

J. F. Lukas y K. Santiago

Evaluación educativa

Segunda edición

Alianza Editorial

**J. F. Lukas
K. Santiago**

Evaluación educativa

Segunda edición

Alianza Editorial

Reservados todos los derechos. El contenido de esta obra está protegido por la Ley, que establece penas de prisión y/o multas, además de las correspondientes indemnizaciones por daños y perjuicios, para quienes reprodujeren, plagiaren, distribuyeren o comunicaren públicamente, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la preceptiva autorización.

Edición electrónica, 2014
www.alianzaeditorial.es

© José Francisco Lukas Mujika y Karlos Santiago Etxebarría, 2004
© Alianza Editorial, S. A. Madrid, 2014
Juan Ignacio Luca de Tena, 15. 28027 Madrid
ISBN: 978-84-206-8883-1
Edición en versión digital 2014

Índice

Introducción	9
1. Naturaleza de la investigación y evaluación en educación	13
1. La investigación científica.....	14
2. La investigación educativa.....	17
3. Paradigmas en la investigación educativa	
3.1. Qué es paradigma	21
3.2. El paradigma positivista	23
3.3. El paradigma interpretativo	26
3.4. El paradigma crítico	28
3.5. Características de los métodos cuantitativos y cualitativos	30
4. Algunas cuestiones acerca de los paradigmas.....	33
5. Los paradigmas de la evaluación	39
6. Evaluación e investigación	43
7. Evaluación y complementariedad metodológica	
7.1. Naturaleza de la complementariedad metodológica	50
7.2. Estrategias de complementariedad	51
7.3. Ejemplos ilustrativos	54
2. Historia de la evaluación educativa.....	63
1. Antecedentes remotos (2000 a. C.-1900).....	64
2. Etapa de la eficiencia y de los tests (1901-1929).....	67
3. Etapa de Tyler (1930-1945)	68
4. Etapa de la inocencia (1946-1957)	
5. Etapa de la expansión (1958-1972).....	71

6. Etapa de la profesionalización (1973-actualidad).....	75
7. Historia de la evaluación educativa en España	
7.1. La Ley General de Educación de 1970.....	77
7.2. La reforma educativa de 1990.....	78
7.3. La Ley de Calidad de 2002.....	81
7.4. La Ley Orgánica de Educación de 2006.....	82
7.4.1. La evaluación diagnóstica	84
3. Concepto, componentes y fases de la evaluación	
1. Concepto	87
2. Componentes	
2.1. La finalidad	93
2.2. El objeto	
2.3. La información.....	95
2.4. El agente.....	96
2.5. El momento	97
2.6. La valoración.....	99
2.6.1. Los criterios.....	100
2.6.2. Los indicadores	101
2.6.3. Los estándares	103
2.7. El informe.....	104
3. Fases del proceso evaluativo	107
4. La evaluación educativa y otras disciplinas	
4.1. Medición y evaluación.....	115
4.2. <i>Assessment</i> y evaluación.....	117
4.3. Diagnóstico y evaluación.....	120
4. Evaluación de programas	
1. Concepto	127
2. Modelos de evaluación de programas	
2.1. Concepto de modelo	130
2.2. Criterios para la clasificación de los modelos	
2.2.1. La finalidad de la evaluación	131
2.2.2. Los criterios de valoración.....	132
2.2.3. El rol de los evaluadores	133
2.3. Algunos modelos representativos de la evaluación de programas	
2.3.1. Modelos de evaluación de costos-resultados	134
2.3.2. El modelo de consecución de metas de Tyler	135
2.3.3. El modelo CIPP (facilitador de toma de decisiones).....	137
2.3.4. El modelo de evaluación respondente de Stake.....	138
2.3.5. El modelo de evaluación iluminativa de Parlett y Hamilton	141
3. El diseño de evaluación	
3.1. Concepto de diseño	
3.2. Criterios para la clasificación de los diseños	143
4. Elección del modelo y del diseño de evaluación	147
5. Ámbitos de la evaluación educativa	
1. Evaluación del alumnado.....	149

1.1. Modelos de evaluación del aprendizaje de los alumnos	151
1.2. El proceso de evaluación del alumnado	
1.3. La evaluación de las competencias	154
2. Evaluación de centros	155
2.1. Enfoques en la evaluación de centros educativos	157
2.2. Criterios que determinan los modelos evaluativos	159
2.3. Estrategias utilizadas para la evaluación de centros	161
2.4. El proceso de la evaluación de centros	163
3. Evaluación del profesorado	
3.1. Concepto y problemática	164
3.2. Propósitos de la evaluación del profesorado	166
3.3. El proceso de evaluación del profesorado	167
4. Evaluación del sistema educativo	
4.1. Concepto y antecedentes	172
4.2. Funciones de la evaluación del sistema educativo	176
4.3. Evaluación del sistema educativo y calidad educativa	177
4.4. Componentes de la evaluación del sistema educativo	
4.4.1. El objeto de evaluación	
4.4.2. Los objetivos de la evaluación	178
4.4.3. Los agentes de la evaluación	179
4.4.4. Las audiencias de la evaluación	
4.4.5. El proceso de la evaluación	180
4.5. La utilización de indicadores en la evaluación del sistema educativo	182
6. Diseños de investigación aplicados a la evaluación	185
1. Diseños descriptivos	186
1.1. Estudios tipo encuesta	188
1.2. Estudios observacionales	189
1.3. Estudios analíticos	
1.4. Estudios de desarrollo	191
2. Diseños correlacionales	193
3. Diseños experimentales	196
4. Método etnográfico	202
5. Estudio de casos	206
6. Investigación-acción	209
6.1. Naturaleza de la investigación-acción	210
6.2. Enfoques de la investigación-acción	212
6.3. El proceso de la investigación-acción	214
7. Procedimientos para la recogida de la información	217
1. La observación	
1.1. Concepto	218
1.2. Modalidades de observación	219
1.3. Planificación de la observación	221
1.4. Ventajas y limitaciones de la observación	222
2. La entrevista	
2.1. Concepto	223
2.2. Clasificación	224

2.3. Procedimiento	224
2.4. Ventajas y limitaciones	
3. El cuestionario	
3.1. Concepto	226
3.2. Clasificación	
3.3. Procedimiento	227
3.4. Ventajas y limitaciones	228
4. Las pruebas de aprendizaje	
4.1. Las pruebas objetivas.....	229
4.2. Las pruebas de ensayo	
4.3. Las pruebas de ejecución	230
5. Los tests estandarizados	
5.1. Concepto	232
5.2. Clasificación de los tests	234
5.3. Procedimiento	235
6. Otros procedimientos para la recogida de información	
6.1. Las escalas de actitudes	237
6.2. La sociometría	
6.3. El grupo de discusión	238
6.4. El autoinforme.....	239
6.5. Las historias de vida	
6.6. Los documentos	
6.7. La técnica Delphi.....	240
7. La calidad de la información.....	241
8. Recursos informáticos para la recogida de la información	
8.1. La edición, corrección y proceso de información de pruebas de evaluación	
8.2. El diseño y presentación de ítems y tests.....	244
8.3. Sistemas de evaluación dinámica y adaptada a las personas	
8.4. Evaluación y diagnóstico de personas con necesidades especiales	
8.5. Recogida de datos	245
8. Procedimientos para el análisis de la información	253
1. Análisis cuantitativos	255
2. Análisis cualitativos	257
3. Recursos informáticos para el análisis de la información	
3.1. Aplicaciones para el análisis cuantitativo de datos.....	262
3.2. Aplicaciones para el análisis cualitativo de datos.....	264
3.3. Aplicaciones de ayuda para la construcción de instrumentos.....	266
9. Actividades prácticas.....	269
Anexo 1. Definiciones de evaluación educativa	289
Anexo 2. Recursos informáticos para la evaluación educativa.....	303
Bibliografía.....	317

Introducción

En estos últimos años se habla de un cambio cuantitativo y cualitativo en el campo de la evaluación educativa. Cuantitativo por la proliferación de trabajos relacionados con el tema de la evaluación que se han y se están realizando desde finales de los años setenta y los ochenta del siglo pasado hasta nuestros días (sin menospreciar los realizados anteriormente). Pero mucho más importante es el cambio cualitativo relacionado con los distintos objetos evaluados y las formas de llevar a cabo esas evaluaciones (Bordás Alsina, 1995; Jiménez Jiménez, 1999a). Si en un principio el único objeto evaluado/medido era el resultado académico del alumnado, se ha pasado a evaluar otra serie de objetos, desde los centros educativos, pasando por los sistemas educativos hasta diferentes programas por citar algunos ejemplos representativos. El segundo cambio cualitativo alude a la metodología utilizada; si en el origen la metodología empleada era exclusivamente cuantitativa, se ha pasado a utilizar también la metodología cualitativa. De esta forma, la evaluación ha alcanzado en pocos años un lugar destacado en el panorama educativo (Tiana Ferrer, 1995). El propio House (1992) señala que cuando a principios de la década de 1970 comenzó su trabajo como evaluador, recopiló todos los artículos existentes sobre evaluación y le entraban en una carpeta. Hoy en día, necesitaría algún inmueble.

Hay autores y autoras¹ que van más allá (De la Orden, 2000) y se atreven a definir esta etapa histórica como «la era de la evaluación» (p. 381), o señalan que la evaluación está de moda (Santos Guerra, 1993) debido al lugar tan amplio y destacado que ha adquirido:

Hoy, en un ambiente de cambio educativo generalizado, efecto tanto de las grandes reformas en marcha como de la innovación permanente, a nivel institucional y de aula, la evaluación educativa cobra también su máximo protagonismo y presenta características similares a las apuntadas para el campo en su conjunto, como demuestra el énfasis puesto en la evaluación del aprendizaje de los alumnos, la evaluación de los sistemas educativos en la perspectiva regional, nacional e internacional, los avances en la evaluación de centros escolares y profesores y especialmente la evaluación de programas educativos específicos y de formación continua de profesionales.

(De la Orden, 2000, p. 382).

Sin embargo, este hecho tiene el riesgo de creer que la evaluación es un fin en sí mismo y no debemos olvidar que la evaluación no es más que un instrumento al servicio de la comunidad educativa.

En esta nueva edición revisada, hemos pretendido mantener la esencia del libro, pero hemos introducido una serie de aspectos que pretenden actualizarlo.

El presente libro está estructurado en tres partes bien diferenciadas que hacen referencia a los aspectos fundamentales de la evaluación educativa. La primera parte, que se refiere a las bases conceptuales y epistemológicas, a la historia de la evaluación y a los componentes de la evaluación, está dividida en tres capítulos. El primer capítulo, «Naturaleza de la investigación y de la evaluación en educación», recoge una reflexión en torno al carácter científico de la investigación educativa y los diferentes enfoques, tradiciones o paradigmas que en la actualidad coexisten en su seno. También se introduce la evaluación como investigación evaluativa dentro de la investigación educativa y se ponen de manifiesto las similitudes y las diferencias que existen entre ambas. Para esta nueva edición se ha añadido un apartado relativo a la complementariedad metodológica.

El segundo capítulo, «Historia de la evaluación educativa», hace un recorrido histórico de la evaluación educativa desde sus orígenes hasta los planteamientos actuales. La reseña histórica a nivel mundial se completa con un acercamiento al camino seguido por la evaluación en España. Además, se han añadido los cambios habidos en los últimos años en la legislación educativa con respecto a la evaluación, mencionándose entre otros la evaluación diagnóstica de los centros escolares aprobada en la LOE.

El tercer capítulo supone una aproximación conceptual a la evaluación educativa. En él, se define la evaluación educativa, se presentan sus componentes y se plantean las fases del proceso evaluativo. Para terminar con este capítulo, se señalan las similitudes y las diferencias entre la medición, el *assessment*, el diagnóstico y la evaluación.

La segunda parte recoge los capítulos cuatro y cinco en los que se presentan los «Ámbitos de la evaluación educativa». Se han distinguido cinco ámbitos fundamentales que son: la evaluación de programas, del alumnado, de centros, del profesorado y del sistema educativo.

La tercera parte, que abarca los capítulos seis, siete y ocho, se centra en los aspectos metodológicos de la evaluación educativa. En el capítulo sexto se presentan los «Diseños de investigación aplicados a la evaluación»; en el séptimo, los «Procedimientos de recogida de información», y en el octavo, los «Procedimientos para el análisis de la información». En estos dos últimos capítulos se presentan también los recursos informáticos de ayuda disponibles en las distintas tareas evaluativas.

