejecutivo o un artículo científico. En todo caso hay que tener en cuenta que la publicación formal retrasa la difusión.

Haya o no publicación formal hay que realizar una difusión del mismo una vez concluido el informe. Tal difusión se puede realizar a partir de un documento “manuscrito” que los actuales programas informáticos editan de una forma muy profesional.

Esta primera difusión implica sesiones de presentación a los agentes sociales, profesionales y ciudadanos involucrados en el programa, con un cierto despliegue de medios de presentación, como diapositivas de los principales resultados y alguna síntesis de los resultados cuando nos dirigimos a la población general. La presentación es siempre una tarea que exige un buen diseño: respondiendo a preguntas como ¿qué, quién, cómo, cuándo, dónde y por qué?, la elaboración del material de apoyo como material escrito, filminas, acetatos o video. Todo ellos de acuerdo a la programación prevista en el cronograma de la evaluación.

Tanto este primer informe como la publicación formal, aunque sólo lo segundo si es inminente, debe ser convenientemente distribuido entre los organismos de juventud nacionales, autonómicos, los locales más relevantes y los locales de la propia región. También a las bibliotecas de las universidades y a los departamentos especializados en el tema. Asimismo conviene remitírselo a profesionales relevantes. Si la publicación es electrónica y está sólo colgada de la página web, conviene preparar unos CDs (es muy barato) para realizar esta distribución, o al menos comunicar que la evaluación ha sido colgada en la página web.

Las publicaciones formales, o aquellas que están en formato CD, tienen la ventaja de que figuran en los registros y bases de datos, así como en las bibliografías especializadas. En la actualidad, colgarlo en la página web sin más tiene la desventaja de que, salvo aquellos que manejan habitualmente dicha página, no se difunde, ya que en general están en páginas institucionales que resultan un tanto opacas a los buscadores. Parece previsible que en el futuro se creen registros y bases de datos electrónicos, pero por ahora, al menos en el ámbito de juventud, son sólo proyectos y en Internet sólo aparecen páginas propias con documentación propia.

Obviamente, la publicación de un artículo científico resuelve estos problemas, ya que las revistas científicas tienen establecido un sistema de difusión y registro muy eficaz. En todo caso, conviene que este artículo científico (y la propia publicación) incluya toda aquella bibliografía que se haya generado en el contexto del programa y en otros programas similares en el ámbito español.

Esta actitud en relación a la bibliografía es, además, la principal diferencia entre aquellos países que se supone son punteros en investigación (y evaluación) y los países como España. Los primeros citan de forma masiva los productos del organismo, el entorno y las evaluaciones de programas similares, mientras que, en general, en España apenas se cita la bibliografía propia (y del conjunto del país) y se recurre a bibliografía externa, en particular norteamericana, para tratar de justificar “que se conoce y hemos hecho un trabajo bien fundado”. No cabe duda de que esta bibliografía es necesaria cuando explicamos los orígenes y la evolución histórica de muchos temas. Pero en la actualidad el “método de las referencias cruzadas” sirve para difundir (y aumentar el prestigio) de las evaluaciones realizadas por autores anglosajones, mientras que “el método de olvidarse de las aportaciones de los similares” invisibiliza nuestras producciones, impidiendo que se difundan de una manera adecuada.

### **14.3.- La meta-evaluación.**

Desde la primera formulación realizada por Scriven en 1969, este concepto ha ido perfilándose gracias a las aportaciones de éste y otros autores. A pesar de que, en los últimos años, ha conocido un auge considerable, su desarrollo no es en modo alguno comparable al de la simple evaluación. Prueba de ello es que algunas de las meta-evaluaciones más significativas que se han llevado a cabo hasta la fecha no son sino aplicaciones de modelos teóricos elaborados para la realización de evaluaciones (House, 1977; Stake, 1986).

El propósito de quienes han propuesto realizar meta-evaluaciones es, ante todo, encontrar una herramienta que permita valorar el proceso de evaluación y verificar el grado de validez de sus resultados y conclusiones. La evaluación, o mejor un conjunto de evaluaciones, se convierte, de esta forma, en el objeto de estudio, en una evaluación de evaluaciones.

Quizá el escaso desarrollo de la meta-evaluación tenga que ver con los problemas de aplicación práctica, porque: ¿qué se entiende por “bondad, valor o méritos” de una evaluación?, ¿cuáles son los criterios conforme a los cuales ha de juzgarse una evaluación? Las respuestas que ofrecen a esta cuestión, la perspectiva de la evaluación fuerte frente a las corrientes más propias de la ciencia experimental, son bien distintas.

En líneas generales, las corrientes científicas o positivistas tienden a hacer hincapié en los aspectos técnico-metodológicos y, en concreto, a hacer de las técnicas cuantitativas la clave esencial de la evaluación. A juicio de los representantes de estas corrientes, el valor de una evaluación sólo queda plenamente garantizado con el empleo de procedimientos acumulativos de recogida y el tratamiento de datos cuantitativos en el marco de un diseño de naturaleza experimental. La meta-evaluación tiene, pues, como objetivo calibrar en qué medida las diferentes evaluaciones se aproximan al “modelo científico” preestablecido. Como consecuencia son muchos los científicos que opinan que por meta-evaluación sólo debemos entender una acumulación de resultados coincidentes que refuerzan la evidencia. Es decir, si el programa se reproduce diez veces, se evalúa diez veces y las diez veces arroja el mismo resultado, esto es una evidencia consistente. El modelo clásico de este proceder aparece en la farmacología y sirve para garantizar los efectos de los mismos. Pero la sociedad no es un organismo biológico estable aunque, en ocasiones, se utilicen metáforas orgánicas.

En resumen, la ciencia trata de saber sobre el funcionamiento del mundo, mientras que en el terreno de la evaluación fuerte las cosas son muy diferentes: se trata de saber sobre la eficacia o la eficiencia de un programa o una intervención concreta, por tanto la meta-evaluación es una comparación en torno a la eficiencia de un conjunto de evaluaciones que valoran programas.

Por este motivo la meta-evaluación fue definida como “*el proceso de diseño, obtención y empleo de información descriptiva y valorativa acerca de la utilidad, la viabilidad, la ética y el rigor técnico de una evaluación con el propósito de orientarla y dar cuenta públicamente de sus puntos fuertes y de sus puntos débiles*” (Stufflebeam 1981). Para lograrlo, de entrada ya fue el propio Michael Scriven quien al formular el concepto explicó que la meta-evaluación puede servirse no sólo de aquellos estándares que han sido específicamente formulados para la misma, sino también de aquellos otros que han sido concebidos para la realización de evaluaciones. De esta consideración surgieron sus dos propuestas: la *Meta-Evaluation Checklist* y la *Key Evaluation Checklist*.

La *Meta-Evaluation Checklist* (MEC) incluye una serie de ítems que el autor obtiene de diversas fuentes, sobre el deber ser del estudio evaluativo. En la MEC se considera que la evaluación ha de explicitar y justificar los estándares de valor utilizados y ser comprensiva, comprensible, creíble, ética, factible, suficientemente precisa desde el punto de vista técnico -pero sin que ello haga olvidar cuestiones como la relevancia de la información proporcionada-, “sensible” desde el punto de vista político -esto es, que atienda y responda a las necesidades de las distintas audiencias-, oportuna, segura, válida y rentable en términos de coste-eficacia.

La *Key Evaluation Checklist* (KEC), en principio elaborada como referencia para la realización de evaluaciones, puede también resultar de utilidad para las metaevaluaciones. La KEC no debe ser considerada como una lista de *desiderata*, sino más bien como un catálogo de cuestiones que deben ser analizadas en el proceso de la misma. Dichas cuestiones se refieren al diseño del estudio, el personal de la evaluación, la procedencia y la validez de los criterios empleados incluso para la elección de los evaluadores, los antecedentes y el contexto de la evaluación, su calendario, la identificación de las audiencias, la determinación de los costes y recursos, la valoración del proceso de investigación y de sus resultados, las alternativas existentes a la evaluación, la posibilidad de generalización de ésta y, por último, la significación global de la investigación evaluativa. Asimismo, en la KEC se contempla la posibilidad de que en el informe se incluyan recomendaciones sobre el uso o el destino de la evaluación (Scriven, 1984).

En el fondo las propuestas de estándares de evaluación (como las que hemos recogido en los anexos) son procedimientos de meta-evaluación, ya que nos permiten “evaluar las evaluaciones”.

Se trata además de estándares acreditados y que han sido propuestos por amplias comisiones después de años de trabajo.

El problema es que este tipo de meta-evaluación basada en estándares sirve en ocasiones para justificar políticas de subvenciones y préstamos, como es común en el Banco Mundial o la Comisión Europea, olvidando que la evaluación debe responder, de forma esencial, a los intereses y necesidades de aquellos que realizan la intervención o reciben los cuidados, así como, aparte de los requerimientos de carácter técnico y metodológico, a las necesidades de información tanto de aquellos que van a tomar decisiones, como del público en general.

Debemos asumir, por tanto, que la meta-evaluación es la lógica culminación analítica de los procesos de evaluación. De la misma forma que las evaluaciones pueden ser consideradas como la conclusión analítica de los programas, los estudios de meta-evaluación han de ser entendidos como el punto final, la culminación de esos procesos de análisis que son las evaluaciones. Ello no quiere decir que toda evaluación deba finalizar necesariamente con un proceso meta-evaluativo, aunque parezca aconsejable tanto desde un punto de vista técnico como político.

Porque la meta-evaluación, al identificar y establecer tanto las cuestiones que han de analizarse en una evaluación como los criterios conforme a los cuales ha de juzgarse su calidad, se convierte en un instrumento imprescindible para el progreso y la consolidación de la actividad profesional de la evaluación.

Finalmente, la meta-evaluación constituye una excelente herramienta para conocer los estilos de evaluar de los poderes públicos: el análisis de las evaluaciones de los programas de juventud realizados en España nos ha proporcionado información de extraordinaria utilidad sobre las actitudes y el comportamiento de las diversas administraciones en las políticas de juventud (Comas, 2007). Así, por ejemplo, hemos podido conocer los enfoques teóricos por los que se han venido decantando, la importancia que han atribuido a ciertas cuestiones, la propiedad de la información producida en la evaluación o la utilidad y la difusión de los resultados.

#### 14.4.- Tipos peculiares de evaluación: la evaluación formativa.

Aparte de todas las formas de evaluación citadas en el capítulo anterior (y en éste) existen otras formas particulares de evaluación, por ejemplo la **evaluación sumativa** (una especie de meta-evaluación que acumula todas las evaluaciones del mismo programa para determinar el logro global de los objetivos o valorar la totalidad del programa); la **evaluación explicativa** que se parece a la investigación porque se utiliza para programas que incluyen hipótesis (se trata de comprobarlas); la **evaluación de auditoría** (eficiencia particular de cada gasto) y la **auto-evaluación**, que es la estrategia que se utiliza en el sistema universitario español como una forma de respetar su autonomía y libertad de cátedra (Mora, 1991). También existen modelos de evaluación que presentan problemas éticos y cuya práctica se suele rechazar, entre otras razones por las consecuencias negativas y las sospechas que su utilización tiene para la propia evaluación, entre tales modelos destacan las **evaluaciones encubiertas** (por ejemplo para evaluar a una entidad o a un equipo sin que este lo sepa para tomar decisiones sobre dicho equipo) y **las evaluaciones de relaciones públicas**, que básicamente consiste en diseñar programas cuyos resultados conocemos o prevemos de antemano para dar prestigio a la entidad o al programa (Aguilar y Ander-Egg, 1992).

De todas estas formas particulares de evaluación queremos destacar la **evaluación formativa**, y lo hacemos en este último y particular apartado porque resulta difícil de incluir en los anteriores.

La evaluación formativa se mueve en el territorio de la evaluación de implementación y procesos, pero se diferencia de estas porque sirve para tomar decisiones **durante** la realización del programa. No es un tipo de evaluación demasiado recomendable, salvo en los casos particulares:

- Casos en los que ha sido imposible o muy difícil hacer un análisis previo de la realidad y muchos aspectos de la misma no serán conocidos hasta que se comienza el programa, por ejemplo, no sabemos si hay necesidades o demanda en relación a un servicio.
- Programas de muy larga duración, que fueron planificados de acuerdo a ciertos parámetros que pueden variar en el tiempo, por ejemplo eran muy pocos los estudiantes que aceptaban trabajos de verano, pero ahora son muchos.

La evaluación formativa implica un alto grado de participación del equipo que realiza la intervención y una estructura predefinida para la toma de decisiones sobre los cambios que deben realizarse en el programa. Debe ser continua e integral y debe medir los efectos de la retroalimentación de estos cambios.

Introduce además un riesgo: una evaluación formativa cambia, de forma más o menos constante las condiciones de aplicación, lo que a su vez implica que introduce dificultades para aplicar los otros tipos de evaluación.

## **BIBLIOGRAFÍA Y DOCUMENTACIÓN COMPLEMENTARIA.**

- SIMONS, H (1999), **Evaluación democrática de instituciones escolares**, Madrid, Morata. El capítulo VI: *“El derecho a conocer: una disputa sobre la divulgación de un informe de evaluación”*.

### 15.1.- La evaluación como acto creativo.

En la parte segunda de Manual hemos descrito las técnicas que ocupan los nueve puestos principales en la tabla redonda de los métodos de investigación. Pero la mesa es muy grande y permite que se sienten en la misma otros procedimientos de recogida y producción de datos. Técnicas peculiares que se adecuan a necesidades muy particulares, o combinaciones metodológicas que resaltan la imaginación (y la preparación) del investigador o el evaluador.

En ocasiones sirven para completar o contrastar la evaluación cuando por sí mismas no suponen una evaluación completa. Entre ellas, en los ámbitos de juventud, podemos destacar las técnicas participativas, que se ajustan a la misión constitucional de los organismos de juventud: promover la participación de los jóvenes (artículo 48 de la Constitución Española).

La combinación entre técnicas y posibles modelos de evaluación genera una infinita gama de posibilidades de actuación. En un reciente trabajo aplicado al ámbito educativo (Zufiaurre y Albertín, 2006), describen y sugieren aplicar múltiples modelos de evaluación, que incluyen diferentes tipos de paneles, de estudios de caso, la llamada “investigación instructiva”, la “evaluación iluminativa”, la “evaluación de prototipos”, la llamada “experiencia vicaria” y otra docena de modelos de los que realizan someras descripciones, con referencias bibliográficas de cada uno de ellos. El detalle de cada uno de estos modelos muestra que muchos de ellos son “adaptaciones y combinaciones”, tanto de métodos de recogida de datos como de procedimientos de evaluación (y de los métodos con los procedimientos).

Es la imaginación combinada con **las necesidades concretas** del ámbito educativo la que ha propiciado toda esta creación de opciones. Cualquier otro ámbito, si de verdad quiere evaluar también tiene que ser imaginativo y adaptarse a sus necesidades. En diferentes momentos del texto hemos hablado de la evaluación como una tarea científica pero que desarrollan “artesanos”. Podemos ahora añadir unos párrafos seleccionados que escribió en su día Wright Mills (1959) en el apéndice sobre “artesanía intelectual” en “**La imaginación sociológica**”, Madrid, FCE, 1993.

*“Para el investigador social individual que se siente como parte de la tradición clásica, la ciencia social es la práctica de un oficio. En cuanto hombre que trabaja sobre problemas esenciales, figura entre los que rápidamente se impacientan y se cansan de discusiones complicadas sobre método-y-teoría-en-general, que interrumpen sus propios estudios. Cree que*

*es mucho mejor la información de un estudioso activo acerca de cómo procede en su trabajo que una docena de "codificaciones de procedimiento" hechas por especialistas que quizá no han realizado ningún trabajo de importancia. Únicamente mediante conversaciones en que pensadores experimentados intercambien información acerca de su manera real de trabajar puede comunicarse al estudiante novel un concepto útil del método y de la teoría. Por lo tanto, creo útil referir con algún detalle cómo procedo en mi oficio. Esto es, inevitablemente, una declaración personal, pero está escrita con la esperanza de que otros, en especial los que se inician un trabajo independiente, la harán menos personal por los hechos de su propia experiencia.*

*Creo que lo mejor es empezar por recordar a los estudiantes principiantes que los pensadores más admirables de la comunidad escolar a que habéis decidido asociarnos no separan su trabajo de sus vidas. Parecen tomar ambas cosas demasiado en serio para permitirse tal disociación y desean emplear cada una de ellas para enriquecer a la otra. Desde luego, esa decisión es la convención que prevalece entre los hombres en general, y se deriva, supongo yo, del vacío del trabajo que los hombres en general hacen hoy. Pero habréis advertido que, como estudiantes, tenéis la excepcional oportunidad de proyectar un tipo de vida que estimule los hábitos de la buena artesanía. El trabajo intelectual es la elección de un tipo de vida tanto como de una carrera; sépalo o no, el trabajador intelectual forma su propio yo a medida que trabaja por perfeccionarse en su oficio; para realizar sus propias potencialidades y aprovechar las oportunidades que se ofrezcan en su camino, forma un carácter que tiene como núcleo las cualidades del buen trabajador.*

*Lo que significa esto es que debéis aprender a usar vuestra, experiencia de la vida en vuestro trabajo intelectual, examinándola e interpretándola sin cesar. En este sentido la artesanía es vuestro propio centro y estáis personalmente complicados en todo producto intelectual sobre el cual podáis trabajar. Decir que podéis "tener experiencia" significa, entre otras cosas, que vuestro pasado influye en vuestro presente y lo afecta, y que él define vuestra capacidad para futuras experiencias. Como investigadores sociales, tenéis que dirigir esa complicada acción recíproca, captar lo que experimentáis y seleccionarlo; sólo de esa manera podéis, esperar usarlo para guiar y poner a prueba vuestro pensamiento, y en ese proceso formaros como trabajadores intelectuales. Pero, ¿cómo podréis hacerlo?*

*Una solución es: debéis organizar un archivo, lo cual es, supongo yo, un modo de decir típico de sociólogo: llevad un diario. Muchos escritores creadores llevan diarios; la necesidad de pensamiento sistemático que siente el sociólogo lo exige.*

*En el archivo que voy a describir, están juntas la experiencia personal y las actividades profesionales, los estudios en marcha y los estudios en proyecto. En ese archivo, vosotros, como trabajadores intelectuales, procuraréis reunir lo que estáis haciendo intelectualmente y lo que estáis experimentando como personas. No temáis emplear vuestra experiencia y relacionarla directamente con el trabajo en marcha. Al servir como freno de trabajo reiterativo, vuestro archivo os permite también conservar vuestras energías. Asimismo, os estimula a captar "ideas marginales": ideas diversas que pueden ser sub-productos de la vida diaria, fragmentos de conversaciones oídas casualmente en la calle, o hasta sueños. Una vez anotadas, esas cosas pueden llevar a un pensamiento más sistemático así como prestar valor intelectual a la experiencia más directa.*

*Habréis advertido muchas veces con cuánto cuidado tratan sus propias inteligencias pensadores consumados, y cuán atentamente observan su desarrollo y organizan su experiencia. La razón de que atesoren sus menores experiencias es que, en el curso de una vida, el hombre moderno tiene muy poca experiencia personal, y sin embargo la experiencia es sumamente importante como fuente de trabajo intelectual original. He llegado a creer que el ser fiel a su experiencia sin fiarse demasiado de ella es una señal de madurez del trabajador. Esa confianza ambigua es indispensable para la originalidad en todo trabajo intelectual, y el archivo es un medio por el que podéis desarrollar y justificar tal confianza.*

*Llevando un archivo adecuado y desarrollando de ese modo hábitos de auto-reflexión, aprendéis a mantener despierto vuestro mundo interior. Siempre que os impresionen fuertemente sucesos o ideas, no debéis dejarlos irse de vuestra mente, antes al contrario, debéis formularlos para vuestro archivo y, al hacerlo, desentrañar todo lo que implican, y demostramos a vosotros mismos la insensatez de aquellos sentimientos o ideas o la posibilidad de articularlos en forma productiva. El archivo os ayuda también a formaros el hábito de escribir. No podéis tener la “mano diestra” si no escribís algo por lo menos cada semana. Desarrollando el archivo, podéis tener experiencia de escritores y cultivar, como suele decirse, vuestros medios de expresión. Llevar un archivo es controlar la experiencia.*

*Una de las peores cosas que les suceden a los investigadores sociales es que sienten la necesidad de escribir sus “planes” sólo en una ocasión: cuando van a pedir dinero para una investigación específica o para “un proyecto. La mayor parte de los “planes” se escriben para pedir fondos, o por lo menos se redactan cuidadosamente para ese fin. Aunque esta práctica está muy generalizada, la considero muy mala: está condenada a convertirse, por lo menos en cierta medida, en un “arte de vender” y, dadas las expectativas que hoy prevalecen, en acabar muy probablemente en afanosas pretensiones; el proyecto quizá va a ser “presentado” después de redondearlo de una manera arbitraria mucho antes de lo que debiera; muchas veces es una cosa amañada, destinada a conseguir dinero para fines diferentes, aunque valiosos, de los de la investigación ofrecida. Un investigador social que trabaja debe revisar periódicamente “el estado de mis planes y problemas”. Un joven, precisamente al comienzo de su trabajo independiente, debe reflexionar acerca de esto, pero no puede esperarse -ni lo esperará él mismo- que vaya muy lejos con eso, y evidentemente no debe entregarse con excesiva rigidez a ningún plan. Todo lo que puede hacer es orientar su tesis, que infortunadamente se supone ser su primer trabajo independiente de alguna extensión. Cuando estéis a la mitad del tiempo de que disponéis para el trabajo, o en su tercera parte, es cuando esa revisión puede ser más fructuosa y hasta quizá interesante para los demás”....*

Después de traducir sociólogo por evaluador y aparte de las referencias negativas a los “procedimientos”, que hoy en día tenemos que leer en una perspectiva de complementariedad, así como el hecho de que muchas de las tareas sugeridas por Mills nos las proporcionan actualmente las TIC, este texto (que se puede leer al completo) nos orienta de una manera clara y toda persona que quiera evaluar debe interiorizarlo.

## **15.2.- Las técnicas participativas.**

Se trata de técnicas específicas del modelo de la IAP (Investigación-Acción-Participación), de las que apenas hay experiencia en el ámbito de la evaluación, aunque como hemos visto la idea de la participación envuelve de forma transversal toda la “teoría de la evaluación fuerte” que hemos desarrollado en este Manual. Por este motivo, aunque todo este Manual aparece de forma trasversal, la idea de una evaluación participativa también parece adecuada incluir de forma concreta en alguna de estas técnicas, en especial por si resulta necesario complementar una evaluación que no ha contado con ningún impulso participativo.