En todos los capítulos anteriores, a su conclusión, se han introducido una serie de preguntas y ejercicios con el objeto de repasar los conceptos tratados.

En el capítulo noveno se presentan una serie de «Actividades prácticas» para cada capítulo, que pueden ser realizadas con los alumnos con vistas a completar la formación teórica recibida. Como se puede comprobar, el objetivo de las actividades prácticas planteadas es lograr un conocimiento y una comprensión más profundos sobre los temas presentados. Pero además, como el conocimiento y la comprensión de una técnica no aseguran la capacidad de aplicación, en algunos casos se plantea la utilización de esas técnicas en situaciones concretas para obtener una formación integral.

Para terminar, además de las referencias bibliográficas, se han incluido dos anexos. En el primero se recogen más de 100 «Definiciones del concepto de evaluación educativa» señaladas por diferentes autores. En el segundo anexo, que se han actualizado desde la edición anterior, se presentan los «Recursos informáticos para la evaluación educativa», en el que se ofrece información acerca de bibliotecas, bases de datos documentales, organizaciones, asociaciones y centros de investigación y evaluación tanto nacionales como internacionales y recursos sobre *assessment*, evaluación y medición.

Nota

1. Al hablar de alumnos y alumnas, profesores y profesoras, autores y autoras, etc., hemos intentado utilizar un término neutro (alumnado, profesorado, etc.). No obstante en algunas ocasiones hemos empleado sólo el masculino en aras a facilitar la lectura, pero sin olvidar que éste recoge también el género femenino.

1. Naturaleza de la investigación y evaluación en educación

Al abordar el tema de la investigación y la evaluación en educación es inevitable remitirse a una serie de conceptos que de una forma u otra pueden crear dudas más o menos importantes, más o menos trascendentales.

Tomando como ejemplo el término «investigación» pueden venirnos a la mente cuestiones tales como: ¿cuáles son las características diferenciadoras de la investigación científica con respecto a otras formas de conocimiento? Ello nos remite a la concepción que tengamos de ciencia, de conocimiento científico y por extensión a la idea propia de lo que es conocimiento.

El término «educativa» puede suscitar una reflexión acerca de si la educación o en su caso la pedagogía es una ciencia o no. Indudablemente dependerá de lo que entendamos por ciencia.

Y si juntamos los dos conceptos, podemos preguntarnos por el tipo de conocimiento que se genera a partir de la «investigación educativa».

Entramos en los terrenos de la Filosofía de la Ciencia y la Epistemología. Si nos estuviéramos moviendo en el terreno de las Ciencias Naturales quizás esto no nos llevaría más de 3 o 4 páginas. Parece que hay mucha unanimidad en la comunidad científica acerca de las cuestiones anteriormente planteadas. En el campo de las Ciencias Sociales y Humanas, y más concretamente en el de las Ciencias de la Educación, esto no es así. Las perspectivas acerca de las finalidades y tareas de investigación y evaluación son muy variadas, con lo cual el desarrollo de estas cuestiones exige una presentación más elaborada.

1. La investigación científica

Etimológicamente, el término «ciencia» proviene del latín *scire* que significa conocer. A lo largo de la historia del mundo, el ser humano ha estado preocupado por conocer e interpretar los fenómenos que acontecían a su alrededor. Para ello, ha utilizado diferentes criterios e instrumentos que nos llevan a diferentes tipos de conocimiento. El cuadro 1.1 (Aliaga, 2000) recoge una clasificación de estos métodos de conocimiento y sus fuentes.

Cuadro 1.1. Métodos de conocimiento y sus fuentes

Tipo de conocimiento	Fuentes de información
Sentido común	Experiencia cotidiana
Religioso-místico	Revelación y fe
Artístico	Intuición o introspección
Filosófico	Pensamiento
Científico	Método científico

Cada uno de estos tipos de conocimiento tiene sus ventajas y limitaciones y aunque no sean comparables entre ellos (no podemos decir que unos sean superiores a otros), en algunos casos sí puede ocurrir que sean complementarios e incluso coincidentes.

A pesar de que la concepción de ciencia no es independiente del momento histórico en que ésta se plantea y de ser un concepto polisémico, en las siguientes líneas hemos tratado de esbozar los ejes fundamentales de lo que se podría entender por ciencia.

Revisando la bibliografía encontramos diferentes definiciones de lo que es ciencia. Es en la época de Galileo, Copérnico, Newton, etc., cuando se produce un cambio cualitativo con respecto a la concepción de lo «científico» y su diferencia con respecto a lo que no lo es, al empezar la ciencia a desvincularse de la metafísica y limitarse a las llamadas ciencias positivas (Física, Astronomía, Biología, etc.).

Desde Galileo se ha mantenido la idea de que «ciencia» es el conocimiento objetivo y sistemático sobre una materia. El Diccionario de la Real Academia define ciencia (1992) como: «Conocimiento cierto de las cosas por sus principios y causas». Según esta definición son los principios de causalidad y certeza los que darán al conocimiento la categoría de científico frente al conocimiento vulgar.

Por **certeza**, concepto que puede ser considerado como sinónimo de verdad u objetividad y que se opone al de opinión o creencia, se entiende la consecuencia del contraste con la realidad a partir de una teoría de partida (una hipótesis fundamentada teóricamente) y por medio de una metodología adecuada. De alguna manera, sería la tan ansiada «objetividad absoluta» por parte de la ciencia

a lo largo de la historia. El otro elemento incluido en la anterior definición de ciencia es el relativo a la **causalidad**. Según este principio todo lo que ocurre en la realidad se puede ordenar de acuerdo a una serie de causas y efectos.

El valor de estos dos principios de objetividad y causalidad ha sido puesto en cuestión en el sentido de que la objetividad absoluta, incluso en las Ciencias Naturales, es imposible y el establecimiento de relaciones causa-efecto en las Ciencias Sociales es realmente problemático.

Además del principio o supuesto de causalidad, que también se conoce como determinismo, y el de certeza, que es similar al de empirismo, Cohen y Manion (1990) identifican otros dos supuestos o características que están implícitas en el trabajo científico: el principio de **economía** y el de **generalidad**. Por principio de economía entienden los autores la idea básica de que los fenómenos deben explicarse del modo más económico posible, mientras que en el caso de la generalidad, el científico que está preocupado por hallar explicaciones a los fenómenos, trata de generalizar sus hallazgos a situaciones más amplias que aquellas en las que se han dado esos hallazgos.

Algunas definiciones de ciencia son las siguientes:

- Forner y Latorre (1996, p. 33) y Lukas, *et al.* (2000a, p. 256) señalan que ciencia es un «conjunto organizado de conocimientos y de las actividades destinadas a alcanzarlos».
- Popkewitz (1988, p. 62), por su parte, señala que la ciencia es «un conjunto de paradigmas o constelaciones de valores, métodos, procedimientos y compromisos que dan forma a la investigación».

Una de las características que va a diferenciar el llamado conocimiento vulgar del denominado científico es, precisamente, la forma en que va a ser obtenido ese conocimiento, metódicamente, a través de la aplicación de un **método**. Dice Bericat que «no puede existir investigación sin método» (1998, p. 101). Dendaluce (1988) afirma que la mayoría de los autores suelen tener en cuenta los tres componentes de contenido, método y producto a la hora de definir lo que es ciencia. Él mismo define la ciencia como un «modo de conocimiento riguroso y metódico con el fin de descubrir leyes en el objeto de estudio y expresar los conocimientos adquiridos sistemáticamente» (1988b, p. 18).

Además de la de tener un método, el conocimiento científico tiene un conjunto de características que lo diferencian de otros tipos de conocimiento como el vulgar (Latorre, Arnal y Del Rincón, 1996). Algunas de estas características son las siguientes:

- *Objetivo*, en el sentido de que es un intento de que el conocimiento obtenido sea una descripción y/o explicación de la realidad tal como es y no tal como nos gustaría que fuese.
- *Racional*. El conocimiento científico es racional, ya que la ciencia utiliza la razón como vía esencial para llegar a sus resultados.

- *Sistemático*. Es un conocimiento ordenado, interrelacionado e integrado en un sistema.
- *Contrastable*. El conocimiento científico es sometido a prueba a través del contraste con la realidad, lo cual permitirá una mayor fiabilidad del conocimiento, a través de la comprobación por distintas personas y en circunstancias variadas.
- *Fáctico*. Su fuente de información y de respuesta son los hechos o fenómenos externos al investigador.
- *Analítico*. Por medio del análisis se rompe la unidad y la globalidad de los fenómenos humanos al seccionar la realidad para acceder a ella.
- *Comunicable*. Es expresado de forma comprensible y con un lenguaje preciso aceptado por la comunidad científica.
- *Metódico*. Es fruto de una metodología rigurosa.

El concepto de ciencia está íntimamente ligado al de **método científico** e incluso hay autores que atribuyen estas mismas características del conocimiento científico al denominado método científico.

El método científico es un proceso sistemático que engloba una serie de pasos interrelacionados con el objetivo de llegar al conocimiento científico.

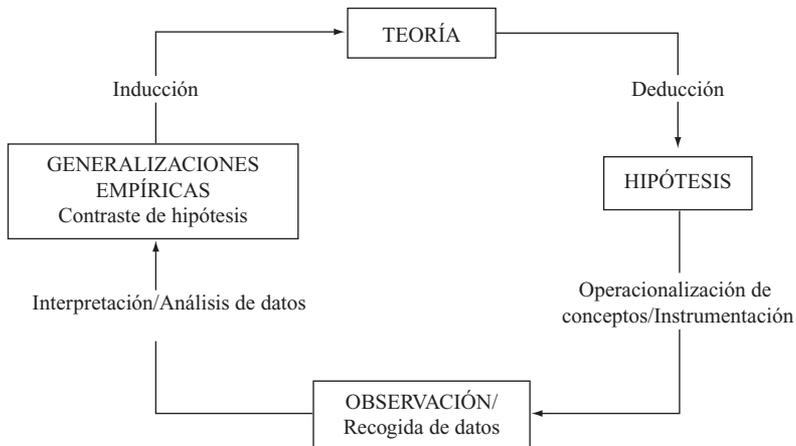
Una forma de entender lo que es el método científico es analizar qué es lo que se hace al aplicarlo. En este sentido podemos encontrar propuestas que se diferencian en el mayor o menor desarrollo de las distintas fases.

Hernández Pina *et al.* (1995) han sintetizado muchas de las propuestas realizadas en distintos manuales de metodología y presentan un modelo configurado en cuatro grandes etapas. Éstas recogen todas las posibilidades de los distintos métodos que podríamos encontrar, sin limitarse solamente al denominado método hipotético-deductivo experimental. Las etapas planteadas son las siguientes:

1. *Preliminares de la investigación*. Plantea actividades tales como identificación de un área de estudio, búsqueda de la información relevante, selección de un tema de estudio y formulación del problema de investigación. El objetivo de esta fase es evaluar una situación de un área de estudio teniendo en cuenta los recursos disponibles y las fuentes de información.
2. *Planificación teórica*. Es el proyecto de investigación propiamente dicho, que podría englobar aspectos como: formulación de objetivos y/o hipótesis, selección o construcción de instrumentos de recogida de datos, selección de participantes y elección del método.
3. *Trabajo práctico*. Es llevar a la práctica lo diseñado o planificado en la fase anterior. Es lo que podríamos llamar el trabajo de campo.
4. *Análisis y resultados*. Es la fase de análisis de los datos e interpretación de los resultados y formulación de las conclusiones.

Uno de los esquemas más conocidos y utilizados del proceso de investigación científica es el que se presenta en la figura 1.1.

Figura 1.1. El proceso de conocimiento científico.
(Tomado de Cea, 1996, p. 66)



2. La investigación educativa

Como se ha podido deducir del punto anterior, otro concepto totalmente relacionado con el de método científico es el de investigación científica. Incluso muchas veces son utilizados como sinónimos.

La investigación científica responde a una actividad rigurosa, sistemática y controlada que tiene como último objetivo la elaboración de teorías. Estos planteamientos responden a las ya clásicas funciones de la ciencia de describir, explicar, prever y controlar la realidad y que están relacionados con el desarrollo de la investigación en las Ciencias Naturales.

Sin embargo, el mundo de las ciencias relacionadas con los fenómenos que ocurren en la naturaleza es diferente del mundo de las ciencias que estudian los fenómenos en los que el protagonismo es del ser humano en cuanto que es integrante de un grupo social. En estos dos mundos los supuestos de partida acerca del carácter de esos fenómenos son diferentes. Esas diferencias van a reflejarse también en las formas de investigar esos fenómenos en cada uno de los casos.