Debe quedar claro que el recurso a las técnicas participativas refleja algún tipo de carencia en la evaluación, porque la participación debe formar parte del diseño de la misma y no debe ser un “añadido”. Además, muchas de las técnicas participativas, por ejemplo el socioanálisis o el análisis de redes, no parecen muy adecuadas para la evaluación de programas de juventud porque requieren la presencia de grupos diana muy cerrados y definidos y las Políticas de Juventud, en cuanto que escenifican derechos sociales, siempre se aplican a poblaciones generales, aunque en algunos programas estén segmentadas en función de determinadas características.

### 15.2.1.- Los talleres de contraste de resultados<sup>32</sup>.

Se trata de una estrategia participativa que puede realizarse tras obtener los resultados de la evaluación y que tiene como finalidad “contrastar el grado de calidad, la bondad o la certeza” de los resultados obtenidos.

El contraste se puede realizar utilizando dos vías diferentes, de una parte la de los expertos y técnicos externos (lo que nos lleva hacia las “técnicas de consultoría” que describimos en el apartado siguiente), y de otra parte el “contraste por devolución”, que implica que la tarea va a ser realizada por personas procedentes del grupo diana de la intervención o que han participado, como agentes informantes, en la evaluación.

Por esto, lo primero que hay que decidir en esta técnica es con quién se realiza esta tarea de devolución, existiendo diversas posibilidades, desde el conjunto de la población objeto de la intervención hasta sólo los agentes que han participado en la evaluación.

- A) A LA POBLACIÓN GENERAL DIANA. Si por población general entendemos los participantes en el programa y que forman parte de la población diana, aparecen muchas dificultades para obtener una muestra representativa, que además sea capaz de entender el sentido de la evaluación y capaces, también, de articular una respuesta crítica. Todo ello requiere un tiempo de selección largo y complejo. Además, requiere un tiempo de información y preparación para los participantes en estos talleres. Asimismo, hay que diferenciar, dentro de esta población diana, los que han recibido el servicio o han participado en el programa, de aquellos que simplemente han oído hablar del mismo, de aquellos que no tienen ni idea de que exista. Unos y otros nos aportaran visiones distintas.
- B) AGENTES PARTICIPANTES EN LOS PROCESOS DE EVALUACIÓN. Si se trata, de forma estricta, de aquellos que han participado de forma directa en la evaluación, los problemas metodológicos son menores, pero en este caso se corre el riesgo de que no aporten ninguna novedad a la evaluación, especialmente si se ha hecho con corrección y cuidado.

La segunda decisión a tomar implica una nueva pregunta: ¿Qué es lo que se va a devolver y cómo se va a plantear la evaluación? Es decir, ¿qué les vamos a contar a los participantes y cómo vamos a recoger sus respuestas y las vamos a analizar?

- A) Podemos entregarles el propio texto con la descripción y resultados de la evaluación y después pedirles un informe.
- B) Podemos hacerles una serie de preguntas (o temas) concretos que aparecen en la evaluación y que se quieren contrastar. En este caso las respuestas se pueden recoger en un cuestionario.
- C) Se puede realizar un juego de rol (o un sociodrama) para contrastar las opiniones, lo que exige la presencia de un analista que interprete lo que está ocurriendo.

En los tres casos se exige una planificación y un esfuerzo previo de organización, pero en ocasiones los resultados lo valen.

La tercera decisión se refiere al objetivo de esta “contraste de resultados”:

- A. Medir el grado de acuerdo o de desacuerdo con las conclusiones obtenidas.
- B. Proponer alternativas a las interpretaciones de los datos.
- C. Detectar carencias o errores metodológicos.

### UN EJEMPLO DE TALLER DE CONTRASTE DE RESULTADOS.

Mientras finalizo este Manual, estoy realizando una evaluación de unas prácticas sociales muy controvertidas realizadas por ciertas ONGs, de las que no voy a

<sup>32</sup> En los textos de referencia se denominan “talleres de contrastación de resultados”, pero “contrastación” es una palabra que no existe en el DRAE, ni ha sido admitida por otro diccionario.

proporcionar ningún dato para que no sean identificadas (no son del ámbito de juventud). En la evaluación he obtenido datos que confirman que dichas prácticas son socialmente necesarias y además estoy manejando la hipótesis de que responden a la falta de respuesta de las administraciones públicas ante tales necesidades. Pero a la vez la mayoría de los profesionales (y de la literatura técnica) considera que no son correctas.

He pensado en celebrar dos talleres de contraste de resultados. En el primero voy a elegir a siete profesionales externos a estas ONGs, pero que tengan alguna relación sectorial con las mismas. No voy a darles el resultado de la evaluación sino que les voy a contar de manera general algunas de mis conclusiones. Después voy a abrir una discusión siguiendo la metodología del grupo de discusión con los mismos. Así podré valorar si los desacuerdos son razonables u obedecen a estereotipos, también podré valorar si el estrecho vínculo (y una cierta empatía por la labor que realizan) hacia los evaluados no han enturbiado mi objetividad. Finalmente el Grupo me dirá si he omitido algún aspecto de la evaluación que debería haber tenido en cuenta. Voy a hacer que me cuenten sobre casos concretos, para visualizar los discursos alternativos de unos y otros.

El segundo taller lo voy a realizar con los propios evaluados tras entregarles un documento provisional con la evaluación (pero sin las aportaciones del taller anterior). En este caso resulta impensable que discutan entre sí (aunque podría ocurrir) pero se les pedirá que se manifiesten en relación al informe provisional. Tendré así en cuenta sus críticas, sus acuerdos, su grado de satisfacción (quizá si es mucho puedo haberme equivocado), los fallos que perciben, los datos que se deberían matizar. Hacer este taller me parece una necesidad técnica y ética, porque no pretendo ser de ninguna manera un “agente de lo institucional o de lo correcto”, sino alguien que les ayude a mejorar lo que están haciendo y si, como consecuencia, pueden normalizar y acreditar su trabajo con las instituciones y otros profesionales pues mejor. Pero es algo que tienen que decidir y hacer ellos.

Ambos talleres me van a permitir entregar el informe (en este caso a la administración pública) con un cierto grado de seguridad. Tanto en el caso de que desdiga el estereotipo como en el caso de que lo confirme.

### **15.2.2.- Las técnicas de consulta externa.**

Las técnicas de consulta externa requieren contar con figuras a las que se supone alguna habilidad o conocimiento, que no forman parte del programa o la evaluación, pero que bien desde el primer momento o bien desde cualquier punto de la evaluación (incluyendo la revisión del informe final) nos pueden aportar su opinión y sugerencias. En algunos casos también pueden contribuir como supervisores a la planificación y la buena marcha de la evaluación.

Asimismo, en el capítulo 2, al hablar de las relaciones entre evaluación y calidad, se han mostrado una serie de estrategias que permiten plantear sinergias entre ambos conceptos. La orientación del Plan de Calidad puede ser considerada como una consulta externa por parte la evaluación y viceversa.

La más conocida de las técnicas de consulta externa es el método DELPHI al que ya le hemos dedicado el capítulo 10, aunque entonces enfocado desde la mediación. Pero hay otras técnicas.

### **LA CONSULTA BIBLIOGRÁFICA GENERAL.**

A través de Internet, de las bases de datos y de las bibliotecas especializadas podemos acceder con facilidad a toda la bibliografía que haya generado un tema en particular.

Se trata de revisarla y consultarla hasta donde podamos. En esta revisión y consulta podemos

ver otras estrategias y resultados, análisis e hipótesis y por supuesto conclusiones. Podremos observar además los desacuerdos. Así podremos situar nuestra evaluación en un campo de conocimiento que incluya, de la tabla 2.1 de Mario Bunge; los contenidos C (la comunidad de sujetos que conocen Cc), D (el dominio, universo o territorio del discurso de Cc) y por supuesto los parámetros K y M.

Es un método de consulta externa porque la bibliografía es de otros autores y además se puede encargar la recopilación aun tercero.

#### **LA INTERPRETACIÓN TEÓRICA O PARADIGMÁTICA.**

Es lo mismo que lo anterior pero mucho más complejo porque exige una cierta especialización en el Campo de Conocimiento que sustenta la intervención. De hecho requiere conocer a fondo la bibliografía y esto significa una cierta dedicación durante largos periodos de la vida al tema.

La idea es situar la evaluación realizada en el campo cognitivo que suponen las diferentes teorías y paradigmas que existen sobre el tema. Aunque en ocasiones puede ser una teoría única, en cuyo caso se mediará el grado de ajuste.

Si el evaluador, o alguien del equipo, posee este background puede realizar esta tarea, sino puede encargarlo a agentes externos. Pero para hacer esto hay poca gente, no es fácil que alguien se comprometa y los teóricos preferían escribir directamente sobre teoría que contrastar evaluaciones.

#### **EL RECURSO A ASESORES ESPECIALIZADOS EN ALGUNA TEMÁTICA.**

El lo mismo que la opción anterior, pero se refiere más bien a profesionales especializados que a teóricos. En la práctica la mejor forma de hacerlo es el método Delphi. Pero si solo contamos con unos pocos asesores posibles (o con uno) la idea es un tipo de contacto de otra naturaleza.

#### **LA EXPOSICIÓN A UN PÚBLICO ILUSTRADO.**

Es una técnica muy similar a los talleres de contraste y también se parece un poco al método Delphi. En España, en el pasado cuando no se conocían otras formas de evaluar, era una manera muy habitual de contrastar un programa. Lo diseñabas o lo realizabas y después se reunía un grupo de iguales, de profesionales, de académicos e incluso de “intelectuales”, se explicaba el proyecto y su desarrollo (no era frecuente ni fácil que hubiera informes o publicaciones) y a partir de la explicación se realizaba un debate abierto sobre como mejorarlo, al tiempo que se transmitían hallazgos y experiencias.

En la actualidad es un trabajo muy poco común. Creo, en una gran medida, porque hay más desconfianza entre iguales, porque, como consecuencia, somos menos tolerantes con las críticas y quizá porque la confianza mutua se ha resentido en ámbitos donde los niveles de competencia entre profesionales han crecido mucho. Además en la actualidad existen métodos alternativos para transmitir resultados sin necesidad de realizar publicaciones.

#### **15.2.3.- El diagnóstico rápido participativo.**

El diagnóstico rápido participativo es una técnica un tanto desconocida, pero que vamos a utilizar como ejemplo de estrategia peculiar e imaginativa. Aunque básicamente puede utilizarse como sustitutivo del análisis de la realidad más clásico y costoso, en circunstancias muy especiales puede también utilizarse como una evaluación global del proyecto, programa e incluso de un Plan de Juventud.

En todo caso, es un recurso extremo al que se acude cuando no se ha previsto o no hay recursos para realizar una evaluación, lo cual, cuando estamos hablando de una política pública resulta un mal síntoma. También es cierto, aunque esta es una experiencia muy personal, que rodeado en la mesa de programas y planes, de memorias, de resultados de encuestas, de transcripciones

de grupos de discusión y de entrevistas, un “diagnóstico rápido participativo” sobre el terreno (de un día) me ha cambiado la perspectiva o me ha permitido ordenar un material que hasta el momento parecía muy confuso.

El diagnóstico rápido supone “movilizar” para un encuentro único y muy limitado a un conjunto de agentes sociales (o sólo agentes que intervinieron) para repasar lo que se está haciendo. Es una metodología que se ha utilizado en Programas de Cooperación ante situaciones que imposibilitan la puesta en marcha de otras metodologías. Por ejemplo, el Banco Mundial ha comenzado a exigir estas estrategias en situaciones socialmente complejas, que incluyen algún grado de represión o violencia que impide conocer la opinión de la población. No se fía de los informes y estadísticas oficiales, pero tampoco tiene manera de contrastarlas con otras metodologías. En realidad es como una Observación Participante, corta pero densa, que se realiza siguiendo una estrategia peripatética de tipo socrático (es decir, paseando en grupo mientras alrededor se van desarrollando las diferentes actividades del programa), hasta llegar al “debate final” entre todos aquellos que han participado en el “paseo”.

Es una estrategia que recibe el nombre de “transectos” cuando la realizan campesinos (en general indígenas) con técnicos agrícolas recorriendo los campos para valorar el impacto de las reformas agrarias y la adopción de nuevas tecnologías (Rodríguez-Villasante, 2007). La visión física inmediata de los acontecimientos, la interacción entre los diversos actores sobre el territorio de la realidad y la intencionalidad manifiesta de la tarea, dotan a estos “paseos” de una densidad interpretativa, que permite establecer conclusiones de una forma rápida y más certera que si la discusión se realiza fuera de contexto.

### 15.3.- Las estimaciones basadas en la evolución de los indicadores formales.

Una forma curiosa de completar la evaluación reside en realizar estimaciones a partir de indicadores formales, es decir, de estadísticas y datos secundarios como los que hemos expuesto en el capítulo 6. No es una técnica fácil.

Se trata de tomar estos datos y realizar estimaciones a partir de extrapolaciones y ponderaciones del peso de determinadas variables que se supone reflejan el impacto particular de la intervención en el contexto de otros impactos que incluso pueden ser más relevantes. Los profesionales más acostumbrados a utilizar estas técnicas son los economistas. Son muy delicadas y las conclusiones nunca pueden considerarse evidencias, salvo que las triangulemos.

#### UN EJEMPLO DE ESTIMACIÓN: EL ANÁLISIS DEL IMPACTO DE UNA POLÍTICA DE EMPLEO.

*Acaba de aparecer, en francés, una publicación en torno a un aspecto clave de las políticas de juventud, posee un notable interés e incluso diría que supone una “evaluación de carácter estratégico”. Se titula “**Las dotaciones de capital para los jóvenes**”, ha sido editada en papel pero también resulta accesible en la dirección. ([http://www.strategie.gouv.fr/IMG/pdf/Rapport\\_no9\\_Web.pdf](http://www.strategie.gouv.fr/IMG/pdf/Rapport_no9_Web.pdf)). Se trata de una evaluación de las acciones relativas a la “asignación de capitales”, para que los jóvenes puedan diseñar un proyecto de vida. Son muy numerosos los países que poseen experiencias en esta materia (Singapur, Inglaterra, Hungría, Corea, Canadá, Estados Unidos, Noruega,...). En cada país la acción es distinta, desde la creación, al nacimiento de cada niño o niña, de “cuentas bancarias” con dinero público y que va creciendo (con aportaciones públicas y privadas por ventajas fiscales) hasta que se puede retirar a partir de los 18 años (algo que en España se hacía mucho en el pasado y que aún mantienen*

algunas Cajas de Ahorros), hasta el modelo de “préstamo único sin interés” de otros países.

*El informe francés sintetiza evalúa tres opciones diferentes: los modelos de dotación universal para todos los jóvenes, los modelos de dotación discriminatoria aplicados a jóvenes con familias de bajas rentas y los modelos de ayudas y préstamos para acciones de formación o proyectos laborales más o menos innovadores.*

*El objetivo es evaluar el impacto de cada uno de ellos y estimar las ventajas que supondría aplicar uno u otro en Francia. La respuesta es clara aunque expresada en un lenguaje ambiguo), el mayor impacto no se corresponde con uno u otro modelo, sino con el grado de ajuste de cada una de las opciones a rasgos culturales (y maneras tradicionales de enfocar el tema) en cada una de los países mencionados. Por tanto, estiman los autores, que la opción de más éxito en Francia sería la tercera (ayudas y préstamos para acciones concretas), ya que en el país vecino existe una larga tradición, conocida y aceptada por la población, de ayudas (desde alquiler de viviendas hasta bolsas de estudios pasando por ayudas a las creación de nuevas empresas) para la emancipación de los jóvenes. Por lo tanto en Francia se estima que la política más eficaz sería la de unificar (y ampliar los recursos) de todas estas ayudas bajo una rúbrica común.*

#### **15.4.- La noción de triangulación.**

La triangulación es una estrategia de investigación social que suele utilizarse en aquellos casos en los cuales los resultados o las evidencias parecen complejos o son muy controvertidos. Se trata de una fórmula que se supone que resulta equivalente a la seguridad que se dice que ofrece la investigación experimental. La idea es muy simple: en vez de utilizar un grupo de control, se repite la investigación utilizando otra metodología y otros entrevistados. Aunque en algunos casos, como una estrategia explícita, podrían ser los mismos, pero con cambios en la metodología.

Al final, las conclusiones y los resultados obtenidos por ambas investigaciones (o una tercera o más) se comparan y si coinciden se supone que los resultados o las evidencias son las correctas. En caso de diferencias hay que tratar de conocer la causa, relacionándolas con el tipo de metodología utilizado y proponiendo explicaciones para las mismas.

La práctica de la triangulación no es muy frecuente, pero de una forma no explícita y aludiendo a la complementariedad de métodos y al logro de una “mayor riqueza de información” en los últimos años se realizan frecuentes “triangulaciones implícitas”, combinando encuestas y grupos de discusión.

Pero la mayor parte de estas investigaciones de “doble método” sin más no responden a la idea de la triangulación, que procede del ámbito de la navegación (antes de que existiera el GPS), cuando se tomaban varios puntos de referencia astronómicos y horarios para establecer la posición del buque. “Triangular” no es por tanto combinar métodos sin más, sino hacerlo buscando un fin preciso.

De hecho, fue un autor procedente del campo de la evaluación (y que ha hecho numerosas aportaciones al mismo), Norman K. Denzin, quien estableció originalmente el concepto como “*la combinación de dos o más teorías, fuentes de datos o métodos de investigación, en el estudio de un fenómeno singular*” (Denzin, 1978). Más adelante añadió que era “*un abordaje más completo de este fenómeno*”, que permitía “*identificar, superar y comprender las limitaciones de uno de los métodos*”, pero que un simple “*método mixto no es necesariamente un método de triangulación, pues la triada debe ser complementaria y servir para el tipo de resultado preciso que se busca*”. La idea es que no por combinar métodos ya se logra triangular sino que es necesario que

el procedimiento de la triangulación responda a un objetivo preciso: en nuestro caso evaluar un programa o un plan de juventud.

Por ejemplo, si, de manera simple, trato de determinar el impacto de una campaña de reducción de riesgos en las relaciones sexuales de los adolescentes promocionado los métodos anticonceptivos adecuados, puedo tratar de triangular partiendo de una investigación mediante una encuesta previa y otra posterior para conocer el cambio de actitud de los adolescentes de la localidad hacia los distintos métodos (y el grado de información que poseen y el uso que hacen de los mismos), con otra investigación sobre la percepción del cambio que han tenido los profesionales que atienden a estos mismos adolescentes.

Pero si los profesionales carecen de protocolos en los que registran las actividades sexuales del motivo de la consulta o la petición de información, para realizar las estimaciones a partir de los mismos, puede ocurrir que la “percepción del cambio” que manifiesten tales profesionales se halle sujeta al rechazo que sienten porque la campaña ha sido realizada por juventud, con las organizaciones juveniles y sin contar con su asesoramiento. En el segundo caso, las ventajas de la triangulación desaparecen y se añade confusión a la evaluación.

Quizá por este motivo Denzin afirmaba que, aparte de darse una triangulación entre metodologías, también podía darse entre datos obtenidos por la misma metodología (por ejemplo mediante preguntas formuladas de manera diferente), entre investigadores (dos o más investigadores trabajando de forma independiente sobre un mismo proyecto) y entre teorías (formulando hipótesis alternativas y tratando de contrastarlas). Se supone que todas estas formas de triangulación son posibles si se han respondido de forma completa las preguntas que, como hemos explicado, formula Kant como condiciones previas a cualquier evaluación.

Para los muy interesados cabe decir también que Denzin define para la triangulación dos categorías antagónicas. De una parte la *within-method triangulation*, que se refiere a la utilización de varias técnicas dentro de una misma aproximación metodológica, lo que le permite garantizar una mejor consistencia interna y fiabilidad; por otra la *between-(or across-) method triangulation*, referida a la utilización de varios métodos (o teorías o investigadores) distintos, la cual persigue la búsqueda de validez externa (Denzin, 1978). Ambas categorías son bien útiles a la evaluación.

## BIBLIOGRAFÍA Y DOCUMENTACIÓN COMPLEMENTARIA.

- David, G y SCOTT, J., (1980), **Estrategias para la creatividad**, Buenos Aires, Paidós.
- ROGERS, C. (1978), **Libertad y creatividad en la Educación**, Buenos Aires, Paidós.



**PARTE IV**  
**APORTACIONES PRÁCTICAS**

*“La primera virtud (para el educador comprometido) que me gustaría subrayar, que no es fácil de ser creada, es la virtud de la coherencia entre el discurso que se habla y que anuncia la opción y la práctica que debería estar confirmando el discurso. Esta virtud enfatiza la necesidad de disminuir la distancia entre el discurso y la práctica”*

Paulo FREIRE

De una conferencia pronunciada en la Asamblea Mundial de Educación de Adultos en Buenos Aires 1985, reproducido en EDUCACIÓN Y CULTURA, n° 11 (1987), Bogota, SNE.



## Un modelo de indicadores para evaluar políticas, planes, programas y actividades de juventud

Como hemos visto, la evaluación requiere seleccionar indicadores adecuados para responder a las preguntas que se formulan en la matriz de evaluación de las intervenciones y sus objetivos. La selección de tales indicadores es un aspecto clave de toda evaluación. Esto significa que los indicadores no deben prefijarse y que no existe un modelo de “indicadores” con validez general para toda intervención, aunque nos refiramos sólo a un ámbito particular como puede ser juventud.

Sin embargo, ni la disponibilidad de indicadores es ilimitada, ni la construcción de nuevos e inéditos indicadores una tarea sencilla. En general se recurre a indicadores ya utilizados en otras evaluaciones, lo que permite además realizar comparaciones y acumular resultados. En este capítulo vamos a presentar una reflexión general y una descripción de algunos indicadores con la finalidad de orientar los trabajos de evaluación. En cualquier caso, debe quedar claro que no se trata ni de una relación exhaustiva, ni tan siquiera de una “recomendación general” sobre los indicadores a utilizar, sino más bien una simple “orientación”.

Por otra parte Martín Hammersley (un evaluador educativo experto en cuestiones de etnicidad) estableció en 1977, en un trabajo sobre tipologías de indicadores, que cada indicador debía referirse a una **dimensión** (sinónimo de “aspecto o faceta de algo” según el DRAE o “cualidad” según el María Moliner) del programa, la intervención o la política. Mostrando que no podemos hablar de indicadores si antes no hemos establecido claramente cuales son las dimensiones de las intervenciones. De hecho fue Hammersley quien primero estableció un modelo de dimensiones del sistema educativo, estableciendo cinco grandes áreas: 1) definición del papel del maestro, 2) conceptualización de la acción del alumno, 3) conceptualización del conocimiento, 4) conceptualización del aprendizaje, 5) técnicas preferidas. Cada área incluye diversas dimensiones y cada una de ellas permite formular uno o varios indicadores (Woods, 1987). Para hablar de indicadores, por tanto, debemos comenzar por hablar de dimensiones.