En primer lugar, hay una serie de características específicas de los fenómenos sociales y humanos que dificultan la aplicación de los métodos utilizados en las Ciencias Naturales a la investigación en las Ciencias Sociales y en la educación. Entre ellos podemos destacar la multivariabilidad de los fenómenos sociales, el papel que juegan los valores en general y la multidisciplinariedad.

La educación no es una ciencia tan desarrollada como las Ciencias Naturales. Sin embargo, no se puede dudar de que puede ser estudiada de una manera científica. Pensamos que la aplicación del concepto de investigación

científica a las Ciencias Sociales debe llevarse a cabo de una forma más amplia, superando la concepción estricta que se le da desde las Ciencias Naturales y teniendo más en cuenta el cumplimiento de ciertas premisas como las de ser una actividad sistemática, controlada, empírica y autocontroladora.

Son múltiples los factores que de una manera u otra influyen en los fenómenos y situaciones educativas, por lo que los ideales del método científico sólo pueden ser aproximados. Según Cardona Moltó:

La investigación en Educación se realiza bajo unas condiciones muy particulares. Los rígidos controles de las situaciones de laboratorio no son posibles en Educación. Por añadidura, la medición en investigación educativa es, a menudo, indirecta e imprecisa, y la verificación de los hallazgos por replicación puede resultar difícil. Por ello, en Educación se suele tomar el método científico como un *estándar* al que se somete la investigación para valorar si es adecuada, pero cabe esperar que algunos estudios no se adhieran rígidamente a las características del método científico.

(2002, pp. 32-33)

El estudio de los fenómenos educativos desde una perspectiva científica es relativamente reciente. Podemos situarlo a principios del siglo xx. Esto es importante para comprender que los actuales debates y problemas metodológicos por los que pasa la investigación educativa pueden ser debidos en cierta medida precisamente a ello, a la todavía juventud de éstas con relación a otras disciplinas como las Ciencias de la Naturaleza en general o la propia Psicología.

Ante la posible científicidad de la Educación, se pueden plantear dos posturas básicas. La primera postura sería la de aquellos que no aceptan que la Educación deba considerarse ciencia. Uno de los argumentos de esta postura es la de no considerar las Ciencias Sociales como ciencias. Pensamos que a fin de cuentas lo que subyace en las diferentes posturas es la concepción que se tiene de lo que es ciencia. La negación del carácter científico de las Ciencias Sociales y por extensión de las de la Educación responde a una concepción de lo que es ciencia que se podría definir como estricta y propia de las Ciencias Naturales. Esa concepción cerrada hace que solamente disciplinas bien definidas y limitadas como la Física o la Química, tengan cabida dentro del calificativo de «ciencia».

La segunda postura es la de quienes partiendo de una concepción más abierta de lo que es ciencia aceptan el carácter científico de la Educación. Dendaluze (1988) planteó una definición de ciencia no basada exclusivamente en el empleo del método hipotético-deductivo experimental, sino como una forma de conocimiento riguroso y metódico que tiene como objetivo descubrir leyes acerca del objeto de estudio y expresar esos conocimientos de una forma sistemática. Si consideramos que la Educación tiene objeto propio diferenciado de otras ciencias o disciplinas, rigor, método o métodos, leyes y teoría sistemáticas podemos concluir que la Educación es una ciencia. Dentro

de ésta, otros autores consideran que la Educación no es una ciencia en el sentido de aquéllas sobre las que discuten la mayoría de los filósofos de la ciencia, sino que es una mezcla de conocimiento teórico y acción práctica, al igual que otras disciplinas sociales.

Tendremos que pensar que, ya que actualmente los planteamientos acerca de lo que es ciencia son cada vez más abiertos, la cientificidad de la Educación cada vez será menos cuestionada.

Todo lo dicho anteriormente acerca de la cientificidad o no de la Educación no puede ser desligado de la cuestión de la interdisciplinariedad. Los problemas educativos y su solución debemos concebirlos en contextos amplios que desbordan el ámbito del aula o del centro escolar. El abordaje de los problemas educativos alcanza también a la Sociología, la Psicología, la Antropología, la Economía, etc. La investigación educativa no puede estar al margen de esta situación y tendrá que tener en cuenta las aportaciones de disciplinas como la Psicología, la Sociología, la Filosofía, etc. Habrá que concluir, pues, que la investigación en Educación es en gran medida interdisciplinar.

Por último, quisiéramos hacer un pequeño apunte acerca de la cuestión de la investigación «básica *versus* aplicada». La pregunta clave en este sentido es ¿cuál es el objeto de la Educación? Nuevamente nos encontramos con una dicotomía entre quienes piensan que el objetivo es la búsqueda de conocimientos científicos acerca de la Educación (investigación básica) y quienes afirman que el objetivo es la búsqueda de soluciones prácticas a partir de la comprensión de los fenómenos que ocurren en los contextos educativos (investigación aplicada). Estas dos posturas, en ocasiones enfrentadas, no son sino el reflejo de dos concepciones de la investigación social y educativa, que tiene como consecuencia la utilización de métodos de investigación diferentes.

El enfrentamiento entre ambas posturas carece de sentido en la Educación, cuyo carácter aplicado es una de sus peculiaridades. La Educación no tiene sentido si no está orientada a la búsqueda de conocimientos que sean aplicables a la práctica educativa.

No es de descartar que se acepte que la investigación básica sirva para solucionar problemas educativos y que la investigación aplicada contribuya al enriquecimiento del cuerpo de conocimientos, para lo cual indudablemente se deberán cumplir los requisitos exigibles a la investigación de calidad. A fin de cuentas, serán las diferentes preguntas de investigación las que demanden diferentes respuestas y, en consecuencia, diferentes tipos de investigación. Pensamos que es la única forma de avanzar, pues mantener este enfrentamiento supone mantener la división entre los prácticos y los teóricos de la Educación, división que por lo visto hasta ahora no ha contribuido significativamente a la solución de muchos de los problemas de la Educación.

¿Qué es, por tanto, la investigación educativa?

La investigación en Educación, como en el resto de las Ciencias Sociales, ha llegado a ser una actividad importante y necesaria en el campo educativo, dando lugar a una disciplina académica propia. De hecho, la investigación

educativa como disciplina surge de la aplicación de nociones como conocimiento científico, ciencia, método científico e investigación científica al ámbito de la Educación. La investigación educativa hoy en día es considerada en un sentido amplio como el estudio y análisis de la Educación, tratando de las cuestiones y problemas relativos a la naturaleza, epistemología, metodología, fines y objetivos en el marco de la búsqueda progresiva de conocimiento en el ámbito educativo (Arnal, Del Rincón y Latorre, 1992).

El concepto de investigación educativa ha ido cambiando a lo largo de la historia en función de los diversos enfoques y modos de entender que se han ido adoptando, tanto acerca de la Educación como de la investigación.

La concepción de la investigación educativa como investigación científica aplicada a la Educación y ceñida a las normas del método científico en su sentido estricto es la que se extrae de las definiciones dadas por autores identificados con la denominada corriente positivista. Veamos algunos ejemplos de estas definiciones o aproximaciones al concepto de investigación educativa:

- «Proceso formal, sistemático e intensivo en el cual se aplica el método científico de análisis» (Best, 1972, p. 6).
- «Investigación sistemática, controlada, empírica y crítica, de proposiciones hipotéticas acerca de las presuntas relaciones entre fenómenos naturales» (Kerlinger, 1987, p. 7).
- «Aplicación del método científico al estudio de un problema [...]. Cuando el método científico se aplica al estudio de problemas pedagógicos el resultado es la investigación educacional» (Ary *et al.*, 1987, pp. 20-21).
- «Una ciencia paradigmática por cuanto propone modelos para analizar, descubrir, explicar, contrastar y predecir los resultados a fin de introducir mejoras en la Educación» (García Hoz, 1975, p. 96).

Con la aparición y el desarrollo de nuevos diferentes enfoques de la investigación surgieron nuevas conceptualizaciones de la investigación educativa.

Para la perspectiva interpretativa, el elemento fundamental de la investigación educativa es el de la **interpretación y comprensión** de los fenómenos educativos a partir de los significados de las personas intervinientes en esos fenómenos. El objetivo no es tanto aportar explicaciones causales como interpretar y comprender esos fenómenos educativos.

En la corriente crítica es el componente ideológico el que marca la diferencia con las otras dos perspectivas. La investigación educativa es concebida como una reflexión permanente acerca de la práctica educativa, lo cual traerá consigo la implicación en el proceso de investigación de las personas participantes en esa práctica educativa. Desde este punto de vista, el conocimiento se construye en la práctica y desde la práctica. Además de buscar la comprensión de los problemas educativos la investigación se centra en la búsqueda de soluciones.

A la hora de conceptualizar la investigación educativa hay dos elementos que de alguna forma están presentes. Esos elementos son la construcción de conocimiento teórico y la solución de problemas prácticos. La mayor o menor importancia que se dé a cada uno de estos elementos condiciona la concepción que se tenga de la investigación educativa. Son las ya mencionadas polémicas entre investigación básica y aplicada, investigación orientada a conclusiones y a la toma de decisiones, investigación teórica y práctica.

3. Paradigmas en la investigación educativa

Desde el punto de vista de la metodología, habría que decir que tanto la investigación educativa como la evaluación educativa engloban diferentes formas de acometer sus tareas: estudios experimentales, etnografía educativa, evaluación de programas, investigación en la acción, etc. Algunas cuestiones como la de qué tiene de peculiar cada tipo de investigación o la de qué tienen de similar diferentes métodos de investigación son más fácilmente discernibles si reflexionamos acerca de los principios teóricos que las inspiran. Aquí es interesante hacer mención de los paradigmas en la investigación educativa en la medida en que nos ayudarán a dar luz al complejo carácter de las concepciones que subyacen dentro de la investigación educativa, ya que el paradigma nos da el marco en el que se hace una investigación y evaluación concreta.

3.1. Qué es paradigma

Cuando hablamos de «paradigma» es inevitable hacer referencia a Kuhn, que lo define como la expresión del modo que en un determinado momento tiene una comunidad científica de enfocar los problemas. Un paradigma es un conjunto de asunciones interrelacionadas acerca del mundo social y proporciona un marco filosófico y conceptual para el estudio organizado del mundo social (1982).

Paradigma es un marco científico específico con unas reglas metodológicas, presupuestos ontológicos, generalidades teóricas que son aceptadas por una comunidad de científicos y a partir de los cuales realizan sus actividades e investigaciones. Hace referencia, por tanto, a un conjunto de prácticas científicas o de investigación que en un espacio de tiempo son aceptadas de una forma generalizada y que proporcionan un marco de trabajo, unos problemas y unas soluciones. Además del carácter de no universalidad, los paradigmas se adscriben a momentos históricos concretos. Otra característica a destacar es la de que un paradigma no es un posicionamiento personal, sino algo compartido por un grupo o comunidad de científicos.

En el campo de las Ciencias Sociales y de la Educación no se ha hecho una aplicación rigurosa del término paradigma, sino flexible. Así, se ha partido de la idea de referir la esencia de paradigma a un conjunto de «creencias y actitudes, como la visión del mundo compartida por un grupo de científicos que implica, específicamente, una metodología determinada» (Alvira, 1982, p. 34) o como una «fuente de métodos, problemas y normas de resolución aceptados por una comunidad de científicos» (Pérez Serrano, 1994, p. 17). Shulman habla de un segundo sentido de paradigma, un sentido más débil, que él prefiere usar y afirma: «Los científicos sociales realizan sus actividades de investigación dentro del marco de una escuela de pensamiento que define los objetivos, puntos de partida, métodos y concepciones interpretativas adecuados para sus investigaciones» (1989, pp. 13-14). Partiendo de esta concepción de paradigma tiene sentido pleno hablar de que es posible abordar el estudio de la Educación desde diferentes paradigmas.

Podríamos decir que un paradigma recoge las creencias siguientes (Nirenberg, Brawerman y Ruiz, 2000):

- *Ontológicas*. ¿Qué tipo de sujeto es el ser humano?, ¿cuál es la naturaleza de la realidad?, ¿existe en los hechos o en el pensamiento?, etc. Es decir, se refiere a la naturaleza de los fenómenos sociales y a su grado de estructuración planteando la controversia de si la realidad social es algo externo al sujeto o es algo creado desde su punto de vista particular (Mateo, 2000a).
- *Epistemológicas*. ¿Cuál es la relación entre el investigador y la persona estudiada?, ¿cuál es la relación entre el investigador y el objeto estudiado? Es decir, se refiere a la forma en la que se adquiere el conocimiento que puede ser tomando una postura externa suponiendo la objetividad del conocimiento o una postura de experiencia compartida con las personas implicadas al suponer que el conocimiento es subjetivo.
- *Metodológicas*. ¿Cómo aprehendemos el mundo y avanzamos en su conocimiento? Es decir, se plantean los problemas relativos a los métodos a utilizar.

Esta concepción de paradigma como marco general de referencia ha sido de gran utilidad para entender los diversos métodos de investigación en profundidad y ha permitido, a través del debate surgido en relación a los distintos paradigmas, que se den notorios avances en el terreno epistemológico.

Otra cuestión es la de cuántos paradigmas existen o, dicho de otra forma, de cuántos paradigmas podemos hablar. En este sentido no ha habido y no hay total unanimidad entre los distintos autores que han reflexionado y escrito acerca de estos temas. Hay quien habla de dos paradigmas, hay quien de tres y en algún caso se habla de un cuarto paradigma e incluso de un paradigma único.

El planteamiento más extendido considera la existencia de dos o tres paradigmas. Desde nuestro punto de vista, hablar solamente de dos paradigmas, positivista y antipositivista por ejemplo, supone un cierto reduccionismo, ya que nos llevaría a un debate basado solamente en los aspectos metodológicos y/o técnicos, con lo cual dejamos a un lado la cuestión ideológica olvidando que muchas investigaciones cualitativas no comparten de ninguna de las maneras algunos de los principios ideológicos del paradigma crítico.

Por lo tanto, y tomando como base la postura de la mayoría de los autores de aceptar la existencia de tres paradigmas diferenciados, pasaremos a exponer las características diferenciales de cada uno de ellos.

3.2. El paradigma positivista

Este paradigma es también conocido con los nombres de *cuantitativo*, *científico*, *empírico-analítico*, *racionalista*, *tecnológico*, etc. Es el paradigma dominante en algunas comunidades científicas y es el modelo que ha dominado en la investigación tanto en Ciencias Sociales y de la Educación desde prácticamente sus inicios hasta nuestros días, siendo en muchos momentos de la investigación educativa el único paradigma. Es en la década 1960 cuando se puede decir que comenzaron a hacer su aparición, de una forma relevante, nuevos planteamientos.

Este paradigma se sustenta en la aplicación de los métodos de las Ciencias Físico-Naturales a la investigación educativa, ya que se considera que son esos métodos los únicos que posibilitan la consecución del conocimiento científico. Ya durante la segunda mitad del siglo XIX, distintos autores defendieron que muchos de los problemas de la Educación sólo podrían resolverse adoptando los métodos experimentales de las Ciencias Naturales.

El positivismo es una escuela filosófica cuyos orígenes se pueden situar en los siglos XVIII y XIX, siendo Comte (1798-1857) quien introdujo el término de filosofía positivista. Uno de los postulados básicos de esta filosofía es el de que sólo puede admitirse como conocimiento válido aquél establecido a partir de la experiencia.

El positivismo parte de unos supuestos acerca de la concepción del mundo y el modo de conocerlo. Entre esos supuestos podemos destacar los siguientes (Latorre, Arnal y Del Rincón, 1996, p. 40).

- a) El mundo natural tiene existencia propia independientemente de quien lo estudia.
- b) Está gobernado por leyes que permiten explicar, predecir y controlar los fenómenos del mundo natural y pueden ser descubiertas y descritas de manera objetiva y libre de valor por los investigadores con métodos adecuados.

- c) El conocimiento que se obtiene se considera objetivo y factual, se basa en la experiencia y es válido para todos los tiempos y lugares, con independencia del propósito de quien lo escribe.
- d) Utiliza la vía hipotético-deductiva como lógica metodológica válida para todas las ciencias.
- e) Defiende la existencia de cierto grado de uniformidad y orden en la naturaleza.

El objetivo general es aumentar el conocimiento. Se busca conocimiento sistemático, comprobable, medible y repetible, de forma que lo que no sea comprobable es rechazado.

Al aplicarse este punto de vista positivista a las Ciencias Sociales se asumen una serie de postulados que básicamente hacen referencia a concebir que, en esencia, los fenómenos sociales tienen las mismas características que los naturales. Así se plantea el supuesto de que los objetivos (explicación, predicción y control), los conceptos, los métodos (método científico) utilizados en las Ciencias Naturales así como las normas lógicas que se siguen en las explicaciones de las Ciencias Naturales pueden utilizarse en las Ciencias Sociales.

En el terreno de la Educación, la aplicación de estos planteamientos lleva a una interpretación de la realidad educativa basada en una serie de principios, algunos de los cuales pasaremos a enumerar.

1. Para el positivismo, **la realidad es algo único** que puede llegar a conocerse de forma objetiva sólo a través del método científico. El mundo educativo es considerado como un sistema de relaciones entre variables que pueden estudiarse independientemente las unas de las otras como partes de ese sistema. A través del estudio de las relaciones entre esos conjuntos de variables se pretende llegar a explicar, controlar y predecir la realidad educativa.
2. A través de la investigación empírico-analítica pueden construirse las teorías. Estas **teorías** resultantes son consideradas **universales** en la medida en que no se circunscriben ni a contextos ni a momentos concretos. Estas teorías nos proporcionarán descripciones y explicaciones de distintas situaciones educativas.
3. Otra característica del punto de vista positivista es la cuestión de la **neutralidad**, en cuanto que la investigación basada en la aplicación del método científico lleva consigo una actitud neutral a través de la rigurosidad de los datos obtenidos y evitando los sesgos ocasionados por preferencias personales o subjetivas.

La investigación de orientación positivista ha realizado importantes aportaciones en cuanto al conocimiento científico sobre la Educación, a pesar de que algunos de esos conocimientos no son inmediatamente aplicables a la realidad escolar.

Desde el punto de vista de la metodología, el positivismo ha aportado una metodología científica muy rigurosa que va desde las técnicas descriptivas de encuesta y observación hasta los diseños experimentales más sofisticados. Este desarrollo ha ido totalmente unido al desarrollo de técnicas de análisis de datos que tampoco hubieran sido posibles sin la ayuda de los procedimientos informáticos. Todo esto ha contribuido a hacer de la metodología positivista una importante fuente de conocimientos. Se configura como la única perspectiva válida para la investigación en relación con determinados propósitos y/o situaciones. Esta situación, sin embargo, ha ahondado también en la separación entre teóricos y prácticos, en el sentido de que la investigación educativa ha quedado totalmente en manos de una elite. Como consecuencia, la función de los prácticos ha quedado reducida a llevar a la práctica las teorías desarrolladas o encontradas por los investigadores.

El modelo positivista ha sido ampliamente criticado y en la actualidad está cuestionado al aplicarse a las Ciencias Sociales en general y a la Educación en particular. El excesivo rigor metodológico que impone el método científico muchas veces es un obstáculo para su aplicación en el estudio de los fenómenos educativos. En esta misma línea Cohen y Manion afirman:

Donde el positivismo tiene menos éxito, sin embargo, es en su aplicación al estudio del comportamiento humano, donde la inmensa complejidad de la naturaleza humana y la cualidad fugaz e intangible de los fenómenos sociales contrasta fuertemente con el orden y la regularidad del mundo natural. Este punto es hoy en día más aparente en los contextos de aula y escuela donde los problemas de enseñar, aprender e interacción humana presentan al investigador positivista un enorme reto.

(1990, p. 37)

Hoy en día nadie defiende el positivismo que se adhiere radicalmente a la objetividad y neutralidad de la ciencia. Más que de positivismo, se puede hablar de pospositivismo. Los pospositivistas mantienen la preferencia por la metodología cuantitativa, pero con planteamientos más flexibles en los que se acepta la variedad metodológica y la diversidad de fuentes y métodos como instrumentos de validación científica.

Entre otras, una de las críticas más severas ha ido en la línea de responsabilizarle de la falta de resultados positivos de cara a la solución de la mayoría de los problemas que tiene planteados la Educación. Aunque no podemos negar una alta dosis de razón en esas críticas, no es menos cierto que la corriente positivista ha tenido grandes repercusiones, tanto a nivel metodológico como a nivel educativo. No podemos olvidar que las aportaciones iniciales de la investigación educativa se dieron bajo el amparo de esta perspectiva.

Éstas y otras limitaciones dieron lugar a un movimiento antipositivista que intentará ofrecer alternativas a la investigación educativa diferentes a las del paradigma positivista.

3.3. El paradigma interpretativo

Este paradigma es también conocido como paradigma *humanista, naturalista, etnográfico, cualitativo, fenomenológico, hermenéutico, etc.*

La primera diferencia con el paradigma positivista estriba en su procedencia, que es muy heterogénea. De hecho, este paradigma engloba un conjunto de corrientes humanístico-interpretativas cuyo interés se centra en comprender los significados de las acciones humanas para las distintas personas implicadas en la acción social.

Las cinco corrientes o escuelas de pensamiento que dominan en las tradiciones de la investigación cualitativa son las siguientes: psicología ecológica, etnografía holística, antropología cognitiva, etnografía de la comunicación e interaccionismo simbólico.

Por **psicología ecológica** se entiende aquella tradición de la metodología cualitativa que pretende describir de forma detallada y objetiva las conductas naturales a fin de descubrir las leyes que las rigen.

El centro de la **etnografía holística** es el interés por la exploración, descripción y análisis de los patrones culturales de un grupo social desde el punto de vista de sus integrantes.

Los **antropólogos cognitivos** pretenden estudiar el sistema semántico de un grupo como vía para conocer su organización cultural.

La **etnografía de la comunicación** estudia los procedimientos de comunicación e interacción tanto dentro de un grupo como entre distintos grupos.

Por último, el **interaccionismo simbólico** centra su interés en la interacción entre individuo y grupo, dando una importancia especial a los significados de esas interacciones.

Como puede verse, si el positivismo tiene sus raíces en las Ciencias Naturales, el paradigma interpretativo las tiene en las aportaciones desde la Antropología y la Sociología y alguna fundamentación psicológica. De todas las maneras, todos estos enfoques o variantes centran sus esfuerzos en la comprensión de los significados de las acciones humanas.

Algunos de los postulados que se asumen desde este paradigma son los siguientes:

- La ciencia no es algo abstracto y aislado del mundo, sino que depende del contexto social en el que se desenvuelve.
- La conducta humana es un fenómeno más complejo que los fenómenos de las Ciencias Naturales por lo que no es posible entenderla utilizando los mismos medios utilizados para conocer el mundo natural.
- Las teorías científicas son relativas puesto que se fundamentan en valores sociales que cambian con el tiempo por lo que los conceptos de universalidad, objetividad y certeza de los conocimientos científicos quedan cuestionados.

- Para el paradigma interpretativo, la realidad es algo múltiple (no único) que solamente puede ser estudiada de forma holística. La investigación de esas realidades múltiples indudablemente tenderá a ser divergente, con lo que los objetivos de predicción y control serán muy inviables.

La finalidad de la investigación educativa bajo este prisma es la comprensión de los fenómenos educativos a través del análisis profundo de las percepciones e interpretaciones de los sujetos intervinientes en las diversas situaciones. La obsesión por lo observable y lo mensurable es desechada y se busca la objetividad en el ámbito del significado intersubjetivo (Martínez Mediano, 1996).

No se buscan leyes generales sino conclusiones válidas para los contextos estudiados. Lo que pretende la investigación cualitativa no es llegar a abstracciones universales, sino a universalidades específicas y concretas que sean en principio válidas para los casos estudiados, aunque no se descarten las posibilidades de generalizar a otras situaciones. Frente a la comprobación se da preferencia a la comprensión.

La influencia continua de los valores es aceptada y asumida como parte indisoluble de la investigación en sus distintas fases.

Las teorías generadas no tienen un carácter universal y explicativo, sino orientativo y comprensivo. Las reglas y las leyes científicas están ligadas a situaciones concretas. Goetz y Lecompte (1988) destacan el carácter inductivo y generativo de las teorías, en el sentido de que se desarrollan a partir de los propios datos y no a través del contraste empírico de hipótesis previamente formuladas.

Como otra característica diferencial de la perspectiva positivista habría que señalar la continua interacción entre el investigador y el objeto de investigación, interacción que es más evidente en los procesos de recogida de datos utilizando estrategias que no posibilitan la independencia total, como por ejemplo en el caso de la observación participante o ciertas variedades de entrevista cualitativa.