### 16.1.- El debate sobre los indicadores.

La cuestión de los indicadores ha generado muchos debates en el área de las ciencias sociales. Se trata de debates que durante un tiempo se realizaron en paralelo al debate entre cuali-

tativo y cuantitativo, ya que se consideraba que el uso de indicadores respondía a estrategias cuantitativas, mientras que las cualitativas no aceptaban la existencia de tales indicadores. En realidad esto ocurría porque los principales defensores de los indicadores han sido siempre los economistas y en el ámbito económico, hasta fechas recientes, los únicos indicadores válidos eran los cuantitativos. En todo caso hay que reconocer que en la práctica el análisis económico no puede realizarse sin indicadores cuantitativos y que la mayoría de las teorías económicas no se habrían formulado sin la ayuda de los indicadores cuantitativos.

Pero la progresiva implantación de la noción de complementariedad de los métodos y las técnicas ha modificado el panorama y ha permitido formular incluso la idea de “indicadores cualitativos”. De hecho, esta noción refleja que no debemos confundir los métodos y las técnicas de recogida y producción de datos con el uso que vayamos a darles en el análisis. Lo que a su vez equivale a afirmar que cualquier dato producido por cualquier método o técnica puede “convertirse en” o “utilizarse como” un indicador (o como un dato por sí mismo sin que lo llamemos indicador). El único requisito para que los datos obtenidos mediante cualquier método o una técnica puedan ser considerados “indicadores” es que se hayan recogido de acuerdo con criterios adecuados a la propia tradición metodológica.

En todo caso, se supone que los indicadores deben poseer ciertas características para que sean útiles. En el caso de la evaluación, la adaptación de estas características nos conduciría a estos seis principios:

1. Haber sido fijados de forma previa en la matriz de datos y ligados a los objetivos que aparecen en la misma.
2. No tener ningún grado de ambigüedad, es decir, que no den la posibilidad de diversas interpretaciones.
3. Poder efectuar mediciones eficaces y poder proporcionar datos de manera fiable y sencilla.
4. Poder ser utilizables en distintas evaluaciones (y por tanto poder compararlas), así como en la evaluación de una misma política a lo largo del tiempo.
5. Poder utilizarlos en diferentes ámbitos y políticas, para comparar procesos y resultados, es decir ser transversales.
6. Poder utilizarlos sin costes adicionales.

Una mera observación de estas seis características demuestra que no son fácilmente alcanzables, lo que en ocasiones convierte la “búsqueda de indicadores” en una tarea sumamente difícil, en una especie de utopía paralizante que impide avanzar en la evaluación. Además, son muchos los autores que opinan que con los indicadores se corre varios riesgos, del tipo como *“burocratizar la evaluación”, “olvidarse de las características particulares del contexto”* y especialmente *“convertir a los propios indicadores en un bucle teórico que obvia otros conceptos o teorías y que se olvida que el método científico es una contrastación abierta de hipótesis”*. A la vez, casi todo el mundo reconoce que utilizar de forma estable los mismos indicadores posee ventajas como establecer series históricas que permitan acumular resultados, medir la evaluación de las mismas y realizar análisis comparativos. Cuestiones que han sido tratadas en los capítulos 6 y 8.

Conviene recordar que de hecho, en España, se produjo un debate histórico (de los más famosos en ciencias sociales) sobre el posible uso de los indicadores sociales en los años 60 (De Miguel, 1967; Díez Nicolás, 1968 y Del Campo, 1972), hasta que se llegó a una especie de “callejón sin salida” por las mutuas críticas otorgadas. Pero en la actualidad el avance de las nuevas tecnologías en relación al registro de datos y una cierta deconstrucción de los indicadores sociales, especialmente en el territorio de las políticas de género (Forum de Política Feminista, 2007), vuelve a colocar el tema en primer plano. Asimismo, para poder analizar la cuestión del cambio climático se ha recurrido a sistemas de indicadores, lo que ha colocado la comprensión del uso de los mismos en el nivel de la opinión pública.

De especial relevancia resulta también el cambio de actitud de las administraciones públicas en esta materia y que hemos expuesto en el capítulo 8 y que de nuevo se relaciona con el avance de las TIC. Desde principios de los años 90 “la revolución de los registros”, permitió formular sistemas de registro de las actuaciones públicas inimaginables una década antes (Torres y Pina, 2004).

Todo este replanteamiento ha permitido en la actualidad que se estén realizando debates y se planteen sistemas de indicadores, aparte del citado ámbito de igualdad y género, así como de aspectos relacionados con el clima, en temas tan variados como salud, educación, sostenibilidad (y en general medio ambiente), ciudadanía, desarrollo urbano, investigación, cooperación (y desarrollo), muchos de ellos temas transversales en relación a las políticas de juventud. Pero hasta ahora, aunque distintos observatorios de la juventud han construido, de facto, sistemas de indicadores sobre la juventud, no se ha establecido ninguna relación medible entre tales indicadores y las políticas de juventud. Ni por supuesto se ha debatido, como ha ocurrido en otros ámbitos, esta cuestión.

## **16.2.- Algunos modelos de indicadores.**

Para comenzar este debate en juventud podemos proponer una doble estrategia, de una parte utilizar los sistemas de indicadores propuestos en otros ámbitos y por otra parte estimar en que punto se encuentra la descripción de las dimensiones de las políticas de juventud y la propuesta de indicadores en relación a estas políticas.

Para desarrollar la primera estrategia podríamos utilizar algunas referencias reconocidas y cercanas, por ejemplo el Department of Development Services del Estado de California, ha implantado un sistema de indicadores de Calidad de Vida Personal, que ha sido adaptado a España (Fernandez, Cabo y García, 1999) y que el Imsero ha “acogido”, quizá porque se adecua muy bien a las necesidades de su población objetivo. Veamos cuales son y así vamos entrando en el tema.

### **Dimensiones e indicadores de Calidad de Vida Personal.**

#### **Opciones**

1. Las personas identifican sus necesidades, deseos, y lo que les agrada y desagrada.
2. Las personas toman decisiones importantes sobre sus vidas.
3. Las personas toman decisiones sobre cuestiones cotidianas.
4. Las personas desempeñan un papel importante en la elección de los proveedores de sus servicios y apoyos.
5. Los servicios y los apoyos requeridos por una persona cambian a medida que cambian sus deseos, necesidades y preferencias.

#### **Relaciones**

6. Las personas tienen amigos y amistades relevantes.
7. Las personas cuentan con apoyos vecinales que pueden incluir a los familiares, amigos, proveedores de servicios, profesionales y a otros miembros del vecindario.

### **Estilo de Vida**

8. Las personas son parte de la vida vecinal normal y viven, trabajan y se divierten en ambientes integradores.
9. Los estilos de vida de las personas reflejan sus preferencias culturales.
10. Las personas son independientes y productivas.
11. Las personas tienen situaciones de vida estables.
12. Las personas están cómodas en el lugar donde viven.
13. Los niños viven en hogares con familias.

### **Salud y Bienestar**

14. Las personas están seguras.
15. Las personas disfrutan de la mejor salud posible.
16. Las personas saben lo que deben hacer en caso de riesgo para su salud, seguridad y bienestar.
17. Las personas tienen acceso a la asistencia sanitaria que necesitan.

### **Derechos**

18. Las personas ejercen sus derechos y responsabilidades.
19. Las personas no sufren malos tratos, abandono ni explotación.
20. Las personas son tratadas con dignidad y respeto.
21. Las personas utilizan servicios públicos y apoyos adecuados.
22. Las personas tienen asistencia legal y/o acceso a servicios de asesoría legal.

### **Satisfacción**

23. Las personas alcanzan los objetivos que se marcan.
24. Las personas están satisfechas con los servicios y los apoyos que reciben.
25. Las personas están satisfechas con sus vidas.

Otro ejemplo, muy completo y que refleja muy bien la actual funcionalidad de los indicadores, la ofrece la reciente propuesta (Schalock y Verdugo, 2002/2003) de un modelo de indicadores de calidad de vida dirigido a impulsar la aplicación del concepto en diferentes tipos de programas y servicios. Estamos ante una propuesta de un modelo global, aunque su génesis empírica se sitúa en un sector muy particular de la intervención social: el ámbito de las personas con minusvalías.

Se trata de un modelo que contempla un enfoque multidimensional compuesto por ocho dimensiones diferentes, que a su vez se hacen operativas en la formulación de indicadores. Las dimensiones propuestas se han obtenido de investigaciones relacionadas con la minusvalía realizadas por los mismos autores, completadas por los 24 indicadores más utilizados por la investigación, que se exponen a continuación. La propuesta está teniendo un gran éxito, porque aúna cuestiones relacionadas con la evaluación con un planteamiento político de bienestar colectivo que produce un fuerte impacto ético, moral y cultural, que legitima, más allá de su aplicación técnica, dicho modelo.

**DIMENSIONES E INDICADORES DE CALIDAD DE VIDA.**  
**Schallock y M.A. Verdugo (2002/2003)**

**Área de Bienestar Emocional**

1. **Satisfacción:** Estar satisfecho, feliz y contento.
2. **Autoconcepto:** Estar a gusto con su cuerpo, con su forma de ser, sentirse valioso.
3. **Ausencia de estrés:** Disponer de un ambiente seguro, estable y predecible; no sentirse nervioso, saber lo que tiene que hacer y que puede hacerlo.

**Relaciones Interpersonales**

4. **Interacciones:** Estar con diferentes personas, disponer de redes sociales.
5. **Relaciones:** Tener relaciones satisfactorias, tener amigos y familiares y llevarse bien con ellos.
6. **Apoyos:** Sentirse apoyado a nivel físico, emocional, económico. Disponer de personas que le ayuden cuando lo necesite y que le den información sobre sus conductas.

**Bienestar Material**

7. **Estatus económico:** Disponer de ingresos suficientes para comprar lo que necesita o le gusta.
8. **Empleo:** Tener un trabajo digno que le guste y un ambiente laboral adecuado.
9. **Vivienda:** Disponer de una vivienda confortable, donde se sienta a gusto y cómodo.

**Desarrollo Personal**

10. **Educación:** Tener posibilidades de recibir una educación adecuada, de acceder a títulos educativos, de que se le enseñen cosas interesantes y útiles.
11. **Competencia personal:** Disponer de conocimientos y habilidades sobre distintas cosas que le permitan manejarse de forma autónoma en su vida diaria, su trabajo y su ocio, sus relaciones sociales.
12. **Desempeño:** Tener éxito en las diferentes actividades que realiza, ser productivo y creativo.

**Bienestar Físico**

13. **Salud:** Tener un buen estado de salud, estar bien alimentado, no tener síntomas de enfermedad.
14. **Actividades de la vida diaria:** Estar bien físicamente para poder moverse de forma independiente y realizar por sí mismo actividades de autocuidado, como la alimentación, el aseo, el vestido, etc.
15. **Atención sanitaria:** Disponer de servicios de atención sanitaria eficaces y satisfactorios.
16. **Ocio:** Estar bien físicamente para poder realizar distintas actividades de ocio y pasatiempos.

**Auto Determinación**

17. **Autonomía/Control Personal:** Tener la posibilidad de decidir sobre su propia vida de forma independiente y responsable.
18. **Metas y valores personales:** Disponer de valores personales, expectativas, deseos hacia los que dirija sus acciones.
19. **Elecciones:** Disponer de distintas opciones entre las cuales elegir de forma independiente según sus preferencias, por ejemplo, dónde vivir, en qué trabajar, qué ropa ponerse, qué hacer en su tiempo libre, quiénes son sus amigos.

## **Inclusión Social**

- 20. Integración y participación en la comunidad:** Acceder a todos los lugares y grupos comunitarios y participar del mismo modo que el resto de personas sin discapacidad.
- 21. Roles comunitarios:** Ser una persona útil y valorada en los distintos lugares y grupos comunitarios en los que participa, tener un estilo de vida similar al de personas sin discapacidad de su edad.
- 22. Apoyos sociales:** Disponer de redes de apoyo y de ayuda necesaria de grupos y servicios cuando lo necesite.

## **Derechos**

- 23. Derechos humanos:** Que se conozcan y respeten sus derechos como ser humano y no se le discrimine por su discapacidad.
- 24. Derechos legales:** Disponer de los mismos derechos que el resto de los ciudadanos y tener acceso a procesos legales para asegurar el respeto de estos derechos.

Como ya hemos dicho el modelo propuesto surge de un colectivo social muy concreto (minusvalías) y por ello hace especial referencia a aspectos relacionados con la salud y la dependencia. Sin embargo, al mismo tiempo estructura las dimensiones e indicadores citados sobre la totalidad del sistema social, lo que le permite contemplar la totalidad de aspectos que deberían confluir en un plan relacionado con esta población, pero también permite considerar todos los indicadores que entrarían a formar parte de su evaluación. Además, relaciona los aspectos personales del individuo (micro), con los funcionales del ambiente que le rodea (meso), y con los indicadores sociales en sentido clásico (macro).

A la vez reitera la necesidad del pluralismo metodológico que hemos defendido en la parte II, ya que requiere una estrategia de evaluación que combine todas las técnicas de recogida y producción de datos citadas para poder medir así la perspectiva personal (valor o satisfacción), evaluación funcional (comportamiento adaptativo y estatus o rol ocupado) e indicadores sociales (medidas a nivel social).

¿Podríamos usar o adaptar estos modelos al ámbito de juventud? De hecho, no podemos obviar los esfuerzos realizados en áreas tan cercanas que nos conducen a **imaginar un modelo de dimensiones e indicadores de Calidad de Vida para los Jóvenes. Pero si descartar esta opción hemos preferido para este Manual proponer un modelo de dimensiones e indicadores para las Políticas de Juventud efectivamente realizadas.** En una gran medida esta decisión tiene que ver con una cierta realidad de los ámbitos de juventud, que en conexión con la noción de evaluación fuerte nos invita a adoptar algunas cautelas.

Se trata del problema de la retórica de las políticas de juventud, en relación a las políticas efectivamente realizadas y que se ha descrito en una reciente publicación (Comas, 2007). Crear un modelo de dimensiones e indicadores sobre Calidad de Vida de los Jóvenes contribuirá al “engorde” de estas retóricas y esto es lo que menos necesitan los ámbitos de juventud para consolidarse, lo que necesitan es **evaluar de verdad** lo que están haciendo. Si además, una vez puesto en marcha un proceso de evaluación realista, se quiere diseñar un modelo de Calidad de Vida para los Jóvenes que recoja un conjunto de “reivindicaciones”, está muy bien. Pero hacer esto sin considerar mejorar lo que efectivamente se está haciendo es caer en la trampa de “las grandes palabras” como sustitutivo de la acción social real.

## 16.3.- Propuesta de indicadores para la evaluación en los ámbitos de juventud.

### 16.3.1. El problema del déficit de evaluación y la cuestión de las “Memorias”.

¿Cómo podemos crear un modelo de indicadores para las políticas reales de juventud? Pues con tiempo y mucha reflexión, de hecho la propuesta (más conceptual además) de Robert Schalock y Miguel Ángel Verdugo ha tardado más de dos decenios en configurarse a partir del planteamiento de partida, aunque se supone que ahora no vamos a necesitar tanto tiempo, ya que otros han realizado interesantes avances. Pero, por otra parte, las políticas y programas en el ámbito de las minusvalías están delimitados por esta condición física, mientras que el ámbito de juventud está sólo acotado por edades y dentro del rango 15-30 años casi todo puede ser una política de juventud. Además las políticas de juventud aparecen cargadas de debates, ambigüedades y confrontaciones de valores.

Tal situación quizá explique la falta de evaluaciones (al menos en sentido estricto) en el ámbito de juventud<sup>33</sup>. Porque el exceso de controversias, como ocurre en el territorio particular como juventud, no ayuda a promocionar la evaluación y mucho menos poner en marcha un modelo de dimensiones e indicadores de estas políticas. Aunque se puede alegar que desde la perspectiva de algunas definiciones de política de juventud, por ejemplo aquellas que las circunscriben a la emancipación, evaluar ciertos objetivos, por ejemplo el acceso a un empleo más o menos estable, se puede hacer con facilidad porque estas políticas ya se evalúan en su conjunto y para las mismas existen indicadores claros y bien recogidos por el INE.

Pero asumir esta definición y esta estrategia tiene dos problemas. El primero que solo evaluamos una parte de las políticas de juventud reales que incluyen otros muchos programas e intervenciones. El segundo que en ciertos temas como el empleo influyen otras políticas muy potentes y que además la evaluación de estas políticas ya se realiza. Obviamente siempre se podrá alegar que esto está bien así porque no se trata de políticas de juventud sino de empleo, por lo que volvemos a la casilla de partida: sean las que sean ¿cómo evaluamos entonces las políticas específicas de juventud? o expresado en términos más administrativos, ¿cómo evaluamos los aproximadamente 650 millones de € de partidas presupuestarias que las tres administraciones dedican al capítulo de organismos y programas específicos de juventud? Pues en esto estamos y en la definición de políticas de juventud que agrupe estas partidas presupuestarias singulares.

Una posibilidad sería evaluar estas políticas, tanto las citadas de empleo como por ejemplo las “jóvenes creadores”, desde una perspectiva transversal que incluya las iniciativas de diferentes ámbitos de las administraciones, por ejemplo para las primeras trabajo y para las segundas cultura, pero entonces seguiríamos sin evaluar las políticas de juventud en particular, y lo incluiríamos en un gran capítulo de políticas sociales (algunas de ellas muy potentes), lo que nos obligaría a proponer una síntesis de indicadores sociales muy globales, algo que, como hemos venido explicando, no se ajusta para nada a la función de *“un canon sistemático de prácticas sobre políticas, programas y otras actividades sociales e institucionales, que nos permite recoger información, valorarla y devolverla a los interesados y afectados, tratando de mejorar estas intervenciones”*. También hemos explicado que los “grandes indicadores formales” que miden la evaluación global de las políticas sociales valen para confirmar la evolución (en términos macro) de estas políticas, pero que no nos dicen nada de las políticas más concretas.

<sup>33</sup> La afirmación no es exacta del todo. Dispongo de algunas evaluaciones de Planes de Juventud en el sentido más estricto del término, en concreto de tres Comunidades Autónomas y de tres Municipios. El problema es que algunas están publicadas formalmente mientras que otras son fotocopias muy precarias. No voy a citar ninguna porque imagino que pueden existir otras en la misma situación y creo que es preferible esperar a la creación de algún tipo de registro. En todo caso esta precariedad y la dificultad en acceder a los resultados de las evaluaciones, que contrasta con la buena difusión de las Memorias, demuestra el “miedo” que aún subsiste en los ámbitos de juventud a las evaluaciones. En todo caso utilizaré esta documentación en el diseño de la propuesta de dimensiones e indicadores.

A la inversa también tenemos que afrontar el problema de la evaluación corporativa más micro. Por ejemplo en el caso de la animación sociocultural, muy presente en los ámbitos de juventud, aparecen numerosas propuestas de evaluación para los profesionales que desarrollan estas actividades (Cenbranos, Montesinos y Bustelo, 1994; Froufe, 1994; Merino, 2000). Se trata de propuestas muy bien desarrolladas, pero ¿qué evalúan?, ¿el trabajo realizado por estos profesionales o una política o un programa? Pues claramente lo primero, que **puede servir** en la evaluación de lo segundo, pero **no es** una evaluación de planes, políticas o programas, que se sitúa en un nivel diferente.

Aclaradas todas estas cuestiones cabe añadir que en los organismos de juventud, en los cuales aparecen una multiplicidad de programas, con o sin concurrencia de otros ámbitos administrativos, se ha venido produciendo un tipo de producto, del que existe una larga tradición y numerosos ejemplos de continuidad histórica: **la Memoria anual del organismo y sus actividades**.

La realización de memorias anuales por parte de programas y organismo de juventud comenzó muy pronto en aquello que hemos llamado (Comas, 2007) “oleada eufórica de políticas de juventud”, en particular municipales, que se desarrolló durante la transición al sistema democrático y en los años siguientes. Los protagonistas de aquella “oleada” tenían claro que, frente al régimen de la dictadura que “controlaba los rendimientos” con una mera estrategia de control monetario y jurídico (Ruano, 2004), debían manifestar un buen grado de transparencia y “pasar cuentas” a la sociedad (y a los jóvenes) mediante las susodichas memorias. Se suponía que, en un momento en el cual la evaluación de políticas públicas no era en nuestro país siquiera una idea, las memorias ofrecían información, valoraciones y transparencia, lo que les lleva a cumplir unos objetivos similares a lo que en los años sucesivos se ha convertido en una de las misiones clave de la evaluación<sup>34</sup>.

La existencia de tal conjunto de memorias anuales, de las cuales los técnicos de juventud que las han elaborado con paciencia y dedicación pueden sentirse orgullosos, nos permiten visualizar cómo la idea de las “políticas de juventud” adoptó, en el momento de la transición a la democracia, un carácter innovador, con prácticas que otros ámbitos administrativos tardaron en asumir (o aún no han asumido). Pero aquella innovación ya no se ajusta a los avances y a las exigencias de la gobernanza en una democracia que propone un futuro con mayores dosis de ciudadanía. Ya no vale hacer sólo memorias, hay que evaluar las políticas y hacerlo desde la perspectiva y los criterios de la evaluación fuerte.

### 16.3.2.- Indicadores seleccionados.

A pesar de las dificultades mencionadas pero también con el apoyo que nos ofrece la experiencia histórica de las memorias anuales, podemos proponer un primer esquema de indicadores, bien entendido que se trata de un primer paso, un paso previo incluso, para iniciar un debate sobre este tema. Un debate que responde a una propuesta de Naciones Unidas contenido en el Programa de Acción Mundial para los Jóvenes y a la que se invita a los propios jóvenes a participar a partir de una guía muy simple pero muy oportuna y bien diseñada (Naciones Unidas, 2004).

El punto de partida requiere contestar a una pregunta: ¿qué tipo de proyectos y políticas se realizan en el ámbito de juventud? Y la respuesta la hemos obtenido en una reciente investigación (Comas, 2007), aunque aparecen dudas a la hora de contestar otras preguntas necesarias, como por ejemplo: ¿cuál es la relación entre las políticas públicas expresadas y los proyectos realizados? ¿Existen criterios de éxito para estas políticas? ¿Quiénes son los agentes que toman

<sup>34</sup> Un ejemplo muy preciso lo constituyen las políticas de igualdad de género, las cuales se comenzaron a impulsar más tarde y además fue un proceso jerárquico muy distinto, porque primero fue la administración central y después la autonómica, mientras que las de juventud comenzaron en los ayuntamientos. Entre las políticas de juventud de los primeros ayuntamientos democráticos y las políticas estatales y autonómicas de género aparece una distancia de 10-12 años, tiempo suficiente para que en España el tema de la evaluación hubiera adquirido un cierto protagonismo, lo que ha propiciado la evaluación de las políticas de género, mientras que en dicho ámbito las Memorias son un documento casi invisible (Bustelo, 2004). También es cierto que en las políticas de juventud no se ha contado, hasta fechas muy recientes, con fondos europeos, mientras que las de género sí los tuvieron lo cual incluía las “correspondientes” obligaciones de evaluar.

decisiones en relación a las mismas? ¿Cómo se adecuan los proyectos a las necesidades o problemas expresados en la abundante documentación sobre la realidad de los jóvenes españoles?

Con respuestas parciales a estas preguntas podemos proponer un esquema de dimensiones e indicadores para las políticas de juventud, asumiendo que dicho esquema no puede (ni debe) incluir la totalidad de las dimensiones de la vida de los jóvenes, porque ni todas ellas forman parte de las políticas de juventud, ni tratamos de establecer un sistema global de indicadores sociales sobre juventud, porque esto sería muy ajeno a las necesidades de un modelo de evaluación fuerte como el expuesto en este Manual. Por ello, nos limitaremos a tres dimensiones, a las tres áreas que componen actualmente las políticas de juventud. Es decir, trataremos como dimensiones para la evaluación de las políticas de juventud la participación, la afirmación y la transición, añadiendo también la investigación.

Cada una de ellas aparece desagregada en una serie de objetivos que reflejarían la síntesis de dichas políticas y dentro de cada uno de estos grupos de objetivos aparecen, hasta sumar 56, las actividades propias de estas políticas, formuladas también como objetivos y cada una de los cuales incluiría uno o varios indicadores.

En la relación aparecen en cursiva, a modo de ejemplo, algunos de estos indicadores finales. No he querido ser exhaustivo, sino sólo sugerente, ya que creo que, como he explicado, para tener una buena base de “posibles indicadores” se requiere muchas evaluaciones previas que los hayan puesto a prueba y esto, en juventud y como he explicado, es una tarea pendiente.

## **PROPUESTA DE DIMENSIONES E INDICADORES PARA POLÍTICAS DE JUVENTUD**

### **1.- ÁREA DE PARTICIPACIÓN: Políticas, planes, programas y actividades relacionadas con el desarrollo del artículo 48 de la Constitución.**

Indicadores de Calidad de Vida relacionados con participación en la vida social y política, roles comunitarios, apoyos sociales y relaciones (en parte también derechos).

#### **1.1.- Objetivos relacionados con la promoción del movimiento asociativo juvenil.**

- 1.- Recursos e infraestructuras suficientes para el desarrollo de la labor asociativa. *Atención a la demanda, incremento de la promoción y eficiencia de los recursos empleados.*
- 2.- Nivel de apoyo a programas para facilitar el compromiso cívico de las asociaciones juveniles. *Eficacia diferencial de los servicios prestados por jóvenes a otros jóvenes.*
- 3.- Fomento del asociacionismo en el ámbito educativo. *Directivas elaboradas, número de centros participantes, razones del rechazo o participación, campañas y acciones de sensibilización, recursos aportados, solicitudes y distribución.*

#### **1.2.- Objetivos relacionados con la promoción de la participación cívica y política de los jóvenes.**

- 4.- Foros y otros procedimientos de participación abierta. *Mesas, reuniones, participaciones y aportaciones realizadas.*
- 5.- Promoción de la afiliación a organizaciones sociales, cívicas y juveniles.
- 6.- Facilidades para la creación y legalización de asociaciones.
- 7.- Subvenciones y ayudas.

#### **1.3.- Objetivos relacionados con la formación de los agentes sociales juveniles.**

- 8.- Detección y captación de agentes y líderes que conforman el tejido asociativo (forma e informal) juvenil.
- 9.- Grado de impulso a la formación de estos agentes.
- 10.- Promoción del voluntariado juvenil.

#### **1.4.- Objetivos relacionados con la convivencia y el desarrollo de una ciudadanía activa.**

- 11.- Conocimiento de derechos y deberes civiles, sociales y culturales.
- 12.- Fomento de la tolerancia y el respeto a la diversidad.
- 13.- Conocimiento de los mecanismos democráticos así como de los procedimientos de representación pública.

#### **2.- ÁREA DE CONDICIONES DE VIDA: Planes, programas y actividades relacionados con las políticas afirmativas.**

Serían indicadores de Calidad de Vida relacionados con desarrollo personal, salud y bienestar físico y psíquico, estilo de vida, así como aquellos relativos a los derechos de información, conocimiento y acceso a la cultura que no están expresados en los modelos anteriores, aunque pueden incluirse en el desarrollo personal.

#### **2.1.- Objetivos relacionados con la información** (visión generalista, no incluye objetivos relacionados con la información de programas concretos, por ejemplo “bolsa de vivienda”).

- 14.- *Centros de información juvenil. Cercanía a los jóvenes. Consultas y usuarios. Documentos registrados, actividades de difusión y mantenimiento de páginas Web, refuerzo del trabajo en red, formación de mediadores para difundir la información,...*
- 15.- Asesorías juveniles. Consultas, resultados, satisfacción usuarios.
- 16.- Publicaciones periódicas.
- 17.- Diseminación de la información mediante soportes publicitarios y tabloides.
- 18.- Guías, dossiers y folletos informativos.
- 19.- Diseminación de la información en los MCM.

#### **2.2.- Objetivos relacionados con la prevención de riesgos.**

- 21.- *Seguridad vial. Materiales diseñados, tipos de cursos o campañas realizadas, difusión y diseminación de la información, cursos de formación y sensibilización realizados, número de agentes y centros participantes, número de participantes, intervenciones individuales, convenios con entidades y auto-escuelas, intervenciones en medios, medida de los efectos, satisfacción de los participantes,..*
- 22.- Drogas, alcohol y tabaco.
- 23.- Sexualidad y ETS.
- 24.- Accidentes diversos.
- 25.- Suicidio juvenil.

#### **2.3.- Objetivos relacionados con la movilidad y el intercambio.**

- 26.- *Centros de Juventud. Actividades y usuarios.*
- 27.- Carnés joven.
- 28.- Campos de trabajo internacional. *Ofertas, acceso, propuestas, demandas, participación, grado de satisfacción, contactos internacionales establecidos, sinergias.*
- 29.- Intercambios y encuentros internacionales.
- 30.- Cooperación joven.
- 31.- Campos de trabajo monográficos.
- 32.- Campamentos.
- 33.- Actividades con y en la naturaleza.
- 34.- Rehabilitación rural, arqueología y
- 35.- Campamentos urbanos.

#### **2.4.- Objetivos relacionados con la creatividad y la innovación.**

- 37.- Inmersión y acceso a las TIC.

- 38.- Certámenes de jóvenes creadores (música, teatro, narrativa, cine, video, arte, poesía, fotografía, danza...). Participación.
- 39.- Difusión de creaciones juveniles. *Tipos de acción, materiales producidos, participación, impacto público.*
- 40.- Acceso y participación en actividades culturales.

**2.5.- Objetivos relacionados con el ocio** (incluye la gestión de los centros e instalaciones y otros equipamientos juveniles, aunque realicen programas y actividades relacionados con otros objetivos).

- 42.- Plazas y recursos en centros de juventud, albergues juveniles.
- 43.- Promoción de actividades deportivas. *Licencias de jóvenes, adecuación instalaciones y equipamientos, campeonatos juveniles, audiencia pública obtenida, satisfacción participantes.*
- 44.- Programas de ocio alternativo. *Propuesta de nuevas actividades, tipo de actividades, participación de voluntarios, audiencia y presencia en la actividad, efectos y consecuencias.*
- 45.- Transporte nocturno.

**3.- ÁREA DE EMANCIPACIÓN: Planes, programas y actividades relacionados con las políticas de transición efectivamente realizadas desde el ámbito de juventud.**

Serían los indicadores de Calidad de Vida relacionados, casi exclusivamente con auto-determinación (emancipación en la terminología de Juventud)

**3.1.- Objetivos relacionados con los déficits de formación.**

- 46.- Compensar los déficits de formación relacionados con el fracaso en el sistema educativo. *Número de cursos, número de alumnos, resultados de la formación, resultados en empleo o inserción, materiales didácticos y otros instrumentos, acciones de promoción del colectivo,...*
- 47.- Becas complementarias para prolongación de estudios. *Tipos de becas, número de becarios, utilidad y satisfacción.*
- 48.- Prácticas en empresas. *Tipo de prácticas, número de casos, funcionalidad de las mismas.*

**3.2.- Objetivos relacionados con el empleo.**

- 49.- Promoción de autoempleo y emprendedores. *Consultas, cursos de formación, participantes, contenidos, empleos o empresas creadas.*
- 50.- Asesoramiento en la búsqueda de empleo. *Consultas, Acceso a currículos, contratos y empleos obtenidos, duración de los mismos.*
- 51.- Creación de sociedades y cooperativas laborales. *Consultas, formaciones, proyectos presentados, entidades creadas y consolidadas en un plazo determinado, proyectos subvencionados.*
- 52.- Medidas de acción positiva dirigidas a jóvenes. *Normativa, beneficiarios, porcentaje de jóvenes en sectores y empresas aludidos o convenidos, porcentaje de formulas de trabajo especiales,...*

**3.3.- Objetivos relacionados con la vivienda y la residencia.**

- 53.- Bolsa de vivienda (alquiler y venta). *Anuncios facilitados, consultas, logros, condiciones,...*
- 54.- Promoción del acceso a la vivienda de alquiler. *Acciones, demandas, logros y subvenciones.*

- 55.- Hipoteca joven.
- 56.- Gestión acceso y promoción VPO. *Promociones, adjudicaciones a jóvenes en programas, cambios en la fiscalidad municipal,....*
- 57.- Alojamientos temporales y convivencia intergeneracional.

#### **3.4.- Objetivos relacionados con la integración social de colectivos con problemas específicos de marginalidad..**

- 58.- Con jóvenes infractores.
- 59.- Programas de calle y otros programas de mediación.
- 60.- Programas con minorías étnicas o grupos específicos de inmigrantes.

#### **3.5.- Objetivos relacionados con la promoción de la igualdad de trato y oportunidades.**

- 61.-Igualdad general de oportunidades.
- 62.-Igualdad económica.
- 63.-Igualdad de género.
- 64.- Igualdad según orientación sexual.
- 65.-Igualdad étnica, cultural y religiosa.
- 66.- Promoción de la interculturalidad.

#### **4.- ÁREA DE INVESTIGACIÓN: programas de investigación sobre juventud.**

Serían indicadores de Calidad de Vida Relacionados con el conocimiento de la realidad del colectivo y su aplicación a la vida individual.

##### **4.1.- Objetivos relacionados con la realización de investigaciones seriadadas y continuas sobre la evolución de la realidad juvenil.**

- 67.- Realización de Estudios. *Estudios finalizados, valoración de su continuidad, cumplimiento de criterios de periodicidad.*

##### **4.2.- Objetivos relacionados con el mantenimiento de centros de documentación de juventud.**

##### **4.3.- Objetivos relacionados con la realización de estudios monográficos sobre cuestiones puntuales o de relevancia.**

- 68.- Realización de Estudios. *Estudios finalizados.*

##### **4.4.- Objetivos relacionados con la difusión de los resultados de los estudios.**

- 69.- Actividades relacionadas con la divulgación. *Número de ejemplares, presentaciones, presencia en Medios,...*

NOTA 1: No todas las políticas de juventud incluyen la totalidad de las dimensiones expresadas en este esquema (aunque sí la mayoría), ni agrupadas de la misma manera. En todo caso este es un esquema estándar que sirve para sistematizarlas.

NOTA 2: No se trata de una propuesta exhaustiva pero ha tratado de ser lo más amplia posible. Los indicadores citados pretenden ser además meramente orientativos, ya que dependen de los objetivos particulares de cada uno de los programas.

NOTA 3: Se han excluido los posibles indicadores relacionados con el control de actividades cruzado entre administraciones.

NOTA 4: Son muchos los indicadores que se superponen lo que implica que el diseño definitivo y concreto de los mismos tiene que hacerse estableciendo un modelo de indicadores globales del conjunto de la evaluación, aunque previamente se hayan establecido matrices de evaluación en la que se indiquen el/los indicadores que se corresponden a cada objetivo.

### 16.3.3.- El manejo de los indicadores.

Este conjunto de dimensiones y los correspondientes indicadores sirven para valorar una política o un plan de juventud, pero también un programa o una acción más concreta, en este sentido son indicadores que miden **la efectividad** global de las políticas o planes, pero también la de programas, actividades o grupos de áreas o actividades unificados por un determinado objetivo general relacionado con la participación, la condición juvenil, la transición a la vida adulta o la investigación.

Podemos utilizar esta relación, índice o modelo (antes) para diseñar análisis de la realidad y entonces deberemos aportar información sobre todos las dimensiones y objetivos citados o que suponemos que van a formar parte de la intervención. Podemos hacer (antes) análisis de diseño o evaluabilidad. Podemos, una vez iniciada la intervención, utilizar el modelo, en concurrencia con los objetivos más específicos y los recursos fijados en la planificación, evaluar la implementación y los procesos (durante). Finalmente estos serían los indicadores para realizar una evaluación de resultados o de impacto (después).

Sobre los indicadores enunciados ¿Qué cosas podemos medir o determinar? En algunos es obvio porque se trata de cantidades recogidas como registro en la propia realización de las actividades. Pero también podemos incluir otras formas de medir y determinar. Por ejemplo:

- Para los análisis de cobertura observar la implantación territorial y social del proyecto o programa.
- Para la implementación y el impacto observar el grado de implicación de los diversos agentes sociales en relación a cada objetivo.
- Para el análisis de procesos observar cosas como el cumplimiento de los plazos, la evolución de los costes, el uso de los recursos el grado e cumplimiento de los objetivos intermedios,...
- Para el análisis de resultados se puede diseñar una investigación más o menos sofisticada que ponga en claro si el objetivo ha sido alcanzado (o superado o no se ha alcanzado).

Para hacer todo esto se puede recurrir a estrictos planteamiento digitales (si/no), a establecer las opiniones registradas en verbatines (que se pueden codificar o no) y por supuesto a construir escalas que otorguen puntuaciones, por ejemplo de uno a cinco puntos, todo para contestar las preguntas como ¿se utilizó o realizó algún estudio previo de la realidad o el diagnóstico? ¿Se tuvieron en cuenta estos resultados? ¿Se aplicaron de forma rigurosa los recursos disponibles?

¿Quiénes van a contestar las preguntas o como vamos a obtener los datos para obtener las respuestas? Pues depende, una evaluación general de una Política o un Plan de Juventud requiere muchos tipos de evaluación, pero podemos sistematizarlos todos (especialmente si hay evaluaciones y registros particulares de cada programa) en un cuestionario a expertos (los técnicos que han desarrollado el programa y algunos agentes que han participado de forma muy directa en el mismo).

Pero si por ejemplo tratamos de medir la audiencia, lo más normal es acudir a los registros preestablecidos, en cambio si tratamos de medir el grado de satisfacción de los usuarios, seguramente deberemos recurrir a una encuesta representativa. Por otra parte la mejor manera de contestar a la pregunta ¿queda algo por hacer o como podemos mejorar?, quizá se pueda resolver mediante grupos de discusión.

Asimismo para analizar todas las medidas y representaciones logradas podemos proceder de muy diversas maneras, Carol Weiss establece tres procedimientos alternativos (Weiss, 1982):

- 1.- Comparar los resultados obtenidos en otras evaluaciones, para establecer un criterio de mejora y realizar una atribución causal. Si son peores es que algo ha fallado y si son mejores es que se han introducido factores (del contexto o de la propia intervención) que explican el éxito. Este es el modelo más académico porque se supone que contribuye a expandir el conocimiento.

- 2.- Establecer los criterios de éxito y puntuar en escalas en función de los mismos. Si se puntúa al máximo se supone que se han alcanzado con brillantez todos los objetivos, si la media se sitúa por debajo de la mitad es que algo no ha funcionado. Esta es una visión típicamente empresarial aunque cada vez la están utilizando más las administraciones públicas.
- 3.- Digitalizar las opciones (si/no) en un protocolo de comprobación. En este caso se trata de determinar si una cosa se ha hecho o no, lo que exige una planificación muy detallada. Este ha sido el típico procedimiento de las administraciones públicas durante años.

Obviamente los dos últimos tipos de procedimientos exigen objetivos muy bien formulados, con plazos bien definidos e indicadores muy cerrados y codificados. Con estas condiciones también se pueden utilizar otras medidas, por ejemplo las de coste o de tiempo para lograr alcanzar un objetivo. No parece necesario aclarar que la información (el dato) que nos llega a establecer todo este tipo de respuestas puede ser tanto cualitativa como cuantitativa.

Sin embargo, la evaluación de los programas, las acciones y las intervenciones más concretas, se miden en términos **de eficacia y eficiencia**, lo que obliga a considerar indicadores concretos de procesos, como costes, dedicación, tiempos, audiencias... En todo caso, la medida de la eficacia y la eficiencia debe moverse dentro del contexto que ofrece la evaluación de la efectividad global.

Para realizar esta medida deben proponerse parámetros y criterios concretos en relación a cada indicador. Dichos parámetros nos sirven para establecer bien el grado de adecuación de la respuesta, la adecuación del marco de la actuación en la perspectiva de la planificación, el grado de cumplimiento o los efectos e impactos producidos. En los próximos años debería desarrollarse en los ámbitos de juventud una cierta cultura (a partir de experiencias) en torno a las maneras de establecer tales parámetros y criterios, que en todo caso deben corresponder al método de recogida y producción de datos que se haya elegido.

## Recopilación de ideas clave sobre la evaluación: la doble faz de la evaluación

Sin pretender establecer unas conclusiones, porque tiene poco sentido tratar de hacerlo en un Manual repleto de pequeñas informaciones y ejemplos, en el que todos tienen su importancia, vamos a tratar de reflejar las seis ideas clave que quizá denoten la identidad de esta propuesta de evaluación para los ámbitos de juventud.

### 17.1.- En torno a la noción de evaluación fuerte.

**El santo patrón de los evaluadores es Jano, un dios con único rostro pero con dos caras.** La evaluación es una práctica que se desarrolla, que ocurre en el terreno de la realidad social, de las políticas y de los programas efectivamente realizados y trata de determinar las condiciones de su efectiva implantación. Se trata de una práctica peculiar que a la vez utiliza de forma estricta y rigurosa el método científico. Esta doble faz refleja una noción fuerte de evaluación, ya que permite a la vez garantizar la fiabilidad y la aplicabilidad de los resultados, una opción lícita para otorgar valor a las políticas en una sociedad democrática y un mecanismo para ampliar la ciudadanía.

La evaluación no es investigación, ni es una tarea exclusiva de los evaluadores externos, sino que exige participación. Pero esta participación es contraria a la noción de espontaneidad: la evaluación es un procedimiento técnico riguroso, con conceptos claros y transparentes, aunque, por su propia complejidad, abierto a la innovación. En términos académicos la evaluación representa un área de conocimiento relativamente nueva pero con una gran proyección en el desarrollo social y la práctica profesional.

### 17.2.- En torno a las estrategias de evaluación.

**Para obtener buenas respuestas se necesita hacer las preguntas pertinentes.** Hay que tratar de comprender que la evaluación es parte del proceso histórico de expansión de la conciencia humana, al menos la parte del proceso que nos remite al aprendizaje de las preguntas pertinentes para contribuir a este proceso histórico. La doble faz de la evaluación se manifiesta también como una práctica de mejora cotidiana del trabajo de los profesionales, pero esta me-

jora solo puede producirse si se sostiene sobre la mejora global del bienestar de la sociedad y la conquista de nuevos derechos sociales.

La evaluación es de forma inevitable, y en especial cuando lo niega, un acto político, de la misma manera que desarrollar un programa social, proponer una innovación tecnológica, crear una infraestructura, abrir nuevas líneas de producción o fijar una nueva legislación, son actos políticos que influyen sobre la sociedad, su grado de bienestar (o de malestar), sus expectativas de futuro y el destino de la ciudadanía. Por este motivo la evaluación, lo mismo que cualquier otra acción política requiere rigor técnico y metodológico pero también un conjunto de criterios éticos que la limiten, la fundamenten y la apoyen. Esto es especialmente cierto en relación a la evaluación de las Políticas de Juventud.

### **17.3.- En torno a las técnicas y a los métodos de recogida y producción de datos.**

**Enrocarse en una metodología supone asumir que hemos perdido la partida del conocimiento.**

No existe un método de recogida y producción de datos estándar o indicado para la evaluación, todos valen aunque hay que tener en cuenta que cada uno produce tipos distintos de datos. A la vez todos son insuficientes. La noción de evaluación invita a aceptar la idea de que los métodos son complementarios, porque son insuficientes. Un investigador puede permitirse el lujo de especializarse en un método, un evaluador debe conocer todos los métodos y utilizarlos a conveniencia, aunque puede recurrir a “expertos” en caso de necesidad.

Todo argumento metodológico, cuantitativo o cualitativo, abierto o cerrado, participativo o jerarquizado, que pretenda presentarse como “el método de la evaluación” se desautoriza a sí mismo y entra en una zona de sospecha. Hace algunas décadas esta fue la pretensión de los diseños experimentales, lo cual les ha conducido hacia el limbo de las estrategias poco recomendables. Algunos intentos de convertir en dogma las estrategias cualitativas deberían correr la misma suerte. Para la evaluación, hay que repetirlo, los métodos **son complementarios porque son insuficientes**. Tras esta afirmación se agazapan dos necesidades: comprender que son complementarios, comprender que **todos** son insuficientes.

### **17.4.- En torno a los estándares de evaluación.**

**Si no podemos mejorar (ni imponer) un sistema alternativo de estándares, resulta poco recomendable rechazar las propuestas de los organismos internacionales.** La descripción de los estándares de evaluación permite entenderlos y comprender cómo se ha producido un proceso de progresiva acumulación y mejora. Existen diferentes sistemas de estándares que podemos colocar sobre una línea evolutiva de progresiva acumulación de mejoras, aunque en los últimos tiempos los cambios y las mejoras se están ralentizando porque los estándares son cada vez más completos y eficientes.

El modelo de estándares de evaluación para el Sistema de Naciones Unidas que figura en el anexo II es, seguramente, la mejor referencia disponible en este momento.

Tratar de utilizar un modelo alternativo requiere una justificación en torno a su mejor calidad o ajuste a los objetivos de la evaluación. Esto no significa que debamos renunciar al objetivo de las mejoras consustancial a la propia evaluación y en particular a la noción de evaluación fuerte.

### **17.5.- En torno a la acción evaluativa.**

**La evaluación requiere rigor, pero el rigor no se sustenta en una teoría o un marco conceptual único o exclusivo. La evaluación es un campo de conocimiento abierto y transversal.**

No existe, ni debe existir, una teoría de la evaluación que sostenga este campo de conocimiento y lo diferencie de otros. La identidad de la evaluación, aparte de lo mencionado en anteriores ideas clave, reside justamente en esta apertura y esta transversalidad.