La aportación más importante en el ámbito de la Educación por parte del enfoque interpretativo ha sido la de iniciar el camino de la ruptura entre el muro que existía entre investigador y educador; entre investigación educativa y práctica docente; entre los teóricos y los prácticos. La razón de esto puede estar en que, técnicamente la metodología cualitativa es más accesible y practicable por los docentes. Unido a esto habría que decir que los informes basados en la metodología cualitativa, principalmente las etnografías, han permitido conocer a fondo las realidades prácticas tal y como suceden a partir de una mayor implicación del investigador en ellas.

Sin embargo, este paradigma no está exento de críticas. Algunas de ellas son verdaderamente duras, como las recogidas por De Miguel (1988) y Fernández Cano (1996). Entre ellas se pueden mencionar las siguientes:

- Rechazo de los hechos frente a la preponderancia de los significados.
- Bajo nivel de elaboración teórica presente en los datos.
- Subjetividad desesperada, que lleva a diferentes intérpretes a obtener significados distintos.
- Irracionalidad de los datos introspectivos.
- Carácter exploratorio, pues no conlleva el contraste científico de hipótesis.
- Ausencia de un método formalizado.
- La precariedad en la instrumentación.
- Su verificación.
- Escasa utilidad ya que las repercusiones sobre la práctica son nimias, no acarrear mejoría.
- Eticidad cuestionable, pues indaga en profundidades íntimas que una persona no está dispuesta, libre y naturalmente, a manifestar.
- Singularidad y privacidad de los métodos de análisis de datos.
- Tamaño muestral escaso, que minimiza la posibilidad de generalizar resultados.

3.4. El paradigma crítico

El paradigma crítico agrupa una serie de enfoques de investigación que a pesar de estar en sus principios básicos muy cercanos al paradigma interpretativo han pretendido responder a lo que según ellos tiene de conservadurismo dicho paradigma, al no cuestionar a través de la investigación el *statu quo* establecido. Como dice Mateo (2000a) pretende superar el reduccionismo del paradigma positivista y el conservadurismo del interpretativo, admitiendo la posibilidad de una Ciencia Social que no sea ni puramente empírica ni únicamente interpretativa.

Otras denominaciones con las que es conocido son paradigma *sociocrítico*, *investigación-acción*, *emancipador*, *ciencia crítica*, *orientado al cambio*, etc.

Básicamente, la investigación bajo el prisma del paradigma crítico se inicia a partir de la crítica a la situación con el objetivo de construir situaciones o realidades (sociales, escolares, comunitarias) más justas. Por ello, muchas veces a este tipo de investigación se le suele llamar investigación como praxis orientada.

Esta teoría tiene una de sus raíces más importantes en la tradición alemana de la Escuela de Frankfurt. Sus principios ideológicos tienen como finalidad el análisis de las transformaciones de las diferentes estructuras de las relaciones sociales y además dar respuesta a los problemas generados por dichas estructuras. Además de la Escuela de Frankfurt se apoyan en el neomarxismo, la teoría crítica social de Habermas, y los trabajos de Freire, Carr y Kemmis, etc.

Colás y Buendía (1998); Arnal, Del Rincón y Latorre (1992), y Fraenkel y Wallen (2000) sintetizan los planteamientos que se realizan desde el paradigma crítico:

- Cualquier búsqueda de conocimiento sobre el mundo real está destinado al fracaso, pues no hay forma de demostrar que algo existe, ya que únicamente existen percepciones individuales de esa realidad. Desde el planteamiento crítico, la realidad es dinámica y evolutiva y son los propios sujetos los agentes activos destinados a configurar y construir esa realidad.
- Las teorías y las leyes son imprecisas porque el lenguaje que utilizamos para identificar las cosas está lleno de imprecisiones. En este sentido, la teoría y la práctica forman un todo inseparable y no se puede hablar de teorías universales. El objetivo de la teoría es la formación del carácter en los hábitos de reflexión.
- En Ciencias Sociales, el objeto de investigación y las cuestiones relacionadas con la investigación no son objetivos sino que están influenciados por los valores de la sociedad. Los valores ocupan un lugar muy importante en la perspectiva crítica.
Se considera que la investigación está al servicio de intereses políticos, por lo cual el principio de neutralidad de la ciencia es cuestionado. Además se achaca tanto al paradigma positivista como al interpretativo de conservadurismo en el sentido de que como mucho, a través de la investigación se pretenden mejorar las prácticas existentes pero sin cuestionarlas. La ciencia y por consiguiente la investigación no deben limitarse solamente a explicar (positivismo) y comprender (naturalismo) esa realidad, sino también a transformarla, a mejorarla.
- Por último, otra de las características que va a diferenciar claramente este paradigma de las posiciones positivista e interpretativa es el hecho de que los participantes se convierten en investigadores y los investigadores participan también en la acción social. Con esto, lo que se crea es un ambiente de colaboración total entre los investigadores y los participantes. Según Tójar (2001), es la indisoluble relación entre investigación y acción el elemento clave para diferenciarse de los otros dos paradigmas y le da su estatus epistemológico y metodológico.

La cuestión de la **ideología** resulta ser un elemento fundamental dentro de este paradigma. Dice López Górriz que «desde esta perspectiva sólo es posible la investigación si se formula explícitamente la ideología sobre la que se sustenta y se dirige a obtener conocimientos orientados a la liberación y emancipación del hombre, a lograr una mejor distribución del poder y de los recursos en nuestra sociedad» (1998, p. 47). Esta aceptación del compromiso ideológico de la investigación es un planteamiento claramente revolucionario que rompe con muchos de los parámetros en que se dan las otras formas

de investigaciones, entre otros el del tipo de conocimiento a lograr a través de la permanente reflexión acerca de la práctica y el de la separación de la dicotomía entre sujeto y objeto de la investigación al implicarse fuertemente los prácticos en el proceso de investigación.

El paradigma crítico no constituye un conjunto uniforme. Ha recibido aportaciones relevantes de diferentes aproximaciones, entre las que cabe destacar el llamado posmodernismo, el feminismo, enfoques posestructuralistas, etcétera. Todos estos elementos son aportaciones que en algunos casos son contradictorias y que hacen difícil la definición de un marco común para todas ellas.

Los mayores éxitos en este tipo de investigación han tenido lugar en campos que, de alguna manera, se pueden considerar marginales, tales como la investigación femenina, etnografía crítica neomarxista e investigación participante (De Miguel Díaz, 1988). En el campo educativo, esta perspectiva está aportando modelos de acción para los campos de formación y trabajo del profesorado, desarrollo de centros educativos y quehacer de agentes externos que tratan de hacer investigaciones con el profesorado en los centros (Escudero, 1990).

En el cuadro 1.2 se exponen resumidamente los principales puntos expuestos que reflejan las diferencias entre los tres paradigmas en cuanto a sus posiciones teóricas. Hemos de aclarar que este cuadro y otros similares no deben entenderse como compartimentos estancos, sino como una forma didáctica de exponer una serie de conceptos teóricos. En realidad, estos cuadros son abstracciones parciales de la realidad, que es algo más compleja y global que las clasificaciones que se ofrecen en ellos. No debemos pensar que los investigadores, a pesar de que puedan defender uno de los paradigmas, son tan coherentes como para ceñirse exclusivamente a lo asignado a cada paradigma en el cuadro. Como dice Tójar, «ni existen objetivistas tan radicales, ni sociocríticos tan puros, ni investigadores naturalistas con todas las dimensiones tan claras como aparecen en la tabla» (2001, p. 54).

Por último habría que mencionar que la mayoría de las características que se mencionan para el paradigma interpretativo son igualmente válidas para el paradigma crítico. El paradigma crítico asume prácticamente todas las aportaciones del paradigma interpretativo, imprimiendo a éstas las especificaciones más relacionadas con el cambio y transformación social. De esta forma las diferencias más significativas se dan en cuanto a la finalidad de la investigación, el papel de los valores en el proceso y la relación sujeto/objeto.

3.5. Características de los métodos cuantitativos y cualitativos

A pesar de que a nivel epistemológico se han planteado tres paradigmas, a nivel metodológico sólo cabría hablar de dos grupos o posibilidades: métodos cuantitativos y cualitativos. Esto ha llevado a que, en algunos casos se hable

1. Naturaleza de la investigación y evaluación en educación

Cuadro 1.2. Paradigmas de investigación educativa. Posturas teóricas

	Positivista	Interpretativo	Crítico
<i>Denominaciones</i>	Empírico-analítico Objetivista, Realista Cuantitativo Racionalista	Humanista, Naturalista Hermenéutico Cualitativo Ecológico Fenomenológico Etnográfico	Sociocrítico Orientado al cambio Investigación acción Racionalidad emancipadora Ciencia crítica de la Educación
<i>Fundamentos teóricos</i>	Positivismo Pospositivismo Empirismo	Fenomenología Teoría interpretativa Antropología	Teoría crítica Praxeología
<i>Naturaleza de la realidad (Ontología)</i>	Única, objetiva Estática, dada Fragmentable Convergente Externa, Observable Sujeta a un orden	Múltiple Dinámica Construida Holística Divergente Interna, Subjetiva	Múltiple, Dinámica Construida, Holística Divergente, Evolutiva Interactiva Compartida, Histórica
<i>Lógica</i>	Hipotética Deductiva, Inductiva	Inductiva/descriptiva Interpretativa	Inductiva
<i>Tipo de conocimiento</i>	Generalizaciones Leyes nomotéticas Técnico	Explicación idiográfica Hipótesis de trabajo Práctico, Inductivo	Explicación idiográfica Emancipador Cambio, Práctico
<i>Objetivos de la ciencia y la investigación</i>	Describir, Analizar Explicar, Predecir Controlar fenómenos Verificar hipótesis Construir teorías Buscar leyes	Comprender Interpretar Descubrir significados Hipótesis de trabajo	Liberar Emancipar Mejorar Transformar Criticar Identificar potencial de cambio
<i>Valores (axiología)</i>	Excluidos Libre de valores Neutralidad	Incluidos Explicitos Influyentes	Integrados Compartidos
<i>Ética</i>	Extrínseca	Intrínseca	Intrínseca
<i>Relación sujeto/objeto</i>	Independientes Distanciados Neutralidad Investigador externo	Dependencia Interrelación Implicación investigado	Interrelación por compromiso Investigador es uno más
<i>Relación teoría y práctica</i>	Independientes Teoría es norma para la práctica	Relacionadas, Unidas Retroalimentación mutua	Indisociables La práctica es teoría en acción Relación dialéctica
<i>Rol de la teoría</i>	Construcción y verificación de las teorías como objetivo central	Construcciones teóricas emergen de la situación	Las construcciones teóricas se llevan a cabo en la comunidad de forma cooperativa

de dos paradigmas, el paradigma cuantitativo y el cualitativo. Desde nuestro punto de vista este planteamiento no es del todo correcto, pues como hemos dejado claro anteriormente, el paradigma crítico, a pesar de ser un paradigma antipositivista y compartir la mayoría de las características de la metodología cualitativa, tiene algunas características diferenciales con el paradigma interpretativo.

Por tanto, a nivel metodológico debemos hablar de métodos cuantitativos y métodos cualitativos.

La **investigación cuantitativa** es una investigación básicamente confirmatoria, inferencial e hipotético-deductiva. El investigador se sitúa fuera de lo investigado, siendo un elemento externo al objeto que se investiga. También se la ha venido denominando como investigación experimental, empírica y estadística.

Los problemas de investigación que se plantean los métodos cuantitativos surgen de los postulados o teorías que ya existen a través de la iniciativa del investigador, siendo el estudio de la bibliografía científica una de las fuentes más habituales. Son considerados como problemas cerrados en el sentido que van a quedar totalmente determinados antes de la recogida de datos.

El diseño de la investigación, igual que el problema, queda totalmente definido antes de la recogida de datos. Es por tanto un diseño cerrado y totalmente estructurado y hecho desde fuera por la persona que va a llevar a cabo la investigación. Es en la fase de diseño cuando deberá quedar garantizada la validez de la investigación, validez que descansa, fundamentalmente, en el control de variables secundarias y posibles errores.

Los datos que se recogen son datos denominados objetivos y son analizados estadísticamente. Muchas veces los sujetos participantes son también elegidos según criterios estadísticos, ya que se pretende generalizar los resultados obtenidos a poblaciones más amplias de las participantes en la investigación.

Las características que mejor definen la **investigación cualitativa** son las de exploratoria, descriptiva, inductiva, cercana a los datos y no generalizable. Se interesa más por la calidad que por la cantidad. El investigador se introduce dentro de las situaciones de investigación, participando de ellas.