La evaluación recoge conceptos, términos, técnicas y métodos, procedimientos y aportaciones de diferentes campos de conocimiento y los agrupa en un procedimiento tras-teórico y tras-disciplinar, que se aplica a partir de unas destrezas básicas y un conocimiento profesional relativamente variado. La pluralidad de los profesionales de la evaluación garantiza la continuidad y la identidad de la misma. Las aportaciones realizadas por todos estos profesionales, siempre que se orienten en relación a la propia pluralidad, refuerzan la noción de evaluación. En cambio las evaluaciones realizadas desde una orientación teórica, técnica o corporativa particular suponen abandonar la misión esencial de la propia evaluación. Ya no son evaluaciones, son investigaciones propias de un campo de conocimiento concreto.

## **17.6.- En torno a los indicadores para evaluar políticas de juventud.**

**Los indicadores no se inventan, no surgen de una inteligencia superior a través de un acto de creación, sino que son el resultado de la continua experiencia de su utilización.**

En el ámbito de las políticas de juventud las evaluaciones han sido escasas y sin apenas referencias mutuas. Los informes de las pocas evaluaciones realizadas no han sido publicados formalmente y los informes no publicados que forman parte de la “literatura gris” apenas han sido difundidos más allá de la propia institución que las ha encargado o realizado. En las bases de datos y en los centros de documentación sobre juventud aparecen muchos informes y libros que incluyen diagnósticos de la realidad juvenil, pero las evaluaciones son escasas y casi todas se refieren a temas muy concretos, como programas de educación o de exclusión social de jóvenes, pero realizados por ámbitos administrativos diferentes a los de juventud. La conclusión es clara: en el ámbito de juventud apenas se ha practicado la evaluación y falta una cultura evaluativa.

Los pocos ejemplos de evaluación de programas, a los que no ha sido fácil acceder, confirman esta conclusión: son casos aislados. Pero aunque están correctamente realizados, no constituyen una “base de partida” común, porque cada uno de ellos utiliza marcos conceptuales y estrategias particulares, sin citar ni las visiones conceptuales de referencia, ni por supuesto, las conclusiones de otras evaluaciones. No hay por tanto ningún trabajo acumulativo y parece que cada una de estas escasas evaluaciones ha sido realizada de forma aislada por equipos externos que carecen de una visión global de lo que significan las políticas y los ámbitos de juventud.

Sin embargo, en los organismos de juventud aparece una experiencia que conviene destacar y que compensa esta falta de implantación de la evaluación. Se trata de la habitual realización de memorias anuales que incluyen los datos de los análisis de la realidad, la descripción de los programas y datos relacionados con los procesos de implantación de los mismos.

En el contexto constituido por la falta de experiencia explícita en evaluación y una cierta experiencia en la preparación de “memorias”, hemos podido avanzar una primera definición de las dimensiones y los indicadores idóneos para evaluar las políticas, los planes y los programas de juventud. En todo caso, la conformación de un primer modelo provisional de indicadores para este ámbito, requiere la puesta en práctica del mismo, la experimentación y la mejora en una serie sucesiva de propuestas.

En el Manual se ha incluido la primera de estas propuestas para que, a partir de su contratación en la práctica, tales indicadores vayan madurando.

## BIBLIOGRAFÍA CITADA.

- AGUILAR, M. J. y ANDER-EGG, E. (1992), **Evaluación de servicios y programas sociales**, Madrid, Siglo XXI.
- AGUINAGA, J. (2004), **El precio de un hijo: los dilemas de la maternidad en una sociedad desigual**, Madrid, Debate.
- AGUINAGA, J. y COMAS, D. (1990), **Infancia y adolescencia: la mirada de los adultos**, Madrid, MAS.
- AGUINAGA, J. y COMAS, D. (1995), **En busca de la complementariedad**, Medellín, Paisajoven.
- AGUINAGA, J. y COMAS, D. (1997), **Cambios de Hábitos en el uso del tiempo**, Madrid, INJUVE.
- ALVARO, M. (1988), “Evaluación externa de la reforma experimental de las enseñanzas medias”, en REVISTA DE EDUCACIÓN, nº 287.
- ALVIRA, F. (1991), **Metodología de la evaluación de programas**, Madrid, CIS.
- ANGUERA, M.T. (1985), **Metodología de la observación en ciencias humanas**, Madrid, Cátedra.
- ARNAU, J (1978), **Métodos de investigación en las ciencias humanas**, Barcelona, Omega.
- AUTORES VARIOS (1993), **Investigación-acción participativa**, monográfico de **Documentación social**, número 92, Madrid, CÁRITAS.
- AZOFRA, M. J. (1999), **Cuestionarios**, Madrid, CIS.
- BABBIE, E. (1995), **Manual para la práctica de la investigación social**, Bilbao, Desdée De Brouwer, 1996.
- BALLART, X. (1992), **¿Cómo evaluar los programas y servicios públicos? Aproximación sistemática y estudio de caso**, Madrid, MAP.
- BAÑON, I. (2002), **La evaluación de la acción y las políticas públicas**, Madrid, Díaz de Santos.
- BARCIKOWSKL R.S. y ROBEY, R.R. (1994), “Uso y abuso de los diseños de medidas repetidas” en AUTORES VARIOS (1994), **Procesos en el análisis de datos para la investigación de intervenciones preventivas**, Madrid, FAD.
- BARDIN, L. (1977), **Análisis de contenido**, Madrid, AKAL.
- BASTIDE, R. (1971), **Antropología aplicada**, Buenos Aires, Amorrortu.
- BERMEJO, J. C. (2007), **La aurora de los enanos: decadencia y caída de las universidades europeas**, Madrid, Akal.
- BERTCHER, H. y MAPLE, F.F. (1996), **Creating Groups**, Londres, SAGE.
- BIRKENBIHL, M. (2000), **Formación de formadores**, Madrid, Paraninfo.
- BOSCH, J. L. y TORRENTE, D. (1993), **Encuestas telefónicas y por correo**, Madrid, CIS.
- BOURDIEU, P. (1997), **Razones prácticas. Sobre la teoría de la acción**. Barcelona, Anagrama.
- BOYER, A. (1994), “De la juste mesure”, en BEAUNE, J.C. (1994), **La mesure, instruments et philosophies**, París, Milieux.
- BUNGE, M. (2001), **La investigación científica**, México, Siglo XXI.
- BUNGE, M. (2002), **Crisis y reconstrucción de la filosofía**, Barcelona, Gedisa.

- BUSTELO, M (2004), **La evaluación de las políticas de género en España**, Madrid, Catarata.
- BUSTELO, M. (1999), “*Diferencias entre evaluación e investigación: una distinción necesaria para la identidad de la evaluación de programas*”, en REVISTA ESPAÑOLA DE DESARROLLO Y COOPERACIÓN, Nº 4.
- BUSTELO, M. (2002), “*¿Qué tiene de específico la metodología de la evaluación?*”, en Bañón, I. (2002), **La evaluación de la acción y de las políticas públicas**, Madrid, Díaz de Santos.
- CALLEJO, J. (1998), “*Sobre el uso conjunto de prácticas cualitativas y cuantitativas*” en REVISTA INTERNACIONAL DE SOCIOLOGÍA, Nº 21, Córdoba, IESA/CSIC.
- CAMPBELL, D. (1979), **Métodos cuantitativos y cualitativos en investigación evaluativa**, Madrid, Morata, 2000.
- CAMPBELL, D.T. y STANLEY, J.C., (1966), **Diseños experimentales y cuasi-experimentales en investigación**, Buenos Aires, Amorrortu, 1973.
- CANALES, M. y PEINADO, A. (1995), “*Grupos de discusión*” en DELGADO, J. M. y GUTIÉRREZ, J. (1995) **Métodos y técnicas cualitativas de investigación en ciencias sociales**, Madrid, Síntesis.
- CAPLOW, T. (1977), **La investigación sociológica**, Barcelona, Laia.
- CARMONA, J.A. (1977), **Los indicadores sociales hoy** Madrid, CIS.
- CEA D'ANCONA, M<sup>a</sup>. A. (1996). **Metodología cuantitativa. Estrategias y técnicas de investigación social**. Síntesis Sociología. Madrid.
- CEMBRANOS, F. (1994), **La animación sociocultural: una propuesta metodológica**, Madrid, Popular.
- CHELIMSKY, E. y SHADISH, W. (1997), **Evaluation for the 21<sup>st</sup> century**, Londres, SAGE.
- CICOUREL, A. V. (1964), **El método y la medida en sociología**, Madrid, Editora Nacional, 1984.
- COHEN, E. y FRANCO, R. (1993), **Evaluación de Proyectos Sociales**, Madrid, Siglo XXI.
- COMAS, D. (1986), “*La medida de la incidencia, prevalencia y problemas causados por drogas ilegales*”, en REÍS, número 34, Madrid, CIS.
- COMAS, D. (1992), **Los estudios de seguimiento: una experiencia metodológica**, Madrid, GrupoGID.
- COMAS, D. (1993), **Los jóvenes y el uso de drogas en la España de los años 90**, Madrid, INJUVE.
- COMAS, D. (1995), “*Los métodos de investigación en la evaluación de programas de prevención*” en AUTORES VARIOS (1995), **Metodología de la evaluación de los programas de prevención del consumo de drogas**, Cuenca, Universidad Internacional Menéndez y Pelayo.
- COMAS, D. (2001), **La evaluación de programas de ocio alternativo de fin de semana**, Madrid, INJUVE.
- COMAS, D. (2002), Coordinador, **Jóvenes y estilos de vida: valores y riesgos de los jóvenes urbanos**. Madrid, INJUVE-FAD
- COMAS, D. (2004), **Los adolescentes en Hirusta, resultados de una evaluación**, Bilbao, Fundación Gizakia.

- COMAS, D. (2005), *“La evaluación de planes, programas e intervenciones: una opción pragmática y realista”*, en AUTORES VARIOS (2005), **Manual para Agentes de Igualdad**, Madrid.
- COMAS, D. (2008), *“Una mirada ideológica (y política) al consumo”*, en AUTORES VARIOS (2008), **Problemas de drogas aquí y ahora**, Madrid FAD.
- COMAS, D.; ROLDAN, G. y GARCIA, (1996), *“Políticas sobre drogas a partir de un estudio longitudinal retrospectivo”*, en ADICCIONES, volumen 8; nº 2.
- COMITÉ DE AYUDA AL DESARROLLO/OCDE (1993), **Principios del CAD para una Ayuda Eficaz**, Madrid, Mundi Prensa, 1995.
- COOK, T. D. y REICHARDT, C. (1979), **Métodos cualitativos y cuantitativos de investigación evaluativa**, Madrid, Morata, 1988.
- CRÉPEL, P. (1994), *“Comment la mesure en arithmétique politique est venue à Condorcet”* en BEAUNE, J.C. (1994), **La mesure, instruments et philosophies**, París, Millieux.
- CRESSWELL, R. y GODELIER, M. (1975), **Útiles de encuesta y de análisis antropológico**, Madrid, Fundamentos, 1981.
- CRONBACH, L. (1982), **Designing evaluations of educational and social programs**, San Francisco, Jossey-bass.
- CROZIER, M. y FRIEDBERG, E. (1977), **L`acteur et le système: les contraintes de l`action collective**, París, Éditions du Seuil.
- DE MIGUEL, A.; DIEZ NICOLÁS, J y MEDINA, A. (1967), **Tres estudios para un sistema de indicadores sociales**, Madrid, Euramerica.
- DEL CAMPO, S. (1972), **Los indicadores sociales a debate**, Madrid, Euramerica.
- DENZIN, N. (1978), **The research act. A theoretical introduction to sociological methods**, Nueva York, Mc Graw Hill.
- DENZIN, N. K. y LINCOLN, Y. (2000), **Handbook of Qualitative Research**, Londres, SAGE.
- DEPARTAMENTO DE DESARROLLO Y SERVICIOS DE CALIFORNIA (1999), **Calidad de vida, calidad de servicio**, Madrid, IMSS.
- DÍAZ, A.; BARRUTI, M. y DONCEL, C. (1992), **Las líneas del éxito**, Barcelona, Laboratorio de sociología de la ICESB/Ayuntamiento de Barcelona.
- DIEZ NICOLÁS, J. (1968), **Análisis comparado de indicadores sociales y políticos**, Madrid, Euramerica.
- DRUKER, P. (1995), **Las cinco preguntas más importantes que debe formularse sobre su ONG: Manual de auto-evaluación**, Buenos Aires, Granica.
- DURKHEIM, E. (1897), **El suicidio**, Buenos Aires, Schapire, 1971.
- ECKHARDT (1977), **Social Research Methods. Perspective, theory and analysis**, New York, Random House.
- EVANS, J. R. y LINDSAY, W.R (2005), **Administración y control de calidad**, Buenos Aires, Thomson.
- FABIA, M. (1994), **Técnicas de grupo para la cooperación**, Barcelona, CEAC.
- FANTOVA, F. (2001), **La gestión de organizaciones no lucrativas. Herramientas para la intervención social**, Madrid, CCS.
- FANTOVA, F. (2004), **El discurso de la calidad en la gestión social: narraciones y cristalizaciones**, Barcelona, Universidad de Barcelona.
- FERNÁNDEZ BALLESTEROS, R. (1995), **Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud**, Madrid, Síntesis.

- FERNANDEZ, S. (2001), "*Análisis de efectividad en la evaluación de programas sociales y educativos*", en REVISTA DE EDUCACIÓN, nº. 324.
- FETTERMAN, D. M. y PITMAN, M. A. (1986), **Educational evaluation. Ethnography in theory, practice and politics**, Londres, SAGE.
- FÓRUM DE POLÍTICA FEMINISTA (2007), **Aplicación de los indicadores de género en las políticas públicas**, Madrid, FPF.
- FOSTER, G. M. (1974), **Antropología aplicada**, México, Fondo de Cultura Económica.
- FOWLER F.J. y MANGIONE, T.W. (1990), **Standardized Survey Interviewing**, London, Sage.
- FREUND, J. (1973), **Las teorías de las ciencias humanas**, Barcelona, Península.
- FROUFE, S. y SANCHEZ, M. A. (1994), **Construir la animación sociocultural**, Salamanca, Amarú.
- FUNES, J. y ROMANÍ, O. (1985), **Dejar la heroína**, Madrid, DGAS/Cruz Roja
- GALVEZ, R.; DELGADO, M. y BUENO, A. (1988), "*Epidemiología descriptiva*" en PIEDROLA, G. (1988), **Medicina preventiva y salud pública**, Barcelona, Salvat.
- GAMELLA, J. F. y ÁLVAREZ, A. (1997), **Drogas de síntesis en España, Patrones y tendencias de adquisición y consumo**, Madrid, Delegación del Gobierno para el PNSD.
- GARCIA, E. (2003), **La evaluación de programas de Reforma Educativa en España: tres estudios de caso desde un enfoque de metaevaluación**, Madrid, UCM.
- GARDE, J. A. (2007), "*La creación de la agencia de evaluación es España*", en REVISTA INFORMACIÓN COMERCIAL ESPAÑOLA, nº. 836, monográfico sobre evaluación de las políticas públicas.
- GARLAND, D. (2001), **La cultura del control**, Barcelona, Gedisa, 2005.
- GARNIER, L y PRZEWORSKI, A. (2004), **Política y gestión pública**, México, FCE.
- GEERTZ, C. (1973), "*Juego profundo: notas sobre la riña de Gallos en Bali*", en GEERTZ, C. (1973), **La interpretación de las culturas**, Madrid, GEDISA, 1987.
- GIDDENS, A. (1993), **Modernidad e identidad del yo: el yo y la sociedad en la época contemporánea**, Barcelona, Península, 1997.
- GOETZ, J. P. y LECOMPTE, M. D. (1984), **Etnografía y diseño cualitativo en investigación educativa**, Madrid, Morata, 1988.
- GONZÁLEZ DEL RÍO, M. J. (1997), **Metodología de la investigación social, técnicas de recolección de datos**, Alicante, Aguaclara.
- GOODENOUGHM, W. (1971), "*Cultura, lenguaje y sociedad*", en KAHN, J.S. (1974), **El concepto de cultura: textos fundamentales**, Barcelona, Anagrama, 1975.
- GOVERNMENT STATISTICIANS COLLECTIVE (1993), "*How official statistics are produced: Views from the inside*", en HAMMERSLEY, M. (1993), **Social Research: Philosophy, politics and practice**. Londres, Sage.
- GRANGER, G.G. (1989), **La mathématique sociale du Marquis de Condorcet**, París, Editions Odile Jacob.
- GREBENIK, E. (1968), "*Estadísticas vitales*", en SILLS, D.L. (1968), **Enciclopedia internacional de las ciencias sociales**, Madrid, Aguilar, 1974.
- GRUPO GID (1994), **La evaluación de programas en drogodependencias**, Madrid, GrupoGID.
- GUASCH, O. (1997), **La observación participante**, Madrid, CIS.
- GUBA, E. e LINCOLN, Y. (1989), **Fourth generation evaluation**, Londres, SAGE.

- HARNOLL, R, y otros (1985), **Problemas de drogas: valoración de necesidades locales**, Madrid, Delegación del Gobierno para el PNsD, 1991.
- HERMAN, J. I., (1987), **Program evaluation Kit**, Beverly Hills, Sage.
- HOUSE, E.R., (1994), **Evaluación, ética y poder**, Madrid, Morata.
- IBÁÑEZ, J. (1979), **Más allá de la sociología. El grupo de discusión. Técnicas y crítica**, Madrid, Siglo XXI.
- INCIARDI, J. A. (1993), **Drug treatment and criminal justice**, Londres, SAGE.
- INCIARDI, J. A., McBRIDE, D.C. y RIVERS, J.E. (1996), **Drugs control and the courts**, Londres, SAGE.
- INCIARDI J. A. (1994), **Screening and assessment for alcohol and other drug abuse among adults in the criminal Justice System**, Washington, CSAT.
- INJUVE (2000), **La juventud con Europa III: Evaluación de su aplicación en España**, Madrid, INJUVE.
- INSTITUTO DE LA MUJER (1998), **Evaluación de las políticas de igualdad**, Madrid, IM.
- ITURRIAGA, R. (2002), *“La evaluación como forma de control externo de las administraciones públicas”*, en BAÑON, I. (2002), **La evaluación de la acción y las políticas públicas**, Madrid, Díaz de Santos.
- JOHNSON, J.C. (1990), **Selecting ethnographic informants**, Londres, SAGE.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION, (1994), **Estándares para la evaluación de programas**, Bilbao, Mensajero, 1999.
- KANT, I. (1783), **Prolegómenos a toda metafísica futura que querrá presentarse como ciencia**, Madrid, Alambra, 1994.
- KETELE, J.M. y ROEGEERS, X. (1993), **Metodología para la recogida de información**, Madrid, La muralla, 1995.
- KEYFITZ, N. (1968), *“Estadísticas oficiales”*, en SILLS, D.L. (1968), **Enciclopedia internacional de las ciencias sociales**, Madrid, Aguilar, 1974.
- KOTTAK, C. P. (1993), **Antropología**, Madrid, McGraw-Hill, 1994.
- LAZARSELD, P. y LANDAU, D. (1968), *“Quetelet, Adolphe”*, en SILLS, D.L. (1968), **Enciclopedia internacional de las ciencias sociales**, Madrid, Aguilar, 1974.
- LAZARSELD, P.; BERELSON, B. y GAUDET, H. (1944), **The people choice: How the voter makes up his mind in a presidential campaign**, New York, Columbia University Press.
- LANDETA, J. (1999), **El método Delphi**. Ariel, Barcelona.
- LE PLAY, F. (1879), **La méthode sociale**, París, Meridiens Klincksieck, 1989.
- LÉCUYER, B. y OBERSCHAL, A. (1968), *“Orígenes de la investigación social”*, en SILLS, D. L. (1968), **Enciclopedia internacional de las ciencias sociales**, Madrid, Aguilar, 1974.
- LIGERO, J. A. (2001), *“Propuesta para un nuevo criterio de evaluación: la cobertura”*, en REVISTA ESPAÑOLA DE DESARROLLO Y COOPERACIÓN, nº 8.
- LIGERO, J. A. (2005), *“La evaluación y la limitación de las políticas de bienestar social”*, en Santiago de Chile, X congreso de la CLAD.
- LIGERO, J. A. (2004), *“Evaluación y calidad de los programas y los servicios públicos”*, en REVISTA NEXO, Nº 11.
- LÓPEZ DE CEBALLOS, P., MERLO, J. y GARCÍA, H. (2001). **Un método de evaluación formativa en el campo social**. Editorial Popular. Madrid.
- LUHMANN, N. (1990), **La ciencia de la sociedad**, Barcelona, Anthropos, 1999

- MACKENZIE, D. L. y DUCMDA, C. (1994), **Drugs and crime: evaluating public policy initiatives**, Londres, SAGE.
- MANZANO, V. G.; ROJAS, A. J. y FERNÁNDEZ, J. S. (1996). **Manual para entrevistadores**, Barcelona, ARIEL.
- MAP (2004), **Guía de autoevaluación para la administración pública. Modelo EFQM de excelencia**, Madrid, MAP.
- MARTÍN SERRANO, M. (1978), **Métodos actuales de investigación social**, Madrid, Akal.
- MARTÍN SERRANO, M. (1986), **La producción social de la Comunicación**, Madrid, Alianza.
- MERINO, J. V. (2000), **Programas de animación sociocultural: tres instrumentos para su diseño y evaluación**, Madrid, Nancea.
- MONDRAGÓN, J. (2002) “*La evaluación de los programas públicos: objeto, contenido y aplicaciones de diferentes instrumentos de evaluación*”, en BAÑÓN, I. (2002), **La evaluación de la acción y las políticas públicas**, Madrid, Díaz de Santos.
- MONNICKENDAM, M. (2002), “*Implementación de un método participativo para crear sistemas informáticos orientados al usuario de los servicios sociales*”, en REVISTA DE TRABAJO SOCIAL, nº. 165.
- MONNIER, E. (1995). **Evaluación de la acción de los poderes públicos**, Madrid, Instituto de Estudios Fiscales.
- MORA, J. G. (1991), **Calidad y rendimiento de las instituciones universitarias**, Madrid, Consejo de Universidades.
- MORENO, J.L. (2007), **Guía de aplicación del modelo EFQM de excelencia en entidades de acción social**, Madrid, Fundación Luís Vives.
- MUNICIO, P. (2002), **Herramientas para la evaluación de la calidad**, Madrid, Praxis.
- MURDOCK, G.P. y otros (1954), **Guía para la clasificación de los datos culturales**, Washington, OEA, 1966.
- NACHMIAS, D.(1976), **Research Methods in the Social Science**, Londres, Arnold.
- NACIONES UNIDAS (2004), **Haciendo realidad los compromisos: una guía juvenil para evaluar políticas nacionales de juventud**, Nueva York, Naciones Unidas.
- NAOHUM, CH. (1986), **La entrevista psicológica**, México, Trillas.
- NAVARRO, P. (1983), “*Las herramientas ‘familiares’ del trabajo de campo: el censo y la genealogía*”, en REÍS, número 21, Madrid, CIS.
- NIRENBERG, O. (2000), **Evaluar para la transformación. Innovaciones en la evaluación de programas y proyectos sociales**, Buenos Aires, Paidós.
- NOELLE, E (1963), **Encuestas en la sociedad de masas**, Madrid, Alianza, 1970.
- NORAD (1990), **El Enfoque del Marco Lógico**, Madrid, UCM.
- ORTÍ, A. (1986), “*La apertura y el enfoque cualitativo o estructural: la entrevista abierta semidirectiva y la discusión de grupo*”, en GARCÍA FERRANDO, M.; IBÁÑEZ, J. y ALVIRA, F. (1986), **El análisis de la realidad social: métodos y técnicas de investigación**, Madrid, Alianza.
- ORTÍ, A. (1993) “*El proceso de investigación de la conducta como proceso integral: complementariedad de las técnicas cuantitativas y cualitativas en el análisis de las drogodependencias*” en AUTORES VARIOS **Las drogodependencias: perspectivas sociológicas actuales**, Madrid, Colegio Licenciados en CP y S, 1993.
- ORTÍ, A. (1995), “*La confrontación de modelos y niveles epistemológicos en la génesis e historia de la investigación social*” en DELGADO, J. M. y GUTIÉRREZ, J. (1995), **Métodos y técnicas cualitativas de investigación en ciencias sociales**, Madrid, Síntesis.