En el caso de los métodos cualitativos los problemas de investigación surgen muchas veces de las necesidades que se dan en los propios grupos sociales. Son problemas conocidos como abiertos, en el sentido de que pueden cambiar en la medida en que la investigación va avanzando.

El diseño o plan es flexible ya que, de la misma manera que el problema de investigación, puede ir variando a lo largo del proceso de investigación. Otra característica es la de, en algunos casos, ser abierto a la intervención de los sujetos participantes en la planificación de la propia investigación. Por esto se suele decir que es un diseño «desde dentro». También se dice que es un diseño emergente en el sentido de que se va reajustando en la medida en que va siendo puesto en práctica.

Los criterios de calidad difieren de los de la investigación cuantitativa y giran alrededor de lo que se denomina credibilidad. Guba (1989) hace una

magnífica exposición de los criterios a utilizar para asegurar la credibilidad de las investigaciones cualitativas y plantea una lista de métodos posibles para ello. Entre estos métodos o técnicas se pueden destacar las siguientes: triangulación de fuentes, métodos, etc.; observación persistente; contrastación con la visión de los sujetos participantes; muestreo teórico; etc. En los últimos años han surgido otros criterios de calidad que, aunque con menor fuerza que los de Guba, poco a poco van encontrando acomodo en la literatura científica.

Los datos recogidos se basan fundamentalmente en observaciones y entrevistas más o menos abiertas que nos van a ofrecer información más textual que numérica. Se les suelen llamar datos profundos, ricos, reales. El investigador es el propio instrumento. El análisis de los datos se realiza por medio de procesos inductivos a través de un análisis profundo de los significados de esos datos para los denominados actores.

Además de las diferencias notorias entre los datos cuantitativos y cualitativos, entre los instrumentos y estrategias para recoger dichos datos y entre los análisis cuantitativos y cualitativos, una de las características que mejor definen la diferencia entre la investigación cualitativa y cuantitativa es la propia naturaleza del proceso de investigación. Indudablemente las diversas fases del proceso de investigación están interrelacionadas entre ellas, pero mientras que en la metodología cuantitativa se da una linealidad entre las distintas etapas (una etapa sigue a la otra y no vuelve a repetirse) en la investigación cualitativa podríamos hablar de un proceso interactivo en el sentido de que una fase de la investigación puede replantear decisiones tomadas anteriormente. En este sentido suele ser muy habitual diseñar una nueva recogida de datos no prevista al principio a partir de la interpretación de datos recogidos.

En el cuadro 1.3. se expone un resumen de las características metodológicas de los distintos paradigmas. En este caso habría que mencionar que las investigaciones realizadas bajo el paradigma crítico están basadas en métodos cualitativos. La investigación-acción es investigación cualitativa y mucha de la investigación evaluativa que se hace hoy en día también lo es. Hubiera cabido la posibilidad de plantear el esquema como las diferencias entre los métodos de investigación cuantitativos y los cualitativos, pero nos ha parecido más interesante hacerlo de esta manera, pues así se recogen los matices inherentes a la investigación bajo el prisma del paradigma crítico. Es por ello por lo que en algunos casos no se repiten las características comunes a los paradigmas interpretativo y crítico.

4. Algunas cuestiones acerca de los paradigmas

Buena parte de las tres últimas décadas del siglo pasado se ha caracterizado por el debate epistemológico entre los diversos paradigmas, fundamentalmente entre lo que podríamos llamar el enfoque cuantitativo y el cualitativo.

**Cuadro 1.3. Paradigmas de investigación educativa.
Características metodológicas**

	Positivista	Interpretativo	Crítico
<i>Metodología general</i>	Cuantitativa	Cualitativa	Cualitativa
<i>Métodos</i>	Experimental Correlacional De encuestas	Etnográfico Estudio de casos Fenomenología	Investigación-Acción Participativo Cooperativo
<i>Problema de investigación</i>	Teóricos Cerrados	Percepciones y sensaciones Abiertos	Vivenciales
<i>Diseño</i>	Estructurado Predeterminado	Emergente Abierto Flexible	Dialéctico Negociado
<i>Muestra</i>	Procedimientos estadísticos	No determinada Informante	El grupo de investigación
<i>Recogida de datos</i>	Instrumentos válidos y fiables Pruebas objetivas Cuestionarios y entrevistas estructuradas Observación sistemática	Técnicas cualitativas Cuestionarios abiertos y entrevistas informales Diarios Observación participante Descriptivas Investigador principal instrumento	Comunicación personal Técnicas dialécticas Estudios de casos Diarios de campo
<i>Análisis de datos</i>	Cuantitativo Técnicas estadísticas descriptivas e inferenciales	Cualitativo Inducción analítica Triangulación	Cualitativo Dialéctica
<i>Conclusiones</i>	Descripciones superficiales Explicaciones Predicciones Leyes	Descripciones en profundidad Hipótesis	Decisiones Valoraciones
<i>Criterios de calidad</i>	Validez interna Validez externa Fiabilidad Objetividad	Credibilidad Transferibilidad Dependencia Confirmabilidad Veracidad Autenticidad	Validez consensuada Intersubjetividad Contextual Estimula la acción

Alvira ya en 1982 decía que la polémica estaba viciada de raíz por las razones siguientes :

- a) Se identifica cada método con un paradigma.
- b) Se afirma la radical oposición entre ambos paradigmas a la vista de unos presupuestos metateóricos.
- c) Se oculta la posibilidad de utilización conjunta de lo cualitativo y lo cuantitativo.

El debate entre defensores y detractores de los distintos paradigmas ha generado diversas posturas. Dentro de esta confrontación, más que cuestiones de fondo, muchas veces lo que han primado han sido descalificaciones con una cierta carga ideológica. Burgess (1993) recogió los términos utilizados por los investigadores participantes en un congreso sobre investigación educativa para referirse a los paradigmas, tanto si estaban a favor como si eran contrarios al paradigma en cuestión. Algunos términos utilizados para referirse a la investigación cuantitativa son los siguientes: duro, abstracto, explicativo, científico, objetivo, libre de valores, deductivo, riguroso, nomotético, malo, bueno. Para referirse a la investigación cualitativa algunos términos recogidos son: blando, flexible, fluido, precientífico, subjetivo, inductivo, político, idiográfico, bueno, malo.

Sin embargo, no todas las actitudes han sido de descalificación. Aparte de estas posturas ha habido serios intentos de situar el debate de una forma más racional y constructiva. A partir de los años ochenta la situación de enfrentamiento virulento en el que se pugnaba por la imposición de un modelo ha dado paso a una situación de mayor tranquilidad en la que se acepta la posibilidad de influencia mutua entre los distintos paradigmas.

Para Walker y Evers (1997), se puede hablar de tres momentos diferentes en el debate epistemológico entre paradigmas:

- a) *Incompatibilidad entre los paradigmas*. Se considera que los distintos paradigmas (dos o tres) son epistemológicamente diferentes; por lo tanto, son inconmensurables e irreconciliables.
- b) *Complementariedad*. Al igual que la postura anterior, se acepta la idea de que existen paradigmas distintos e inconmensurables, pero que cada uno de ellos es defendible.
- c) *Unidad epistemológica*. Se rechaza el concepto de enfrentamiento entre paradigmas y se plantea que lo que se da es un enfrentamiento entre teorías que se superará en la medida que cada una de las teorías demuestre una mejor adecuación a los diferentes casos.

Como se puede observar, algunos han defendido la elección de un único paradigma, bien porque lo consideran superior a otro u otros a los niveles ontológico y epistemológico (Smith y Heshusius, 1986), o bien porque a nivel

práctico lo consideran el más adecuado a la investigación educativa (Carr, 1989; Barbier, 1996). Para algunos de éstos, el único paradigma válido será el positivista porque es el único que va a permitir la obtención de conocimiento científico libre de juicios de valor. Para otros en cambio, el interpretativo es el único válido, pues sus métodos son los únicos que posibilitan un acercamiento a la comprensión profunda de los fenómenos sociales. Por último tenemos los que defienden que la única forma de llegar a transformar la sociedad va a ser a través del paradigma crítico pues los otros dos, aunque utilizando métodos diferentes, perpetúan el actual estado de cosas.

En un segundo bloque se encuentran los que han venido defendiendo la opción de no tener que definirse por uno de los paradigmas como el único o el mejor, defendiendo la complementariedad paradigmática, en el sentido de compaginar los métodos cuantitativos y cualitativos en función del problema a estudiar e incluso buscando vías de superación de la situación de enfrentamiento.

Otra postura dentro de este bloque es la de aquellos que abogan por una serie de alternativas que integren o superen los tres paradigmas. Para algunos, esta alternativa sería un nuevo paradigma, lo que De Miguel Díaz (1988) llama un «paradigma para el cambio», o Nisbet (1988) denomina «investigación orientada a la decisión y al cambio».

Independientemente de que a esta alternativa se le dé o no el rango de paradigma, lo que sí está claro es que este enfoque persigue la aplicación de los resultados de la investigación con vistas a la mejora educativa a través de la toma de decisiones para la acción. Teniendo en cuenta que para tomar decisiones o para actuar es necesaria la adopción de algún tipo de marco teórico o de interpretación, desde este planteamiento se puede considerar que se pretenden integrar las dos grandes visiones o paradigmas, las denominadas positivista y la naturalista.

A la hora de llevar a la práctica estos planteamientos, el mismo Nisbet (1988) plantea cuatro propuestas:

- a) *Estudios tipo encuesta* que, tras una recogida de datos relevantes acerca de los hechos, permiten tomar decisiones
- b) *Estudios experimentales* similares a los de encuesta pero diseñados para la comprobación de hipótesis.
- c) *Estudios de desarrollo* para la mejora de las políticas educativas.
- d) *Estudios de evaluación* de decisiones tomadas en el pasado. En esta línea de investigación dirigida a la toma de decisiones se puede incluir también la investigación evaluativa en el sentido de que persigue tomar decisiones acerca de la eficacia de los programas sociales y/o educativos.

En esta misma línea de superar o integrar los distintos paradigmas están las aportaciones realizadas por Howe en diferentes artículos en los años 1985,

1988 y 1992. El punto de partida es la negación de la obligatoriedad de elegir entre métodos de investigación cualitativos y cuantitativos sugiriendo la combinación de ambos. Más adelante planteará la necesidad de combinar lo cuantitativo y lo cualitativo estableciendo para ello los niveles de los datos, diseño, análisis e interpretación. Por último, y argumentando que el positivismo es insostenible y el interpretativismo incompleto, sostiene una perspectiva epistemológica capaz de superar los antagonismos producidos por los paradigmas tradicionales integrando rasgos de un enfoque y otro.

Otra alternativa es la superación del debate paradigmático centrando la discusión a los niveles metodológico y técnico (los planteamientos de las líneas anteriores han estado centrados en los aspectos más teóricos o epistemológicos del debate paradigmático), para llegar a situaciones de integrar en una misma investigación métodos cuantitativos y cualitativos. Así, Cook y Reichardt (1986), que son también favorables a la integración de perspectivas a nivel de objetivos, plantean esta complementariedad también a nivel metodológico reclamando flexibilidad para un uso combinado de las metodologías asociadas a los dos paradigmas.

En la línea anterior estarían las aportaciones de Dendaluce (1995, 2001) en lo que él denomina «pluralismo integrador». Algunas características de esta propuesta son las siguientes:

- No es una síntesis de dos metodologías que crean una nueva.
- Se propone una mejora de la metodología que más nos convence en general o que nos parece más apropiada para el problema que estamos estudiando.
- Se trata de profundizar en la metodología propia e integrar en ella aportaciones de otras.
- Una de las implicaciones de esta propuesta es la de que se buscará trabajar en equipos de investigación interdisciplinares.

Este autor propuso una serie de aspectos formales básicos y comunes que comparten todas las metodologías de investigación a los que se denomina «aspectos críticos». En el cuadro 1.4, tomado de Rodríguez Conde (1999) se recogen los puntos más significativos en los que debería coincidir cualquier tipo de investigación.

Se puede afirmar que el debate a nivel metodológico está más superado o en vías de superación en relación al nivel ontológico y epistemológico. Así, es en el ámbito de la evaluación donde se dan las más claras experiencias de integración de métodos cuantitativos y cualitativos, básicamente en lo relacionado con la recogida y análisis de la información.

Aunque parece ser que epistemológicamente la compatibilidad no es muy coherente, hay que aceptar que se produzcan investigaciones en esta línea porque además de favorecer el debate y discusión pueden ayudar en la búsqueda de soluciones prácticas para la investigación educativa.