- OSBORNE, D. y PLSTRIK, P. (2003), **Herramientas para transformar el gobierno**, Barcelona, Paidós.
- OVEJERO, F. (1987), **De la naturaleza a la sociedad**, Barcelona, Península.
- PALLARES, J. (1996), **El placer del escorpión**, Lérida, Mileniura.
- PALLARES, M. (1981), **Técnicas e instrumentos de evaluación**, Barcelona, CEAC.
- PALUMBO, D.J. (1987), **The politics of program Evaluation**, Newbury, SAGE.
- PASCUAL, J. (1999), **La evaluación de políticas y proyectos: criterios de valoración económicos y sociales**, Barcelona UAB / Icaria
- PATTON, M.Q. (1990), **Qualitative evaluation and Research Methods**, Londres, SAGE.
- PATTON, M.Q. (1997), **Utilization focuset evaluation**, Londres, SAGE.
- PATXOT, C. y FARRE, R. (2007), **Evaluación de la sostenibilidad del Estado de Bienestar en España**, Madrid, Ediciones 2010.
- PAWSON, R. y TILLEY, N., (1997), **Realistic evaluation**, Londres, SAGE.
- PLAYER, M. (1990), **Employment Discriminaron Law: Cases and Materials**, San Paul, West Publising.
- POMED, L.A. (1989), **El derecho de acceso de los ciudadanos a los archivos y registros municipales**, Madrid, INAP.
- PONS, I. (1993), **Programación de la investigación social**, Madrid, CIS.
- POPPER, K. (1934/1959), **La lógica de la investigación científica**, Madrid, Tecnos, 1973.
- PUJADAS, J.J. (1992), **El método biográfico: el uso de historias de vida en las ciencias sociales**, Madrid, CIS.
- QUINTERO, M. y PEREZ, A. (1988), **Evaluación: un instrumento de superación**, Bogota, Alcaldía Mayor.
- RIESSMAN, C.K. (1993), **Narrative analysis**, Londres, SAGE
- RODRIGUEZ-VILLASANTE, T. (2007), **Desbordes creativos**, Madrid, Catarata.
- ROOTMAN, I. y HUGHES, P.H. (1983), **Abus des drogues: systémes de notification**, Ginebra, OMS.
- ROSELLÓ, D. (2004), **Diseño y evaluación de proyectos culturales**, Barcelona, Ariel.
- ROSSI, I. y OGIGGINS, E. (1980), **Teorías de la cultura y métodos antropológicos**, Barcelona, Anagrama, 1981.
- ROSSI, P.H. y FREEMAN, H. E. (1989), **Evaluación. Un enfoque sistemático para programas sociales**, México, Trillas.
- RUANO, J. (2002), "*La doble naturaleza técnica y política de la evaluación de rendimiento*", en BAÑON, I. (2002), **La evaluación de la acción y las políticas públicas**, Madrid, Díaz de Santos.
- RUBIO, M.J. y VARAS, J. (1997), **El análisis de la realidad en la intervención social. Métodos y técnicas de investigación**, Madrid, CCS.
- RUIZ DE OLABUÉNAGA, J. I. (1996), **Metodología cualitativa**, Bilbao, Universidad de Deusto.
- RUIZ OLABUÉNAGA, J. e ISPIZUA , M. A. (1989), "*La técnica Delphi*", en RUIZ OLABUÉNAGA, J. e ISPIZUA , M. A. (1989), **La descodificación de la vida cotidiana. Métodos de investigación cualitativa**. Bilbao, Universidad de Deusto.
- SAGARDOY DE SIMÓN, I. y ALONSO, M. (1996), **La relación de trabajo de los empleados de los registros**, Madrid, Civitas.
- SANZ, A. (1956), **Resumen histórico de la estadística en España**, Madrid, INE.

- SANZ, E. (1994), **Manual de estudios de usuarios**, Madrid, Fundación Sanchez Ruperez / Pirámide.
- SARABIA, B. (1989), "*Documentos personales: historias de vida*" en GARCÍA FERRANDO, M.; IBÁÑEZ, J. y ALVIRA, F. (1989), **El análisis de la realidad social**, Madrid, Alianza.
- SARTORI, G. y MORLINO, L. (1991), **La comparación en las ciencias sociales**, Madrid, Alianza.
- SCHALOCK, R. y VERDUGO, M.A. (2003), **Calidad de vida. Manual para profesionales de la educación, salud y servicios sociales**, Madrid, Alianza Editorial.
- SCHNAÏTH, N. (1999), **Paradojas de la representación**, Barcelona, Café Central.
- SCHWARTZ, H. y JACOBS, J. (1976), **Sociología cualitativa. Método para la reconstrucción de la realidad**, México, Trillas, 1984.
- SCRIVEN, M. (1991), **Evaluation thesaurus**, Newbury Park, SAGE
- SEGRAEUS, V. (1994), **Compulsory measures for treatment and rehabilitation in sweden**, Estocolmo, SiS.
- SELLTIZ, C. y WRIGHTSMAN, S.W. (1975), **Métodos de investigación en las Ciencias Sociales**, Madrid, RIALP, 1980.
- SELLTIZ, C.; JAHODA, M y otros. (1964), **Métodos de Investigación en las relaciones sociales**, Madrid, RIALP, 1974.
- SHAW, I. F., (1999), **La evaluación cualitativa**, Barcelona, Paidós, 2003.
- SIMON, H. (1999), **Evaluación democrática de las instituciones escolares**, Madrid, Morata.
- STAKE, R. (1975), **Program Evaluation, Particularly Responsible Evaluation**, Michigan, Western University,
- STOETZEL, J. y GIRARD, A. (1973), **Las encuestas de opinión pública**, Madrid, IOP.
- STUFFLEBEAM, D.L. (1971), **Educational Evaluation an Decision Making**, Itaska, Peacock.
- STUFFLEBEAM, D.L. y SHIRKFIELD, A. J. (1987), **Evaluación sistemática: guía teórica y práctica**, Barcelona, Paidós.
- SUBIRATS, J. (1989), **Análisis de políticas públicas y eficacia de la administración**, Madrid, INAP.
- SWEDISH NATIONAL BOARD OF INSTITUTIONAL CARE -SiS- (1994), **Action plan for research and development and basic facts**, Estocolmo, SiS.
- SWOBODA, H. (1975), **El libro de la estadística moderna**, Barcelona, OMEGA.
- TAYLOR, S. J. y BOGDAN, R. (1984), **Introducción a los métodos cualitativos de investigación**, Buenos Aires, Paidós, 1986.
- THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981), **Normas de evaluación para programas, proyectos y material educativo**, México, Trillas, 1988.
- THOMAS, I. y ZNANIECKI, F. (1918), **El campesino polaco en Europa y América**, Madrid, CIS, 2004.
- TOBLER, N. (1994), "*Aspectos del meta-análisis en la investigación de intervenciones preventivas*" en AUTORES VARIOS (1994), **Procesos en el análisis de datos para la investigación de intervenciones preventivas**, Madrid, FAD.
- TORRES, L. y PARRA, V. (2002), "*Construcción de indicadores para las políticas públicas*", en BAÑON, I. (2002), **La evaluación de la acción y las políticas públicas**, Madrid, Díaz de Santos.
- TRINIDAD, A. y PÉREZ, M. (1998), "*Evaluación de programas de intervención social*", en REVISTA DE TRABAJO Y ASUNTOS SOCIALES, nº 10.

- VALLÉS, M. (2002). **Entrevistas cualitativas**. CIS. Madrid.
- VEDUNG, E. (1997), **Evaluación de políticas públicas y programas**, Madrid, MTAS.
- VIDAL, J. (2007), **Instrumentos para la gestión de calidad en las ONGs: perspectiva internacional**, Madrid, Fundación Luís Vives.
- VIDAL, P.; VILLA, A.; SUREDA, M; ALBINYANA, M. y VIDAL, L (2006), **El movimiento asociativo juvenil: Escuelas de ciudadanía y valoración social de los aprendizajes en las organizaciones juveniles**, Madrid, CJE.
- VINER, J. (1968), "*Pensamiento mercantilista*", en SILLS, D.L. (1968), **Enciclopedia internacional de las ciencias sociales**, Madrid, Aguilar, 1974.
- VIÑAS, V. (2001), "*El uso de técnicas cualitativas en la evaluación de programas: los programas de desarrollo regional financiados por la Unión Europea*" en REIS, N° 95.
- WAGNER, P.; WEIS, C. y otros, (1991), **Las ciencias sociales y los estados modernos**, México, FCE, 2001.
- WEISS, C. (1982), **Investigación evaluativa**, México, Trillas.
- WEISS, C. (1998), **Evaluación**, Nueva Jersey, Prentice-Hall.
- WERT, J. I. (1996), "*La encuesta telefónica*", en GARCÍA FERRANDO, M.; IBÁÑEZ, J. y ALVIRA, F. (1996), **El análisis de la realidad social. Métodos y técnicas de investigación**, Madrid, Alianza.
- WHOLEY, J. ; HATRY, H. y NEWCOMER, K. (1994), **Handbook of practical program evaluation**, San Francisco, Jossey-Bass.
- WINNICOTT, P. (1971), **Realidad y juego**, Barcelona, Gedisa, 1979.
- WOODS, P. (1986), **La etnografía en la investigación educativa**, Barcelona, Paidós/MEC, 1987.
- WRIGHT, CH. (1968), "*Evaluación de resultados*", en SHILLS, D. (1968), **Enciclopedia Internacional de las Ciencias Sociales**, Madrid, Aguilar, 1974.
- YOUNG, J. (2003), **La sociedad "excluyente". Exclusión social, delito y diferencia en la modernidad tardía**, Madrid, Marcial Pons.
- ZUFIAURRE, B. y ALBERTÍN, A. M. (2006), **Evaluación democrática y proyectos cooperativos**, Pamplona, UPN.

## ANEXO 1

### ESTÁNDARES DE EVALUACIÓN DEL JOINT COMITEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981-1994)

#### ESTÁNDARES DE UTILIDAD

La evaluación debe ayudar a las personas o grupos interesados a que éstos tengan acceso a las informaciones que necesitan para el desempeño de su actividad. Debe ayudarles a identificar los defectos y virtudes de lo evaluado y proponer soluciones a fin de mejorar el objeto que se evalúa. Este aspecto de utilidad abarca los siete estándares siguientes:

- U1 Identificación de los interesados.** Debería identificarse a las personas implicadas o afectadas por la evaluación para poder dar respuesta a sus necesidades.
- U2 Credibilidad del evaluador.** Las personas que llevan a cabo la evaluación han de ser dignas de confianza y competentes como evaluadores, de modo que sus resultados tengan una credibilidad y aceptación máxima.
- U3 Foco y selección de la información.** La información recogida debería seleccionarse en términos generales para responder a cuestiones pertinentes del programa y a los intereses de los clientes y otros interesados que se especifiquen.
- U4 Identificación de valores.** Deberían describirse cuidadosamente los enfoques, procedimientos y fundamentos utilizados para interpretar los resultados de modo que queden claras las bases de los juicios de valor.
- U5 Claridad del informe.** Los informes de evaluación deberían describir claramente el programa evaluado, incluyendo el contexto y el propósito, procedimientos y resultados de la evaluación, de modo que se proporcione información esencial y se entienda fácilmente.
- U6 Oportunidad y difusión del informe.** Deberían divulgarse los resultados provisionales significativos y los informes de evaluación a los usuarios destinatarios, de manera que puedan utilizarlos en el momento oportuno.
- U7 Impacto de la evaluación.** Las evaluaciones deberían ser planificadas, llevadas a cabo, e informar de ellas de modo que fomenten su seguimiento por parte de los interesados, de manera que aumente la probabilidad de que se utilice la evaluación.

## ESTÁNDARES DE VIABILIDAD

Los procedimientos para llevar a cabo la evaluación no deben ser excesivamente complicados y sí eficientes. Han de garantizar que la evaluación se lleve a cabo de un modo realista, prudente y diplomático, que se hayan ponderado todos los puntos y tenido en cuenta los costes. Los estándares de viabilidad son los siguientes:

- V1 Procedimientos prácticos.** Los procedimientos de evaluación deberían ser prácticos para reducir al mínimo las interrupciones y molestias mientras que se obtiene la información necesaria.
- V2 Viabilidad política.** La evaluación debería planificarse y llevarse a cabo anticipando las diferentes posiciones de los diversos grupos de interés, de manera que pueda lograrse su cooperación y puedan evitarse o contrarrestarse los posibles intentos de cualquiera de estos grupos para limitar las actividades de evaluación o para sesgar o usar indebidamente los resultados.
- V3 Eficacia de costos.** La evaluación debería ser eficiente y producir informaciones con el suficiente valor, de modo que puedan justificarse los recursos empleados.

## ESTÁNDARES DE PROPIEDAD O DECORO

La evaluación debe estar basada en compromisos explícitos que aseguren la necesaria cooperación, la protección de los derechos de las partes implicadas y la honradez de los resultados, además debe proporcionar un informe equilibrado que ponga de manifiesto tanto aciertos como errores del objeto que se evalúa. Este aspecto de propiedad abarca los ocho estándares siguientes:

- P1 Orientación hacia el servicio.** Las evaluaciones deberían diseñarse para ayudar a las organizaciones a atender y responder con eficacia a las necesidades de toda la variedad de participantes destinatarios.
- P2 Acuerdos formales.** Las obligaciones de las partes formales de una evaluación (qué debe hacerse, cómo, por parte de quién, cuándo) deberían ser acordadas y redactadas de manera que las partes estén obligadas a adherirse a todas las condiciones de acuerdo o renegociarlo formalmente.
- P3 Derechos de las personas.** Las evaluaciones deberían diseñarse y llevarse a cabo respetando y protegiendo los derechos y el bienestar de las personas.
- P4 Relaciones humanas.** Los evaluadores deberán respetar la dignidad y el valor de las personas en sus interacciones con otras personas relacionadas con la evaluación, de manera que los participantes no estén amenazados ni resulten dañados.
- P5 Evaluación completa y justa.** La evaluación debería ser completa y justa en su examen y registro de los puntos fuertes y débiles del programa evaluado, de modo que se puedan desarrollar los puntos fuertes y atender a las áreas problemáticas.
- P6 Revelación de los resultados.** Las partes formales de una evaluación deberían asegurarse de que la totalidad de los resultados de la evaluación, teniendo en cuenta las limitaciones pertinentes, resulten accesibles a personas afectadas por la evaluación y a cualquier otra con derecho legal reconocido a recibir los resultados.
- P7 Conflicto de interés.** Los conflictos de intereses deberían afrontarse abiertamente y de un modo honesto, de modo que no comprometan los procesos y los resultados de la evaluación.
- P8 Responsabilidad fiscal.** La disposición y el gasto de recursos por parte del evaluador deberían llevarse a cabo mediante procedimientos de rendición de cuentas acertadas y además debe ser prudente y éticamente responsable de manera que los gastos sean apropiados y se responda de ellos.

## ESTÁNDARES DE PRECISIÓN

El objeto, su evolución y su contexto deben estar claramente descritos. Han de garantizar que la evaluación presente y transmita las informaciones técnicas adecuadas referentes al programa objeto de evaluación, resaltando tanto los aspectos positivos como negativos. Los estándares son los siguientes:

- Pr1 Documentación del programa.** El programa evaluado debería ser descrito y documentado con claridad y precisión, de manera que se identifique claramente el programa.
- Pr2 Análisis del contexto.** Debería examinarse con el suficiente detalle el contexto en el que existe el programa, de modo que puedan identificarse las posibles influencias sobre el programa.
- Pr3 Propósitos y procedimientos descritos.** Debería realizarse un seguimiento y una descripción suficientemente detallada de los propósitos y procedimientos de la evaluación, de modo que puedan ser identificados y valorados.
- Pr4 Fuentes de información defendibles.** Deberían describirse con el detalle suficiente las fuentes de información utilizadas en la evaluación de un programa, de modo que pueda valorarse la adecuación de la información.
- Pr5 Información válida.** Los procedimientos de recogida de información deberían ser elegidos o elaborados y posteriormente implementados de modo que aseguren que la interpretación a la que se llegue es válida para el uso pretendido.
- Pr6 Información fiable.** Los procedimientos de recogida de información deberían ser elegidos o elaborados y posteriormente utilizados de modo que aseguren que la información obtenida es suficientemente fiable para el uso pretendido.
- Pr7 Información sistemática.** La información recogida, procesada y de la que se informa en una evaluación debería ser revisada sistemáticamente y todos los errores hallados deberían ser corregidos.
- Pr8 Análisis de la información cuantitativa.** La información cuantitativa de una evaluación debería ser analizada de un modo apropiado y sistemático para que se responda eficazmente a las preguntas de la evaluación.
- Pr9 Análisis de la información cualitativa.** La información cualitativa de una evaluación debería ser analizada de un modo apropiado y sistemático para que se responda eficazmente a las preguntas de la evaluación.
- Pr10 Conclusiones justificadas.** Las conclusiones a las que se llegue en una evaluación deberían ser justificadas explícitamente, de modo que los interesados puedan valorarlas.
- Pr11 Informes imparciales.** Los procedimientos seguidos para la elaboración del informe deben estar protegidos contra la distorsión causada por los sentimientos personales y el sesgo de cualquier parte en la evaluación, de manera que los informes de la evaluación reflejen con justicia los informes de la evaluación.
- Pr12 Metaevaluación.** Debería evaluarse la propia evaluación, formativa y sumativa, en referencia a estos u otros estándares pertinentes, de manera que se guíe adecuadamente su proceso y, además, los interesados puedan examinar detalladamente sus puntos fuertes y débiles tras su conclusión.



## ANEXO II

### Grupo de Evaluación de las Naciones Unidas (UNEG) Estándares de evaluación en el Sistema de las Naciones Unidas. 2005

#### Preámbulo

El Grupo de Evaluación de las Naciones Unidas (UNEG), en su calidad de grupo de evaluadores/as profesionales, emprendió la tarea de definir normas y estándares para contribuir a la profesionalización de la función de evaluación y proveer orientación a las oficinas de evaluación en la elaboración de sus políticas de evaluación u otros aspectos de sus operaciones. Esta iniciativa se acometió, en parte, como respuesta a la Resolución A/RES/59/250<sup>1</sup> de la Asamblea General de diciembre de 2004, que exhortaba a UNEG a profundizar la cooperación en materia de evaluación entre todas las organizaciones del sistema, especialmente en lo que se refiere a la armonización y simplificación de las metodologías, normas, estándares y ciclos de evaluación.

Estos estándares que complementan las Normas de Evaluación para el sistema de Naciones Unidas (NU), se derivan de prácticas idóneas de los miembros de UNEG<sup>2</sup>. Los estándares han sido diseñados para guiar el establecimiento del marco institucional, la gestión de la función de evaluación y la realización y el uso de evaluaciones. Representan, asimismo, un referente para las competencias de los profesionales y la ética de trabajo en materia de evaluación, y la intención es que sean aplicados dentro de cada organización según resulte apropiado. UNEG actualizará, ampliará y expandirá periódicamente la cobertura de estos estándares, al servicio de las organizaciones del sistema de las NU<sup>3</sup>.

<sup>1</sup> Documento A/C.2/59/L.63 del 17 de diciembre de 2004, párrafo 69.

<sup>2</sup> Adicionalmente a las políticas y lineamientos de evaluación existentes en las distintas organizaciones del sistema de las NU, los estándares se derivan también de las siguientes fuentes: los principios de evaluación del Comité de Asistencia al Desarrollo de la Organización para la Cooperación y el Desarrollo Económicos (OCDE/CAD); los estándares nacionales de los países de la OCDE; las políticas de evaluación de las instituciones financieras internacionales; las políticas de evaluación de la Unión Europea; los estándares de las asociaciones de evaluación; la Guía de Evaluación desarrollada por The Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP, por sus siglas en Inglés) para la asistencia humanitaria.

<sup>3</sup> El término organismos u organizaciones de las NU, en adelante se usará para denominar a todas las organizaciones, fondos y programas, así como agencias especializadas, que forman parte del Sistema de las NNUU.

## Marco institucional y gestión de la función de la evaluación.

### Marco institucional

---

Estándar 1.1: Las organizaciones de las NU deben contar con un marco institucional adecuado para la gestión efectiva de su función de evaluación.

---

1.- Un marco institucional integral para la gestión de la función de evaluación y la realización de evaluaciones, es crucial para garantizar un proceso de evaluación efectivo.

2.- Dicho marco institucional debe responder a los siguientes requerimientos:

- Proveer conocimiento y apoyo gerencial e institucional de alto nivel para el rol clave de la función de evaluación cual es, contribuir a la efectividad de la organización.
- Garantizar que la evaluación forme parte de las funciones de gobierno y gerencia de la organización. La evaluación realiza una contribución esencial a la gestión por resultados.
- Promover una cultura que valore la evaluación como una base para el aprendizaje.
- Facilitar un proceso de evaluación independiente e imparcial, garantizando que la función de evaluación se realice de manera independiente de otras funciones gerenciales. El jefe/a de la unidad de evaluación deberá reportar directamente al órgano de gobierno de la organización o al jefe/a de ésta.
- Garantizar recursos financieros y humanos adecuados para la evaluación, que permitan una prestación de servicios eficiente y efectiva y el fortalecimiento de la capacidad de evaluación, por parte de una función de evaluación competente.
- Incentivar las asociaciones y la cooperación en evaluación entre las organizaciones del sistema de las NU así como con otras instituciones relevantes.