Cuadro 1.4. Puntos críticos de coincidencia entre la metodología cuantitativa y cualitativa (Dendaluce, 1995, adaptado por Rodríguez Conde, 1999)

Puntos críticos	Aspectos básicos comunes
1. <i>Bases conceptuales</i>	<ol style="list-style-type: none"> 1. Partir de una concepción coherente del concepto de ciencia. 2. Postura definida en el debate paradigmático y en el debate cuantitativo-cualitativo. 3. Situar el papel de los valores. 4. Tener su ontología y epistemología. 5. A nivel metodológico, situar al investigador frente al objeto, decidir el papel de los participantes, la temporalización, las técnicas a emplear, el tipo de análisis, etc.
2. <i>Planteamiento del problema de investigación</i>	<ol style="list-style-type: none"> 1. Explorar la realidad a investigar. 2. Plantearse objetivos o interrogantes sobre los que averiguar, describir o transformar. 3. Fundamentar y contextualizar el problema a través de la revisión bibliográfica. 4. Sobre qué problema investigar, unas metodologías se adaptan mejor que otras, pero tendría que existir cierta flexibilidad; así como en el cómo investigar, incluyendo el proceso de negociación para unos (cuantitativos) y reconstruyendo los objetivos en cada fase abierta, flexible y cíclica (cualitativos).
3. <i>Planteamiento metodológico para la resolución del problema</i>	<ol style="list-style-type: none"> 1. Calidad del diseño: los cuantitativos buscan la validez y fiabilidad; los cualitativos persiguen la credibilidad, transferibilidad, dependencia y confirmabilidad. 2. Ganancias en el diseño: los cuantitativos, a través de adaptarse flexiblemente a problemas más abiertos e incluir a otros participantes en la toma de decisiones; los cualitativos, sistematizando explicaciones alternativas estudiando variables relevantes, así como definir el diseño en cada cambio, reconstruyendo el diseño en el informe. 3. Muestras intencionales: evitando el sesgo en ambos casos, describiéndola suficientemente. Cuando el estudio requiere de poblaciones, utilizar el metaanálisis. 4. Datación (recogida y elaboración de todo tipo de datos): Intentar sistematizar la información recogida, clasificando por distintos criterios los datos cuantitativos y cualitativos. Datación y análisis de datos englobaría, por un lado, la medición y datación cualitativa y, por otro, el análisis estadístico y el análisis cualitativo. Tomar en cuenta los avances en medición para ambos tipos.
3.1. <i>Diseño/plan</i>	
3.2. <i>Datos: Sujetos y datación</i>	<ol style="list-style-type: none"> 5. Análisis cuantitativo de datos cuantitativos (análisis estadístico de datos resultado de la medición nominal, ordinal, de intervalo y de razón). 6. Análisis cuantitativo de datos cualitativos (análisis estadístico de datos cualitativos cuantificados; por ejemplo, análisis estadístico de textos). 7. Análisis cualitativo de datos cuantitativos (a veces por razones de cumplimiento de supuestos estadísticos o problemas de poder). 8. Análisis cualitativo de datos cualitativos. 9. Metaanálisis.
3.3. <i>Análisis de datos</i>	
4. <i>Globalización (y conclusiones) de la investigación</i>	<ol style="list-style-type: none"> 1. Interpretar lo que de verdad representan en su conjunto los resultados respecto a los objetivos propuestos. 2. Integrar las distintas aportaciones teóricas, históricas, legislativas y empíricas entre sí y con lo que los autores han concluido sobre el tema. 3. Discutir la investigación en su conjunto. 4. Sacar conclusiones como el producto natural de la globalización.

En este sentido son elocuentes las palabras de la profesora Pérez Serrano, quien hace una llamada a no prescindir de las valiosas aportaciones que pueda ofrecernos cualquiera de los paradigmas en estos tiempos, que ella denomina de crisis:

En las Ciencias Sociales existe una gran preocupación teórica y práctica, de modo que es difícil separar el conocimiento y la acción como camino válido para acercarse al estudio de la realidad. La preocupación por el saber va unida a la preocupación por el actuar, por mejorar la acción pues lo que interesa es la transformación y la mejora de la realidad social. En este sentido podemos afirmar que la tarea es ingente y necesitamos contar con las aportaciones de los distintos paradigmas.

(Pérez Serrano, 1994, p. 41)

5. Los paradigmas de la evaluación

Las reflexiones acerca de los paradigmas propuestos en los apartados anteriores, hacen referencia a la investigación educativa en general. Pero ¿estos mismos planteamientos pueden ser extrapolados a la evaluación educativa? O, dicho de otra manera, ¿es posible distinguir los tres paradigmas de investigación (positivista, interpretativo y crítico) en la evaluación educativa? Tradicionalmente, la respuesta a las dos cuestiones ha sido en sentido afirmativo. Así podemos encontrar que la mayoría de los autores hablan de los paradigmas de la evaluación trasladando los mismos postulados que se proponen en la investigación educativa. Esto es debido a que la evaluación educativa, entendida como evaluación de programas adaptó como sinónimo el concepto de investigación evaluativa¹.

La historia de la investigación evaluativa va unida a la de la misma evaluación de programas y al surgimiento de los debates paradigmáticos y metodológicos.

(Martínez Mediano, 1996, p. 36)

De esta forma estaríamos ante un tipo de investigación cuyo propósito es atender a las necesidades de su objeto o programa, pero es ante todo un TIPO DE INVESTIGACIÓN.

[...] la investigación evaluativa dado que es un tipo de investigación aplicada del ámbito de las Ciencias Sociales y de la Educación.

(Martínez Mediano, 1996, p. 36)

Al utilizarse de forma indistinta evaluación de programas e investigación evaluativa es por lo que tradicionalmente se ha hablado de los paradigmas de la evaluación.

Sin embargo, en una reciente aportación, De Miguel (2000a) critica esta extrapolación. Vamos a intentar sintetizar las ideas principales expuestas por

el citado autor, ya que consideramos que pueden contribuir a aclarar las bases teóricas de la evaluación.

En primer lugar, De Miguel distingue entre evaluación de programas e investigación evaluativa, ya que si bien aluden a un mismo campo de trabajo «la investigación evaluativa constituye esencialmente una estrategia metodológica orientada a la búsqueda de evidencias respecto a un programa, pero la evaluación de un programa requiere, además, abordar otros problemas relacionados con los criterios a utilizar para emitir juicios de valor y en los procesos relativos a la toma de decisiones» (De Miguel, 2000a, p. 290).

A partir de esa distinción, el autor rechaza la traslación que comúnmente se realiza de los paradigmas propios de la investigación social a la evaluación de programas como si fueran equivalentes. Esta extrapolación carece de sentido puesto que la evaluación de programas exige necesariamente contemplar el mundo de los valores y de las consecuencias y, por ello, requiere abordar ese tema desde una perspectiva más comprensiva.

Ninguno de los paradigmas clásicos ofrece una base epistemológica adecuada al tipo de conocimiento que requiere la evaluación y, por ello, propone avanzar en marcos paradigmáticos novedosos como el propuesto por Pawson y Tilley (1997) orientado hacia la explicación social realista cuyos supuestos teóricos se sitúan entre el positivismo y el relativismo. En esta misma línea Mateo (2000a, p. 138) argumenta que:

Aun con todo, no queda claro en la actualidad el punto de corte entre las aproximaciones evaluativas. Parece que las tendencias en los últimos años más bien se inclinan por el uso conjunto y mixtificado de ellas, atendiendo más a la naturaleza del problema y a criterios de practicidad y pragmatismo que a principios epistemológicos.

En segundo lugar, señala De Miguel (2000a) que, aunque los evaluadores obvian en muchos casos la justificación teórica para explicitar únicamente la opción metodológica elegida, en esta elección subyace la concepción epistemológica y ontológica que tienen los evaluadores sobre los conceptos de evaluación y programa. A partir de aquí se podrán señalar las ventajas y los inconvenientes de cada aproximación metodológica, pero lo que debería ser esencial es que el evaluador justificara el marco conceptual asumido en la elección de dicha aproximación puesto que éste debe estar supeditado a los planteamientos teóricos. Ante esta situación, señala que para elegir un determinado enfoque teórico y metodológico para evaluar un programa, el evaluador deberá aclarar las siguientes cuestiones básicas:

- El concepto de programa como estrategia de intervención social.
- Las bases epistemológicas que determinan la construcción de las evidencias.
- Los criterios a utilizar para formular los juicios de valor.
- La utilidad y los usos de los informes de evaluación.
- El enfoque metodológico a seguir para realizar el proceso evaluativo.

En tercer lugar, aludiendo al concepto de programa² como estrategia de intervención social, el evaluador debe explicitar la concepción teórica desde la que aborda el programa. De Miguel (2000a) distingue tres diferentes:

a) *El programa como verificación de teorías*

Desde este punto de vista, el programa es considerado como una estrategia que permite la comprobación de teorías que sirven para resolver problemas sociales. En este caso, el objetivo de la evaluación es probar las relaciones causa-efecto que se dan entre unos tratamientos (programas) y sus resultados. Es decir, a través de los programas se someten a prueba determinadas teorías para ver si son eficaces para la resolución de problemas sociales prácticos. Para ello, es necesario utilizar todos los componentes de la experimentación científica con diseños metodológicos que intentan garantizar la validez interna y externa. El evaluador asume el rol de experimentador.

b) *El programa como tecnología sistemática*

Desde este punto de vista, la evaluación de programas debe centrarse en los procesos de instrumentación de las estrategias de intervención, puesto que son las contingencias que se establecen las que determinan las diferencias, dejando a un lado los problemas de causalidad de las teorías. Por ello, la evaluación de programas deberá tener los objetivos siguientes en cada una de las fases:

1. Estimación del mérito de las metas que se propone cada estrategia.
2. Calidad del diseño y planificación de las actuaciones a realizar.
3. Grado de cumplimiento y adecuación con que dicha planificación ha sido llevada a cabo (implementación).
4. Calidad y utilidad de los resultados e impactos generados.

El evaluador asume el rol de programador social.

c) *El programa como compromiso con el cambio social*

Desde este punto de vista, la justificación teórica del programa es promover el cambio y la transformación social por lo que también éste debe ser el objetivo de la evaluación. Por ello, el evaluador debe cuestionar los valores y las consecuencias que subyacen en cada estrategia de intervención social, es decir, además de su supuesta capacidad técnica debe poseer un compromiso activo en la resolución de problemas sociales. El evaluador asume el rol de agente de cambio.

En cuarto lugar situamos el debate de la utilización de la metodología cuantitativa (metodología deductiva) o la metodología cualitativa (metodología inductiva). Como bien señala De Miguel (2000a), la elección de una u otra vendrá determinada por el tipo de conocimientos que se necesitan y el

porqué de esos conocimientos. En este sentido ambos tipos de procedimientos son adecuados en función de las necesidades:

- Cuando buscamos relaciones entre inputs y outputs o resultados generalizables utilizaremos metodología deductiva.
- Cuando buscamos la comprensión y el análisis de una intervención para contribuir a optimizar el programa emplearemos metodología inductiva.

Por ello, dado que los propósitos de la evaluación van a condicionar la adecuación más óptima de la metodología a utilizar (Martínez Mediano, 1996), esto significa que desde la perspectiva epistemológica no existe una supremacía de un método sobre el otro.

De hecho, una de las principales características de la evaluación de programas es que sus métodos cubren toda la gama de paradigmas que dominan en la investigación social [...]. De ahí que, la evaluación de programas constituye actualmente el ámbito de la investigación social aplicada en el que se hace más patente la superación del debate entre metodología cuantitativa y cualitativa, lo cual significa que clasificar los modelos evaluativos según este criterio carece de sentido.

(De Miguel, 2000a, p. 298)

En quinto lugar, hemos recogido la problemática que plantea De Miguel (2000a) respecto a los criterios de valor. Tras la recolección de la información pertinente se deben formular juicios de valor basados en algunos criterios o marcos de referencia. Tradicionalmente se distinguen dos tipos de criterios:

- *Prescriptivos*, cuando el propio evaluador define los criterios.
- *Descriptivos*, cuando se asumen los valores de otros como criterios para enjuiciar el programa.