---

Estándar 1.2: Las organizaciones de las NU deben elaborar una política de evaluación y actualizarla regularmente, teniendo en consideración las Normas y los Estándares de Evaluación del sistema de las UN.

---

3.- La política de evaluación debe ser aprobada por los órganos de gobierno de las organizaciones y/o el jefe/a de éstas y debe ser coherente con las Normas de Evaluación de UNEG que sean aplicables, así como con los objetivos y estrategias corporativos. La política de evaluación debe incluir:

- una explicación clara del concepto y el rol de la evaluación al interior de la organización;
- una definición clara de los roles y responsabilidades de los profesionales directamente vinculados a la función de evaluación, de la alta gerencia y de los gerentes de programas;
- un énfasis en la necesidad de que los lineamientos de evaluación de la organización, sean adoptados;

- una explicación de la forma en que se priorizan y planifican las evaluaciones;
- una descripción de la forma en que se organizan, gestionan y presupuestan las evaluaciones;
- un énfasis en los requerimientos para el seguimiento de las evaluaciones;
- un enunciado claro en materia de divulgación y difusión.

---

Estándar 1.3: Las organizaciones de las NU deben asegurar que los planes de evaluación de actividades de evaluación sean presentados a sus órganos de gobierno y/o a los jefe/as de organizaciones para su revisión y/o aprobación.

---

4.- Los órganos de gobierno y/o el jefe/a de la organización deben recibir no solamente el plan de evaluación sino también un informe de avance acerca de la ejecución tanto del plan de evaluación como de las recomendaciones que emanan de las evaluaciones.

---

Estándar 1.4: Las organizaciones de las NU deben garantizar la existencia de mecanismos apropiados para el seguimiento de las evaluaciones; y formular una política explícita en materia de divulgación.

---

5.- Deben existir mecanismos apropiados para el seguimiento de las evaluaciones al interior de la organización, de tal suerte que garanticen que las recomendaciones de las evaluaciones sean debidamente utilizadas oportunamente implementadas, y que sus hallazgos sean vinculados a actividades futuras.

6.- Una política de divulgación debe garantizar la difusión transparente de los resultados de las evaluaciones, lo cual incluye poner los informes a disposición general de los órganos de gobierno y del público, salvo en aquellos casos en que se requiera protección y confidencialidad razonables para algunas de las partes interesadas.

### **Gestión de la función de evaluación**

---

Estándar 1.5: El jefe/a de evaluación ejerce un rol de liderazgo en garantizar que la función de evaluación se encuentre plenamente operativa y que el trabajo en materia de evaluación, se lleve a cabo de acuerdo con los más altos estándares profesionales.

---

7.- Dentro del marco institucional global, la gestión de la función de evaluación, encomendada al jefe/a de evaluación, debe garantizar que:

- una política de evaluación sea elaborada y actualizada regularmente; el presupuesto para evaluaciones sea administrado de manera eficiente;
- un plan de actividades, en materia de evaluación, sea elaborado una vez al año o cada dos años como parte del ciclo de planificación y presupuestación de la organización. Dicho plan priorizará aquellas áreas donde haya mayor requerimiento de una evaluación, y especificará los recursos adecuados para la planificación, realización y seguimiento de evaluaciones;
- metodologías de evaluación adecuadas sean adoptadas, desarrolladas y actualizadas en forma frecuente;
- las evaluaciones sean llevadas a cabo de acuerdo a estándares de calidad definidos y

de una manera oportuna, para que sirvan como una herramienta valiosa para las partes interesadas y para los usuarios previstos;

- los informes a la alta dirección sean oportunos y relevantes a sus necesidades, impulsando con ello un proceso informado de toma de decisiones de gestión y de política institucional;
- informes periódicos de avance sobre la ejecución del plan de evaluación y/o la implementación de las recomendaciones que emanan de las evaluaciones ya realizadas, sean compilados y presentados a los órganos de gobierno y/o jefe/as de las organizaciones;
- las lecciones extraídas de las evaluaciones sean destiladas y difundidas según se juzgue conveniente.

---

Estándar 1.6: El jefe/a de evaluación es responsable de asegurar la elaboración de lineamientos de evaluación.

---

8.- Se elaborarán lineamientos de evaluación, que incluyan lo siguiente:

- metodologías de evaluación que reflejen los más altos estándares profesionales;
- procesos de evaluación que aseguren que las evaluaciones se lleven a cabo de una manera objetiva, imparcial, abierta y participativa, con base en evidencia verificada empíricamente que sea válida y confiable, y en que los resultados se pongan a disposición;
- consideraciones éticas, asegurando que las evaluaciones se lleven a cabo con el respeto y la consideración debidos a quienes se está evaluando.

---

Estándar 1.7: El jefe/a de evaluación debe cerciorarse de que la función de evaluación sea dinámica, adaptándose a los nuevos desarrollos y a las necesidades cambiantes tanto dentro como fuera de la organización.

---

9.- En particular, la gestión de la función de evaluación debe incluir:

- generar conciencia y/o construir capacidad en evaluación;
- facilitar y gestionar redes de evaluación; diseñar e implementar metodologías y sistemas de evaluación;
- garantizar el mantenimiento de una memoria institucional de las evaluaciones, mediante mecanismos que sean amigables con el usuario;
- promover la compilación de lecciones de una manera sistemática.

### **Competencias y ética.**

1. Todas las personas involucradas en el diseño, la realización y la gestión de actividades de evaluación deben aspirar a realizar un trabajo de alta calidad, que se rija por estándares profesionales y principios éticos y morales

### **Competencias**

---

Estándar 2.1: Las personas involucradas en el diseño, la realización y la gestión de actividades de evaluación deben poseer competencias medulares en materia de evaluación.

---

2.- Las competencias en materia de evaluación se refieren a las calificaciones, destrezas, experiencia y atributos requeridos de quienes trabajan en la función de evaluación para desempeñar sus funciones de la manera estipulada y para garantizar la credibilidad del proceso.

3.- Las competencias se requieren para todas las personas que participan en el diseño, realización y gestión de actividades en materia de evaluación, la supervisión de evaluadores/as, la realización de actividades de capacitación y de desarrollo de capacidades y en el diseño e implementación de metodologías y sistemas de evaluación.

4.- Algunas destrezas son especialmente útiles para aquellas personas que realizan evaluaciones asumiendo el rol de “evaluadores/as”, mientras que otras son necesarias para las personas que son responsables de coordinar evaluaciones, asumiendo el rol de “gerentes de la evaluación”. El término “evaluadores/as” que se utiliza de aquí en adelante, engloba ambos roles.

5.- Los evaluadores/as deben comunicar a los clientes cualquier conflicto de interés antes de emprender un proyecto de evaluación, así como en cualquier momento en que dicho conflicto ocurriera. Esto incluye conflictos de interés ya sea del evaluador/a o de la parte interesada.

6.- Los evaluadores/as deben demostrar, con precisión, su nivel de destrezas y conocimiento. Del mismo modo, los evaluadores/as deben ejercer dentro de los límites de su formación y competencia profesionales, y declinar la realización de evaluaciones que recaigan, de manera significativa, fuera de dichos límites.

---

Estándar 2.2: Los evaluadores/as deben poseer educación, calificaciones y capacitación relevantes, en materia de evaluación.

---

7.- Los evaluadores/as deben poseer, de preferencia, un grado universitario avanzado o estudios equivalentes en ciencias sociales u otras disciplinas pertinentes, con capacitación especializada en áreas tales como evaluación, gestión de proyectos, estadísticas sociales, investigación y análisis estadístico avanzado.

8.- Los evaluadores/as deben esforzarse permanentemente por mantener y mejorar sus competencias, a fin de ofrecer el más alto nivel posible de desempeño en sus evaluaciones. Este desarrollo profesional permanente podría incluir, seminarios y talleres formales, auto-estudio, evaluaciones de sus propias prácticas y trabajo con otros evaluadores/as a fin de aprender de sus destrezas y experticia.

---

Estándar 2.3: Los evaluadores/as deben contar con experiencia de trabajo profesional relevante en materia de evaluación.

---

9.- Los evaluadores/as deben poseer además experiencia profesional en:

- diseño y gestión de procesos de evaluación, incluyendo aquellos que involucren a múltiples actores;
- diseño y conducción de encuestas;
- investigación científica social;
- planificación, seguimiento y gestión de proyectos, programas y/o políticas.

---

Estándar 2.4: Los evaluadores/as deben tener conocimiento técnico específico de, y estar familiarizados con, la metodología o enfoque que se requerirá para la evaluación específica a ser ejecutada, así como ciertas destrezas gerenciales y personales.

---

10. Especialmente útil será poseer experiencia especializada y/o conocimiento técnico-metodológico, incluyendo algunas destrezas específicas de recolección y análisis de datos, en las siguientes áreas:

- comprensión de enfoques de programación basados en los derechos humanos;
- comprensión de las consideraciones de género;
- comprensión de los principios de la gestión basada en resultados (GBR);
- análisis centrado en el modelo lógico/ marco lógico;
- evaluación en tiempo real, centrada en el uso, conjunta, sumativa y formativa;
- recolección y análisis de datos cuantitativos y cualitativos;
- procedimientos de evaluación rápida;
- enfoques participativos.

11.- El evaluador/a, cuyas responsabilidades incluyan la coordinación o supervisión de evaluaciones, requiere destrezas gerenciales específicas:

gestión del proceso de evaluación;  
planificación, establecimiento de estándares y seguimiento o monitoreo;  
gestión de recursos humanos y financieros;  
liderazgo de equipos de trabajo;  
pensamiento estratégico y global;  
previsión y resolución de problemas.

12. El evaluador/a requiere, asimismo, ciertas destrezas personales que resultan especialmente valiosas para la evaluación:

trabajo en equipo y cooperación;  
capacidad para congregarse a actores diversos;  
comunicación;  
fuerte capacidad de redacción;  
capacidad de análisis;  
destrezas de negociación;  
conocimiento de idiomas, adaptado a la región donde tendrá lugar la evaluación.

## Consideraciones éticas

---

Estándar 2.5: Los evaluadores/as deben ser sensibles a las creencias, usos y costumbres y conducirse con integridad y honestidad en sus relaciones con todas las partes interesadas.

---

13. En línea con la Declaración Universal de los Derechos Humanos de las NU y otras convenciones de derechos humanos, los evaluadores/as deben operar en concordancia con valores internacionales.

14. Los evaluadores/as deben ser conscientes de las diferencias en términos de cultura, costumbres locales, creencias y prácticas religiosas, interacción personal y roles de género, discapacidad, edad y etnicidad, y deben estar atentos a las potenciales implicaciones de dichas diferencias al planificar, llevar a cabo e informar sobre las evaluaciones.

15. Los evaluadores/as tienen que garantizar la honestidad e integridad de todo el proceso de evaluación. Los evaluadores tienen asimismo la responsabilidad primordial de garantizar que las actividades de evaluación sean independientes, imparciales y exactas.

---

Estándar 2.6: Los evaluadores/as deben cerciorarse de que sus relaciones con las personas se caractericen por el respeto.

---

16.- Los evaluadores/as deben evitar ofender la dignidad y el amor propio de las personas con las que entran en contacto en el curso de la evaluación.

17.- Conscientes de que la evaluación a menudo podría afectar negativamente los intereses de algunas partes interesadas, los evaluadores/as deben llevar a cabo la evaluación y comunicar el propósito y los resultados de la misma de una manera claramente respetuosa de la dignidad y del amor propio de las partes interesadas.

---

Estándar 2.7: Los evaluadores/as deben proteger el anonimato y la confidencialidad de informantes particulares.

---

18. Los evaluadores/as deben anunciarse con la máxima anticipación, minimizar las exigencias sobre el tiempo de las personas y respetar el derecho a la privacidad.

19. Los evaluadores/as deben respetar el derecho de las personas a proporcionar información de manera confidencial, y deben garantizar que para aquellas informaciones de carácter sensible, no se rastreen las fuentes. Deben informar, asimismo, a los participantes acerca del alcance y los límites de la confidencialidad.

20. Los evaluadores/as no deben evaluar a las personas y deben mantener un equilibrio entre la evaluación de las funciones de gestión y este principio general.

21. Los evaluadores/as tienen la responsabilidad de mencionar aspectos y hallazgos que pueden no estar directamente vinculados a los términos de referencia. Deben consultar con otras entidades de supervisión competentes en caso de que surgiera alguna duda acerca de la conveniencia de informar sobre dichos aspectos - por ejemplo, evidencia de infracción - y sobre la forma de cómo hacerlo.

---

Estándar 2.8: Los evaluadores/as son responsables de su desempeño y de su(s) producto(s).

---

22.- Los evaluadores/as son responsables de la presentación escrita y/u oral clara, precisa y justa de las limitaciones, hallazgos y recomendaciones del estudio.

23.- Los evaluadores/as deben ser responsables de culminar la evaluación dentro de un plazo razonablemente programado, dejando constancia de cualesquier demora imprevista que se produjera como consecuencia de factores fuera del control del evaluador/a.

## Realización de evaluaciones

### Diseño

---

Estándar 3.1: La evaluación debe ser diseñada de manera tal que garantice información oportuna, válida, confiable y relevante sobre la intervención que está siendo evaluada.

---

1.- La realización de las evaluaciones se rige por la planificación cíclica a varios niveles, la cual se compone de distintas etapas: planificación, diseño, ejecución y seguimiento ex post.

---

Estándar 3.2: Los términos de referencia deben precisar el propósito y describir el proceso y el producto de la evaluación.

---

2.- El diseño de una evaluación debe describirse de la manera más precisa posible en los términos de referencia, que deben incluir los siguientes elementos:

- contexto de la evaluación;
- propósito de la evaluación;
- alcance (descripción de lo que cubre y no cubre la evaluación);
- criterios de evaluación (entre otras cosas, relevancia, eficiencia, efectividad, impacto, sostenibilidad);
- preguntas clave de la evaluación;
- metodología - enfoque para la recolección y el análisis de datos y para la participación de las partes interesadas;
- plan de trabajo, organigrama y presupuesto;
- productos e informes;
- uso de los resultados de la evaluación, lo cual incluye responsabilidades en relación con dicho uso.

---

Estándar 3.3: El propósito y el contexto de la evaluación deben ser claramente descritos, proporcionando una justificación específica del por qué se emprenderá la evaluación en una fecha determinada.

---

3.- El propósito de la evaluación debe ser definido de una manera clara y precisa, teniendo presentes las principales necesidades de información de los usuarios previstos de la evaluación. El propósito refiere a por qué se está efectuando la evaluación, qué la desencadenó y de qué manera será utilizada. El propósito refiere también a la oportunidad de la evaluación en diferentes coyunturas del ciclo de gestión. Esto añade a la claridad de la evaluación y debe ofrecer una orientación general, la cual se desarrollará luego en los objetivos específicos y en el alcance de la evaluación.

---

Estándar 3.4: La intervención que será evaluada debe ser claramente descrita.

---

4.- La intervención que será evaluada, debe ser descrita en términos de lo que aspira a alcanzar, la forma en que los diseñadores pensaron que se abordaría el problema que habían identificado, las modalidades de ejecución, así como cualesquier cambios, intencionales o no, en la ejecución.

5.- Otros elementos incluyen la importancia o los parámetros de la intervención que será evaluada; entre otras cosas, el costo y el peso relativo de la misma en relación con, por ejemplo, las actividades generales de la organización. Como mínimo indispensable, la descripción debe incluir el número de participantes o personas cubiertas por la intervención.

---

Estándar 3.5: Los objetivos específicos de la evaluación deben ser realistas y alcanzables, a la luz de la información que puede ser recolectada en el contexto de la intervención. El alcance de la evaluación debe ser también claramente definido.

---

6.- Los objetivos específicos deben desprenderse del propósito de la evaluación. Deben ser claros y gozar del consenso de todas las partes involucradas.

7.- El alcance determina los límites de la evaluación, adaptando los objetivos específicos y criterios de evaluación a la situación dada. Debe consignar, asimismo y de manera explícita, la cobertura de la evaluación (el plazo, la fase en ejecución, el área geográfica y las dimensiones de la participación de las partes interesadas que está siendo examinada). Los límites de la evaluación también deben ser reconocidos en el alcance.

8.- Las evaluaciones también pueden ser orientadas por preguntas de evaluación. Éstas añaden mayor detalle a los objetivos específicos y contribuyen a definir el alcance.

9.- Los objetivos específicos y el alcance de la evaluación constituyen referentes críticos para determinar la metodología de evaluación y los recursos requeridos.

---

Estándar 3.6: El diseño de la evaluación debe detallar claramente los criterios a partir de los cuales se evaluará la intervención.

---

10. Los criterios de evaluación más comúnmente aplicados son los siguientes: relevancia, eficiencia, efectividad, impacto, valor por dinero, satisfacción del cliente y sostenibilidad. Los criterios para respuestas humanitarias deben incluir asimismo: cobertura, coordinación, coherencia, conectividad y protección. No todos los criterios son aplicables a todas las evaluaciones.

---

Estándar 3.7: Las metodologías de evaluación deben ser lo suficientemente rigurosas como para evaluar la intervención y garantizar una evaluación exhaustiva, justa y libre de sesgos.

---

11.- Las metodologías de evaluación que se utilicen para la recolección y el análisis de datos y para la participación de las partes interesadas, deben ser correspondientes con la intervención que será evaluada, garantizando que la información recolectada sea válida, confiable y suficiente para cumplir con los objetivos específicos, y que la evaluación sea exhaustiva, justa y libre de sesgos.

12.- Los métodos de evaluación dependen de la información que se busque y del tipo de datos que se esté analizando. Los datos deben provenir de una variedad de fuentes a fin de garantizar la exactitud, validez y confiabilidad de los mismos, y que todas las personas afectadas y/o partes interesadas sean tomadas en cuenta. La metodología debe incidir explícitamente en los aspectos de género y de grupos subrepresentados.

13.- Deben reconocerse también, las limitaciones de los métodos de evaluación que fueron seleccionados.

---

Estándar 3.8: Una evaluación debe medir, en lo posible, el costo-efectividad.

---

14.- Recurriendo a una gama de enfoques para el análisis de los costos, desde el elaborado análisis de costo-efectividad y de costo-beneficio, pasando por el análisis de costo-eficiencia, hasta la comparación rápida de costos, una evaluación debe intentar responder, en la medida de lo posible, las siguientes preguntas generales:

- ¿Cómo se comparan los costos reales con otros costos referenciales similares?
- ¿Cuál es la forma más económica o eficiente de obtener los resultados esperados?
- ¿Cuáles son las implicaciones de aumentar o reducir el impacto en términos de costos?
- ¿Cuáles son los costos de replicar la intervención que está siendo evaluada, en un entorno diferente?
- ¿Merece la pena la evaluación de esta intervención en particular? ¿Los beneficios económicos son mayores que los costos?
- ¿De qué forma afectan los costos la sostenibilidad de los resultados?

15.- El análisis de costos en la evaluación se basa en información financiera, pero puede entrañar asimismo calcular “costos económicos” tales como: recursos humanos, trabajo en especies, costos-opportunidad, etc.

16.- El alcance del análisis de costos - es decir, si se hace una comparación de los costos en relación con los impactos, los efectos o los productos - dependerá del propósito de la evaluación y de las preguntas de evaluación planteadas. El análisis de costos debe ser explícito en cuanto a las diferentes perspectivas a partir de las cuales se analizan los costos (donantes, una sola organización, interesados primarios) y las limitaciones - complejidad de la intervención (objetivos programáticos múltiples, socios, sistemas financieros), disponibilidad de datos, así como el tiempo y los recursos invertidos.

17.- El análisis de costos no siempre es factible. Cuando la evaluación no haya incluido un análisis de costos, deberá proporcionarse alguna explicación de las razones de dicha exclusión en la sección sobre los objetivos específicos o en la metodología.

18.- Se espera que los evaluadores/as identifiquen aquellas áreas en las que el uso de los recursos es obviamente ineficiente.

---

Estándar 3.9: El diseño de la evaluación debe incluir, donde sea relevante, consideraciones sobre hasta qué punto el compromiso del sistema de las NU con el enfoque basado en los derechos humanos ha sido incorporado en el diseño de la intervención que será evaluada.

---

19.- Los organismos de las NU se rigen por la Carta de las Naciones Unidas y tienen la responsabilidad y la misión de ayudar a los Estados Miembros a honrar sus obligaciones frente a la realización de los derechos humanos de las personas que viven dentro de su jurisdicción. Los tratados, mecanismos e instrumentos de derechos humanos otorgan a las entidades de las NU un marco de referencia rector y un fundamento legal de principios éticos y morales, y deben guiar el trabajo de evaluación. Debe prestarse atención, además, a los aspectos de género y a los grupos vulnerables y difíciles de acceder.

20.- El diseño de la evaluación podría incluir además, algún proceso de revisión ética del diseño original de la intervención que será evaluada., que incluya lo siguiente:

- un balance de costos y beneficios para los participantes, incluyendo el potencial impacto negativo;
- las consideraciones éticas que subyacen a quién se incluye o excluye de la evaluación y cómo se lleva a cabo este proceso;
- el manejo de la privacidad y la confidencialidad;
- las prácticas para obtener acuerdos informados;
- la retroalimentación a los participantes;
- los mecanismos para determinar y monitorear la conducta y la práctica de los evaluadores/as y recolectores/as de datos.

## Proceso

---

Estándar 3.10: La relación entre el evaluador/a y la institución o instituciones que comisionan la evaluación tiene que caracterizarse, desde el principio, por el respeto mutuo y la confianza.

---

21.- Las responsabilidades de las partes que convienen en realizar una evaluación (especificando qué, cómo, por quién y cuándo deberá realizarse qué) deben establecerse en un convenio escrito con la finalidad de obligar a las partes contratantes a cumplir con todas las condiciones acordadas o, en caso contrario, renegociar el acuerdo. Deberán establecerse convenios - por ejemplo, términos de referencia al menos en las siguientes áreas: financiación, marco cronológico, personas involucradas, informes a ser producidos o publicados, contenido, metodología, y procedimientos. Un convenio de esta naturaleza reduce la probabilidad de que surjan malentendidos entre las partes contratantes y facilita la resolución de los mismos en caso de que llegaran a surgir. Presentar un informe inicial al comienzo de la evaluación, es una manera provechosa de formalizar dicho convenio y de garantizar una interpretación adecuada de los términos de referencia.

22.- Los evaluadores/as deben consultar con la institución o instituciones que comisionan la evaluación acerca de decisiones contractuales tales como: confidencialidad, privacidad, comunicación y propiedad de los hallazgos e informes.

---

Estándar 3.11: Las partes interesadas deben ser consultadas en las etapas de planificación, diseño, realización y seguimiento de las evaluaciones.

---

23.- Es necesario identificar a las partes interesadas y consultar con ellas (temas clave, metodología, plazos, responsabilidades), al planificar la evaluación así como mantenerlas informadas a todo lo largo del proceso de evaluación. El enfoque de evaluación debe considerar oportunidades de aprendizaje y participación (por ejemplo, talleres, grupos de aprendizaje, reuniones para informar sobre procesos concluidos o “debriefings”, participación en visitas de campo), a fin de asegurar que actores clave sean integrados plenamente al proceso de aprendizaje que implica la evaluación.

24.- Ahí donde sea factible, puede crearse un grupo medular de aprendizaje o coordinación, integrado por representantes de las diversas partes interesadas en la evaluación. El rol del mencionado grupo consiste en actuar como caja de resonancia así como facilitar y revisar el trabajo de evaluación. Dicho grupo puede asumir, asimismo, la tarea de facilitar la difusión y aplicación de los resultados y otras actividades de seguimiento.

---

Estándar 3.12: Una revisión de pares o un grupo referencial integrado por expertos externos, puede ser especialmente valioso.

---

25. Dependiendo del alcance y de la complejidad de la evaluación, puede ser útil establecer una revisión de pares o un grupo referencial integrado por expertos en los tópicos técnicos cu-

biertos por la evaluación. Dicho grupo daría una orientación sustantiva al proceso de evaluación (por ejemplo, aportando insumos respecto de los términos de referencia y ejerciendo control de calidad sobre el borrador del informe).

### **Selección del equipo**

---

Estándar 3.13: Las evaluaciones deben ser llevadas a cabo por equipos de evaluación altamente calificados.

---

26.- El número de evaluadores/as de un equipo dependerá del tamaño de la evaluación. Las evaluaciones que incluyan facetas múltiples, tienen que ser realizadas por equipos multidisciplinarios.

27.- Los evaluadores/as deben ser seleccionados sobre la base de sus competencias y por medio de un proceso transparente.

28.- Los evaluadores/as seleccionados deben introducir distintos tipos de experticia y experiencia en el equipo. De ser posible, al menos un miembro del equipo debe contar con experiencia en el sector o en las áreas técnicas que aborda la evaluación, o debe tener un sólido conocimiento de la intervención que será evaluada. De preferencia, al menos un miembro adicional debe ser especialista en evaluación y tener experiencia en el uso de las metodologías de evaluación específicas que se emplearán en esta evaluación. El equipo de evaluación debe poseer, asimismo, un amplio conocimiento y comprensión de los principales temas y problemas de desarrollo económico y social en el país o países donde se desarrolla la evaluación o en países similares de la región. Es posible que se requieran, asimismo, antecedentes o familiaridad con situaciones de emergencia, tanto para la realización del ejercicio en sí como para entender el contexto particular de la evaluación.

---

Estándar 3.14: La composición de los equipos de evaluación debe ser balanceada en términos de género, geográficamente diversa e incluir profesionales de los países o regiones concernidos.

---

29.- Empresas o individuos profesionales calificados, competentes y experimentados de los países concernidos deben participar en la realización de las evaluaciones cada vez que sea posible, a fin de garantizar, entre otras cosas, que el conocimiento y la información a nivel nacional y/o local sean debidamente tomados en cuenta en las evaluaciones y para contribuir a la construcción de capacidad en evaluación en los países en desarrollo. La realización de las evaluaciones también puede ser tercerizada a organizaciones nacionales del sector privado y la sociedad civil. Del mismo modo, deben incentivarse las evaluaciones conjuntas con los gobiernos y otras partes interesadas.

30.- Los miembros del equipo de evaluación deben familiarizarse también con los valores y características culturales y sociales de los receptores y beneficiarios previstos. De este modo, estarán mejor preparados para entender y respetar las costumbres, creencias y prácticas locales a todo lo largo del trabajo de evaluación.

## Ejecución

---

Estándar 3.15: Las evaluaciones deben llevarse a cabo de una manera profesional y ética.

---

31.- Las evaluaciones deben llevarse a cabo de una manera participativa y ética, y debe darse el respeto y la consideración debidos al bienestar de las partes interesadas (derechos humanos, dignidad y justicia). Las evaluaciones deben ser sensibles a los aspectos culturales y de género y respetar los aspectos de confidencialidad, protección de la fuente y dignidad de los entrevistados.

32.- Los procedimientos de evaluación deben llevarse a cabo de manera realista, diplomática, consciente de los costos y costo-efectiva.

33.- Las evaluaciones deben ser precisas, estar sólidamente documentadas y emplear métodos transparentes que generen información válida y confiable. Los miembros del equipo de evaluación deben tener la posibilidad de disociarse de opiniones y recomendaciones particulares. En el informe se dejará constancia de cualesquier diferencias de opinión que no hubieran sido resueltas al interior del equipo.

34.- Las evaluaciones deben llevarse a cabo de una manera integral y balanceada, de manera tal que las diferentes perspectivas sean consideradas y analizadas. Los hallazgos clave deben ser sustentados por medio de la triangulación. Cualquier conflicto de interés debe ser abordado de una manera abierta y honesta, de modo tal que no socave el resultado de la evaluación.

35.- Los evaluadores/as deben discutir, de una manera apropiada para el contexto, aquellos valores, supuestos, teorías, métodos, resultados y análisis que pudieran afectar significativamente la interpretación de los hallazgos de la evaluación. Esto se aplica a todos los aspectos de la evaluación, desde su concepción inicial hasta el uso eventual de los hallazgos.

36.- Los derechos y el bienestar de las personas no deben verse afectados negativamente durante la planificación y ejecución de una evaluación. Esto tiene que ser comunicado a todas las personas que participan en una evaluación, y sus consecuencias previsibles para la evaluación deben ser discutidas.

## Informes

---

Estándar 3.16: El informe final de evaluación debe contar con una estructura lógica, contener hallazgos basados en evidencia, conclusiones, lecciones y recomendaciones y estar libre de información irrelevante para el análisis general. El informe debe ser presentado en un formato que permita que la información sea comprensible y accesible.

---

31.- El lector o lectora del informe de evaluación debe poder entender:

- el propósito de la evaluación;
- qué, exactamente, se evaluó;
- cómo se diseñó y llevó a cabo la evaluación;

- qué evidencias se encontró;
- qué conclusiones se extrajeron;
- qué recomendaciones se hicieron;
- qué lecciones se derivaron.

38.- Si los evaluadores/as identificaran fraude, malos manejos, abuso de poder y violaciones de derechos deberán derivar el asunto, confidencialmente, a las autoridades competentes de las NU para la investigación de estos hechos. Las evaluaciones no deben sustituir, o ser usadas para, la toma de decisiones respecto de asuntos individuales relacionados con los recursos humanos.

39.- Los evaluadores/as deben permitir a todos los actores pertinentes, acceso a información procedente de la evaluación que sea apropiada, y, de ser posible, deben difundir activamente dicha información entre las partes interesadas. Las comunicaciones a una parte interesada deben incluir, siempre, los resultados importantes que pudieran afectar sus intereses. En todos los casos, los evaluadores/as deben esforzarse por presentar los resultados con la mayor claridad y sencillez posible, de manera tal que los clientes y otras partes interesadas puedan entender fácilmente el proceso y los resultados de la evaluación.

## **Seguimiento**

---

Estándar 3.17: La evaluación requiere de una respuesta explícita por parte de las autoridades de gobierno y gerencia afectadas por sus recomendaciones.

---

40.- Al igual que en el caso de las Normas, dicha respuesta puede adoptar la forma de una respuesta de gerencia, un plan de acción y/o un acuerdo especificando claramente las responsabilidades y rendiciones de cuentas.

41.- Debe hacerse un seguimiento sistemático de la implementación de las recomendaciones de la evaluación que hubieran sido aceptadas por la gerencia y/o los órganos de gobierno.

42.- Debe realizarse, asimismo, un informe periódico acerca del estado de ejecución de las recomendaciones de la evaluación. Dicho informe debe ser presentado a los órganos de gobierno y/o el jefe/a de la organización.

## **Informes de evaluación**

---

Estándar 4.1: La portada y las páginas iniciales del informe deben ofrecer información básica clave.

---

1.- La siguiente información debe ser fácilmente accesible en las primeras páginas del informe:

- nombre de la intervención (actividad, programa, política, etc.) que está siendo evaluada;
- fecha;
- índice, incluyendo anexos;

- nombre y organización(es) de los evaluadores/as;
  - nombre y dirección de la(s) organización(es) que comisionaron la
  - evaluación.
- 

Estándar 4.2: El informe de evaluación debe contener un resumen ejecutivo.

---

2.- El resumen ejecutivo debe ofrecer una sinopsis de los aspectos sustantivos del informe de evaluación. Para facilitar la lectura, el resumen ejecutivo debe ser corto, de dos o tres páginas, y auto-contenido. El nivel de la información debe permitir al lector/a no letrado, una comprensión clara de los hallazgos, recomendaciones y lecciones de la evaluación.

3.- El resumen ejecutivo debe constar de:

- una breve descripción de la intervención que está siendo evaluada;
  - el contexto, la situación actual y una descripción de la intervención en relación con otros asuntos vinculados;
  - el propósito de la evaluación;
  - los objetivos específicos de la evaluación;
  - el público objetivo del informe;
  - una breve descripción de la metodología, incluyendo la razón fundamental para haberla seleccionado, las fuentes de datos utilizadas, los métodos de recolección y análisis de datos empleados; y las principales restricciones relacionadas con la metodología;
  - los hallazgos y conclusiones más importantes;
  - las principales recomendaciones.
- 

Estándar 4.3: La intervención que está siendo evaluada debe ser claramente descrita, incluyendo el modelo lógico y/o la cadena de resultados y el impacto esperados, la estrategia de ejecución y los supuestos centrales.

---

4.- El informe de evaluación debe describir claramente el propósito de la intervención que está siendo evaluada y de qué manera pensaron los diseñadores que ésta respondería al problema identificado. Otros elementos importantes incluyen: importancia, alcance y escala de la intervención que se está evaluando; descripción de los receptores a los beneficiarios previstos y de las partes interesadas; y cifras presupuestarias.

5.- La descripción de la intervención que se está siendo evaluada debe ser lo más sucinta posible, al tiempo de asegurar que contendrá toda la información pertinente. Si se estima necesario incluir detalles adicionales puede adjuntarse, en un anexo, una descripción que incluya el modelo lógico.

---

Estándar 4.4: El rol y las contribuciones de las organizaciones de las NU y de otras partes interesadas de la intervención que está siendo evaluada deben describirse claramente.

---

6.- El informe debe describir quiénes participan en la intervención que está siendo evaluada, sus roles y sus contribuciones, incluyendo recursos financieros, aportes en especies, asistencia técnica, participación, tiempo del personal, capacitación, liderazgo, incidencia política, cabildeo y cualesquier contribuciones de los interesados primarios, por ejemplo, las comunidades. Se debe intentar aclarar qué socios contribuyeron a qué efecto en particular.

7.- Los usuarios querrán comparar esta información con los actores que participaron en la evaluación para evaluar la forma en que se incluyeron distintos puntos de vista.

---

Estándar 4.5: El propósito y el contexto de la evaluación deben consignarse en el informe.

---

8.- El propósito debe indicar por qué se está efectuando la evaluación, cómo será utilizada y qué decisiones se tomarán una vez culminada. El contexto debe ser descrito a fin de permitir la comprensión del entorno en el cual se desarrolló la evaluación.

---

Estándar 4.6: El informe de evaluación debe ofrecer una explicación de los criterios de evaluación que fueron utilizados por los evaluadores/as.

---

9.- No todos los criterios son aplicables a todas las situaciones. Las razones fundamentales para no emplear un criterio en particular deben figurar en el informe, así como cualesquier restricciones a la aplicación de criterios de evaluación. Se debe describir también los estándares o parámetros de desempeño que se utilizaron en la evaluación.

10.- Es importante exponer de forma transparente los fundamentos que subyacen a los juicios de valor.

---

Estándar 4.7: El informe de evaluación debe ofrecer una explicación clara de los objetivos específicos, así como del alcance de la evaluación.

---

11.- Los objetivos específicos originales de la evaluación deben ser consignados, así como cualesquier cambios efectuados al diseño de la evaluación.

12.- Debe describirse el alcance de la evaluación, explicitando la cobertura de la misma. Igualmente, deben mencionarse los límites de la evaluación.

13.- Deben consignarse las preguntas originales de la evaluación, así como aquéllas que hubieran sido añadidas en el transcurso de ésta. Se trata de referentes básicos contra los cuales se contrastará el contenido del informe.

14.- Los objetivos específicos y el alcance de la evaluación son, asimismo, referentes fundamentales para establecer si la metodología seleccionada y los recursos asignados fueron los adecuados.

---

Estándar 4.8: El informe de evaluación debe indicar hasta qué punto fueron incorporadas consideraciones relevantes de género y de derechos humanos ahí donde fueran aplicables.

---

- 15.- El informe de evaluación debe incluir una descripción de lo siguiente, entre otros aspectos:
- si las consideraciones de género fueron abordadas como un tema transversal en la programación, y si la intervención que está siendo evaluada dedicó atención suficiente a la promoción de la igualdad de género y la sensibilidad de género;
  - si la evaluación que está siendo evaluada prestó atención a los efectos sobre los grupos marginados, vulnerables y difíciles de acceder;
  - si la intervención que está siendo evaluada fue informada por los tratados e instrumentos referentes a los derechos humanos;
  - hasta qué punto, la intervención que está siendo evaluada, identificó las reivindicaciones y obligaciones de derechos humanos relevantes;
  - cómo se identificaron los vacíos en la capacidad de los titulares de derechos para reivindicar sus derechos, y de los responsables de hacer cumplir los derechos para cumplir con sus obligaciones, incluyendo un análisis de género y de los grupos marginados y vulnerables: y cómo el diseño y la ejecución de la intervención que está siendo evaluada, respondieron a dichos vacíos;
  - cómo la intervención que está siendo evaluada, monitoreó y enfocó los resultados desde la perspectiva de este marco de derechos.
- 

Estándar 4.9: La metodología de evaluación aplicada debe describirse de una manera transparente, incluyendo cualesquier limitaciones inherentes a ella.

---

16.- El informe de evaluación debe contener una descripción exhaustiva, más no excesiva, de los aspectos esenciales de la metodología, que permita al usuario o usuarios de la evaluación extraer sus propias conclusiones respecto de la calidad de la información. La descripción de la metodología debe incluir:

- fuentes de información;
- descripción de los métodos de recolección y análisis de datos (que incluya el nivel de precisión requerido para los métodos cuantitativos, así como las escalas de valores o codificaciones empleadas en el análisis cualitativo);
- descripción del muestreo (área y población por representar, razón fundamental de la selección, mecánica de selección, números seleccionados de entre los potenciales sujetos, restricciones de la muestra);
- indicadores y parámetros referenciales, ahí donde sea relevante (indicadores previos, estadísticas nacionales, etc.);
- el equipo de evaluación, incluyendo la participación de miembros individuales del equipo;
- el plan de evaluación;
- restricciones clave.

Los anexos deben incluir:

- mayores detalles sobre cualquiera de los puntos anteriores;
- instrumentos de recolección de datos (encuestas, listas de verificación, etc.);
- sistema para garantizar la calidad de los datos mediante el monitoreo y la supervisión de la recolección de los datos;
- una discusión más detallada de las restricciones según sea necesario.

---

Estándar 4.10: La evaluación debe proporcionar una descripción detallada de la participación de las partes interesadas.

---

15. Debe incluirse una descripción del nivel de participación de las partes interesadas en la evaluación, incluyendo la razón fundamental para haber seleccionado dicho nivel de participación en particular. Si bien no todas las evaluaciones pueden ser participativas en la misma medida, es importante otorgar la debida importancia a la participación de las partes interesadas, dado que dicha participación es crecientemente reconocida como un factor crucial en el uso de las conclusiones, recomendaciones y lecciones. El enfoque de programación basado en los derechos humanos añade énfasis a la participación de los interesados primarios. En muchos casos, esto supone claramente la participación de la población y de las comunidades. Adicionalmente, la inclusión de determinados grupos de interesados puede ser necesaria para garantizar una evaluación completa y justa.

---

Estándar 4.11: El informe de evaluación debe incluir un análisis de la medida en que el diseño de la evaluación incluyó salvaguardas éticas donde fuera apropiado.

---

16. El informe debe contener una descripción satisfactoria de las consideraciones éticas, incluyendo la base lógica detrás del diseño de la evaluación y los mecanismos para proteger a los participantes donde fuera apropiado. Esto incluye la protección de la confidencialidad, dignidad, derechos y bienestar de los seres humanos, incluyendo a la niñez, y el respeto por los valores de las comunidades beneficiarias.

---

Estándar 4.12: En la presentación de los hallazgos es necesario, en la medida de lo posible, medir los insumos, productos y efectos / impactos (o dar una explicación apropiada de por qué no se hace).

---

19.- Los hallazgos relativos a los insumos para el cumplimiento de las actividades o los logros a nivel de proceso, deben ser claramente diferenciados de los productos, los efectos y del impacto.

20.- Los efectos e impactos deben incluir cualesquier resultados no intencionales, ya sean beneficiosos o perjudiciales. En adición, deberá incluirse cualesquier efectos multiplicadores o a nivel micro, de la intervención que está siendo evaluada. En la medida de lo posible, cada uno de éstos efectos deberá ser medido ya sea cuantitativa o cualitativamente. Al utilizar dichas mediciones, deberá hacerse referencia a parámetros referenciales.

21.- El informe debe hacer una distinción lógica en los hallazgos, mostrando la progresión de la ejecución hacia los resultados, incluyendo una medición y un análisis apropiado de la cadena de resultados, o una explicación de por qué este análisis no habría sido incluido.

22.- Los datos no tienen que consignarse en su totalidad; únicamente merecen ser enunciados aquellos datos que respalden un hallazgo. La versión completa de los datos puede adjuntarse en un anexo. En adición, los informes no deben segregar los hallazgos por fuente de información.

23.- Los hallazgos deben cubrir todos los objetivos específicos de la evaluación y hacer uso de los datos recolectados.

---

Estándar 4.13: El análisis debe incluir una discusión apropiada de las contribuciones relativas, de las partes interesadas a los resultados.

---

24. Los resultados atribuidos a la intervención que está siendo evaluada, deben ser vinculados a las contribuciones de las diferentes partes interesadas. Debe haber un sentido de proporcionalidad entre las contribuciones relativas de cada parte interesada y los resultados observados. Éste es un elemento integral de la rendición de cuentas frente a los socios, donantes e interesados primarios.

25. Si dicho análisis no es incluido en el informe, deberá indicarse claramente la razón por la que no se hizo. Por ejemplo, si una evaluación se lleva a cabo en una etapa temprana del ciclo de gestión, posiblemente no se hallen resultados o los vínculos de estos con una parte interesada.

---

Estándar 4.14: Deben ser identificadas, en la medida de lo posible, las razones de los logros y las dificultades de la intervención que está siendo evaluada, especialmente los factores de restricción y habilitación.

---

26.- El informe de evaluación debe ir más allá de una simple descripción de la implementación y de los efectos e incluir un análisis, basado en los hallazgos, de las causas subyacentes, las restricciones, las fortalezas que hay que capitalizar y las oportunidades. Deben identificarse y analizarse, en la medida de lo posible, los factores externos que contribuyen a los logros y dificultades, entre ellos la situación social, política o ambiental.

27.- Una explicación del contexto contribuye a la utilidad y precisión de la evaluación. La comprensión de los factores externos que contribuyeron al éxito o fracaso de la intervención que está siendo evaluada, ayuda a determinar la forma en que dichos factores afectarán el futuro de la intervención o la posibilidad de replicarla en otro lugar.

---

Estándar 4.15: Las conclusiones tienen que ser sustentadas por hallazgos que sean consistentes con los datos recolectados y por la metodología; asimismo, deben aportar elementos de juicio conducentes a la identificación y/o solución de problemas o temas importantes.

---

28.- Las conclusiones deben agregar valor a los hallazgos. La lógica detrás de las conclusiones y la correlación de éstos con los hallazgos deben ser evidentes.

29.- Las conclusiones deben concentrarse en temas trascendentes para la intervención que está siendo evaluada. Deben estar determinadas por los objetivos específicos y por las preguntas clave de la evaluación. Las conclusiones simples, que ya se conocen o resultan evidentes, no son útiles y deben evitarse.

30.- Las conclusiones referidas a la atribución de los resultados, que en la mayoría de los casos son de carácter tentativo, exigen detallar lo que se sabe y lo que puede ser asumido como plausible, a fin de hacer más transparente la transición lógica de los hallazgos a las conclusiones y fortalecer con ello la credibilidad de las conclusiones.

---

Estándar 4.16: Las recomendaciones deben estar firmemente basadas en evidencia y análisis y ser relevantes y realistas, con prioridades para la acción claramente enunciadas.

---

31.- Por razones de precisión y credibilidad, las recomendaciones deben ser implicaciones lógicas de los hallazgos y de las conclusiones. Asimismo, las recomendaciones deben ser relevantes para la intervención que está siendo evaluada, y pertinentes a los términos de referencia y los objetivos específicos de la evaluación; además, deben estar formuladas de una manera clara y concisa. Adicionalmente y en la medida de lo posible, las recomendaciones deben ser priorizadas.

32.- Las recomendaciones deben especificar, en la medida de lo posible, las responsabilidades y el marco cronológico para su implementación.

---

Estándar 4.17: Las lecciones, al ser presentadas, deben incluir generalizaciones que trasciendan la intervención inmediata que está siendo evaluada poniendo énfasis en la relevancia más amplia que podrían tener.

---

33.- No todas las evaluaciones generan lecciones. Las lecciones deben ser extraídas solamente si representan contribuciones al conocimiento general. Deben estar sólidamente respaldadas por los hallazgos y conclusiones de la evaluación. Pueden fortalecer o complementar lecciones comúnmente aceptadas, pero no deben ser una mera repetición del conocimiento común.

34.- Un buen informe de evaluación identifica correctamente las lecciones que se desprenden lógicamente de los resultados, presenta un análisis de cómo éstas pueden ser aplicadas a diferentes contextos y/o sectores y toma en cuenta limitaciones evidentes tales como generalizaciones a partir de observaciones puntuales.

---

Estándar 4.18: Los anexos deben ser completos y relevantes.

---

35.- Información complementaria adicional que debe ser contenida en anexos incluye:  
Relación de personas entrevistadas (siempre que lo permita la confidencialidad) y sitios

visitados; instrumentos de recolección de datos (copias de cuestionarios, encuestas, etc.); los términos de referencia originales de la evaluación; relación de abreviaciones.

36.- Los anexos incrementan la probabilidad de uso y la credibilidad del informe.