En la evaluación de programas sociales, usualmente, se utilizan criterios descriptivos, sin embargo, estos pueden ser obtenidos atendiendo a una de las opciones siguientes:

- a) *Criterios políticos*; es decir, coinciden con los objetivos que pretenden alcanzar los organismos que patrocinan los programas. En este caso, el evaluador acepta los valores del patrocinador.
- b) *Criterios democráticos*; es decir, las diferentes audiencias implicadas en el programa son las que mejor conocen el funcionamiento del mismo y, por ello, sus criterios constituyen el marco de referencia para emitir los juicios de valor. En este caso, se habla de una valoración descriptiva pluralista basada en criterios democráticos.

- c) *Criterios técnicos*; es decir, aquella en la que se aplican instrumentos y procedimientos científicos apropiados, propios de la teoría normativa sobre evaluación de programas: la calidad de las metas propuestas, la coherencia del diseño, la adecuación de la implantación, la efectividad de los resultados e impactos y la oportunidad y viabilidad de las decisiones tomadas en relación con los procesos de mejora.

6. Evaluación e investigación

Aunque en el capítulo 3, relativo al concepto de evaluación, se expondrá más pormenorizadamente, es preciso que señalemos lo que entendemos por evaluación para confrontarlo con la evaluación de programas (aspecto que se tratará más en profundidad en el capítulo 4), la investigación evaluativa y la investigación científica.

Entendemos por evaluación el proceso de identificación, recogida y análisis de información relevante, que podrá ser cuantitativa o cualitativa —de manera sistemática, rigurosa, planificada, dirigida, objetiva, creíble, fiable y válida— para emitir juicios de valor basados en criterios y referencias preestablecidos para determinar el valor y el mérito del **objeto educativo** en cuestión, a fin de tomar decisiones que ayuden a optimizar el mencionado objeto. Hemos remarcado el **objeto educativo** puesto que éste puede variar, es decir, cualquier componente de la educación podría ser susceptible de ser evaluado. De tal manera que cuando ese objeto es un programa educativo, hablaríamos de evaluación de programas. En este caso estaríamos aceptando la tesis defendida por Martínez Mediano (1996) de que la evaluación de programas educativos es una subárea de la evaluación educativa. Sin embargo, en muchos casos se utiliza el concepto de «programa» dentro de la evaluación de programas como el objeto de la evaluación. En este sentido señala Bisquerra (1989, p. 144): «en general se utiliza el término programa para referirse a los objetos³ de la evaluación». De tal forma que desde esta perspectiva la evaluación educativa y la evaluación de programas⁴ serían utilizados como sinónimos.

Antes de adentrarnos en las diferencias entre la evaluación y la investigación, debemos todavía presentar la investigación evaluativa. En las últimas décadas, la evaluación ha experimentado un desarrollo histórico que la ha llevado a superar las resistencias iniciales de los investigadores para considerarla como una disciplina de conocimiento que poseía su propia teoría (Martínez Mediano, 1996). De esta manera, en 1974 aparece por primera vez el término de investigación evaluativa en el *Educational Index* (De la Orden, 1985a) vinculándose al campo de la evaluación de programas.

[...] proceso sistemático de recogida y análisis de información fiable y válida para tomar decisiones sobre un programa educativo. Este proceso en la medida en que es riguroso,

controlado y sistemático constituye un modo de investigación que se conoce con el nombre de investigación evaluativa.

(De la Orden, 1985a, p. 134)

En las últimas tres décadas, la investigación evaluativa ha tenido un desarrollo espectacular debido a su vocación por conocer la eficacia de los múltiples programas desde un intento político, social y del quehacer pedagógico (Pérez Carbonell, 2000). Esta autora señala que los términos de evaluación de programas e investigación evaluativa son sinónimos y, por lo tanto, pueden incluso utilizarse indistintamente. En esta misma línea se sitúan la mayoría de los autores puesto que ofrecen definiciones semejantes para los dos conceptos.

La evaluación de programas es un proceso sistemático, diseñado intencional y técnicamente, de recogida de información rigurosa —valiosa, válida y fiable—, orientado a valorar la calidad y los logros de un programa, como base para una posterior toma de decisiones de mejora tanto del programa como del personal implicado, y de modo indirecto, del cuerpo social en que se desarrolla.

(Pérez Juste, 2000, p. 272)

La evaluación de programas es el proceso sistemático de recogida, análisis e interpretación y de información relevante y fiable para describir un programa educativo, o una faceta, significativa del mismo, y formular un juicio sobre su adecuación a un criterio o patrón, que representa un valor aceptado, como base para la toma de decisiones sobre el programa o faceta programática.

(De la Orden, 2000, p. 383)

La investigación evaluativa es el proceso de recoger información pertinente, válida y fiable que nos permite emitir juicios de valor sobre la calidad y rentabilidad de un programa o sistema a fin de tomar las decisiones oportunas.

(De Miguel, 1995, p. 169)

La investigación evaluativa es un modo de investigación que implica un proceso riguroso, controlado y sistemático de recogida y análisis de información fiable y válida para tomar decisiones sobre un programa educativo.

(Tejedor, 2000, p. 320)

Por investigación evaluativa se entiende una modalidad destinada, fundamentalmente, a la evaluación de programas, sea en su sentido más tradicional, en el que la evaluación se considera una actividad encaminada a determinar su eficacia, sea, como se piensa en la actualidad, en un sentido mucho más amplio capaz de incluir el programa tanto en sus diferentes momentos —inicial, desarrollo y final— como en sus distintas dimensiones, como puede ser su contenido, su formulación técnica, su adecuación o cualquier otro aspecto considerado relevante. El concepto, asimismo, puede ser aplicado a innovaciones, centros, organizaciones o profesionales, como es el caso de los profesores.

(Pérez Juste, 1994a, p. 404)

Como puede observarse, en las definiciones precedentes, puede intercambiarse indistintamente en cualquiera de ellas «evaluación de programas» por «investigación evaluativa», y viceversa, y las definiciones siguen siendo igualmente válidas. Esto es debido a que, como hemos mencionado previamente, ambos términos son utilizados indistintamente por la mayoría de los autores.

Sin embargo, en una reciente publicación de De Miguel (2000a), ya mencionada previamente, se señala que aunque ambos pertenecen a un mismo campo de trabajo, su significado, *sensu stricto*, es distinto:

La investigación evaluativa (evaluation research) constituye la aplicación de los principios y procedimientos de la investigación social para comprobar sistemáticamente la eficacia de un programa de intervención social (p. 289).

Por contra, **la evaluación de programas** (*program evaluation*) es el

conjunto de principios, estrategias y procesos que fundamentan la evaluación de toda acción o conjunto de acciones desarrolladas de forma sistemática en un determinado contexto con el fin de tomar las decisiones pertinentes que contribuyan a mejorar las estrategias de intervención social (p. 289).

Según él, cuando se define la evaluación como «un proceso sistemático de búsqueda de evidencias para formular juicios de valor que orienten la toma de decisiones» (p. 290) se apuntan tres aspectos clave:

- *Las evidencias*; es decir, el proceso metodológico que se lleva a cabo para recoger e interpretar las evidencias relacionadas con el programa.
- *Los valores*; es decir, los criterios a utilizar para emitir los juicios de valor.
- *Las decisiones*; es decir, la utilización que se va a hacer de los resultados.

De estos aspectos, la investigación evaluativa se preocuparía únicamente del primero; sin embargo, la evaluación de programas, recogiendo también esa herramienta metodológica a utilizar en una de las fases del proceso, necesariamente ha de recoger también el segundo y el tercero. No obstante, como hemos visto anteriormente, muy pocos autores realizan la distinción reseñada por De Miguel, y, al definir investigación evaluativa, definen en realidad lo que el autor señala como evaluación de programas.

Hechas estas aclaraciones terminológicas, veamos cuáles son las similitudes y las diferencias entre la evaluación y la investigación.

La evaluación educativa surgió vinculada a la tradición positivista de la investigación educativa y, por ello, muchas veces durante un largo periodo

han llegado a utilizarse de forma entrelazada y en ocasiones confundidas en los textos (García Llamas, 1995).

Entre la evaluación y la investigación aparecen importantes puntos de contacto, ya que en los dos casos hablamos de un proceso sistemático, riguroso, controlado y además utilizamos técnicas de recogida y de análisis de datos. Como señala Jiménez (1999b), muchos de los pasos, etapas, instrumentos, metodología, etc., que se emplean en la evaluación están tomados de la investigación. Conocemos que en la investigación deben recogerse datos siguiendo con rigor las condiciones que exige el método científico (cualitativa y cuantitativamente entendido). Los métodos, pues, no son la dimensión que distingue la evaluación de la investigación. Las diferencias debemos buscarlas en los propósitos y los objetivos de cada una de ellas.

La investigación se caracteriza por aspectos tales como probar, extraer conclusiones, no implicar al investigador, generalizar y verdad científica. Bien al contrario, la evaluación se caracteriza por mejorar, tomar decisiones, no impliar al evaluador, valorar el fenómeno, etc. (García Llamas, 1995).

La siguiente cita de Nirenberg, Brawerman y Ruiz (2000, p. 37) resume de manera clara la diferencia entre los dos conceptos:

Una evaluación no es equivalente a una investigación, aun cuando comparte muchos de sus procedimientos. La investigación busca fundamentalmente incrementar el cuerpo de conocimientos, estableciendo relaciones entre hechos, fenómenos, situaciones, etc., mientras que la evaluación se preocupa ante todo por el perfeccionamiento de la acción y compromete esencialmente un juicio de valor. La investigación termina con conclusiones acerca de cómo ocurren los hechos, alimenta la teoría y el cuerpo de conocimientos sobre cierto segmento del saber, mientras que una evaluación debe terminar con recomendaciones que lleven a modificar lo que se venía haciendo, o bien agregar nuevas acciones.

Popham (1980) reduce a tres las diferencias entre evaluación e investigación:

- I. *Definición del problema*: a los evaluadores les interesan las decisiones, mientras que los investigadores pretenden extraer conclusiones.
- II. *Posibilidad de generalizar los resultados*: mientras que la evaluación se ciñe a un fenómeno educativo en particular, la investigación pretende extrapolar (generalizar) los descubrimientos obtenidos a otras situaciones.
- III. *Papel de la valoración en la investigación*: mientras que en la evaluación es obligatorio valorar un fenómeno educativo con el propósito de tomar decisiones, en la investigación se busca la verdad científica libre de estimaciones de valor.

De la Orden (1985a) considera que la investigación evaluativa (evaluación de programas utilizando la distinción realizada por De Miguel) exige el mismo grado de control y rigor que la investigación básica, la diferencia entre

1. Naturaleza de la investigación y evaluación en educación

ellas es la orientación. En el primer caso se busca valorar una situación concreta para tomar decisiones, y en el segundo, la producción de la teoría. Este mismo autor señala las características peculiares de la investigación evaluativa:

- a) Los juicios de valor están presentes en todo el proceso.
- b) La formulación de hipótesis precisas es difícil y a veces inapropiada.
- c) La replicación no es posible.
- d) La recogida de información está condicionada por la viabilidad del proceso.
- e) Multitud de variables relevantes escapan del control riguroso.
- f) La decisión de la continuidad, modificación, ampliación o sustitución del programa es responsabilidad del administrador del programa.
- g) El informe de evaluación debe adaptarse a las exigencias del patrocinador.

Casanova (1995a, p. 59) plantea las diferencias entre investigación y evaluación presentando en el cuadro 1.5 las fases de cada uno de ellas.

Como puede observarse, la investigación parte de una hipótesis, que mediante el estudio posterior termina con un informe descriptivo en el que se especifica el proceso seguido y las conclusiones obtenidas. La evaluación, sin embargo, no parte de una hipótesis previa, sino de la decisión de evaluar algún objeto. Sigue un camino paralelo al de la investigación en lo que se refiere a la planificación y a la ejecución, pero no termina con el informe, sino que debe valorarse el objeto evaluado para tomar las decisiones pertinentes. Como señala la autora, la evaluación y la investigación coinciden en las fases intermedias, pero se diferencian al comienzo y al final.

Cuadro 1.5. Fases de la investigación y de la evaluación

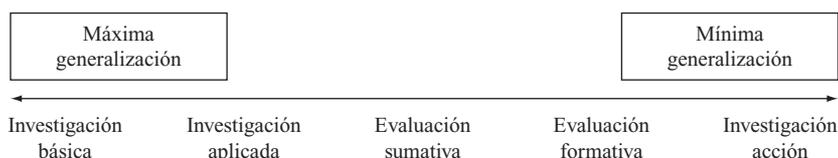
Fases de la investigación	Fases de la evaluación
Definición del problema.	Definición del problema.
Planteamiento de las hipótesis.	
Planificación: diseño, técnicas, instrumentos, tiempos, agentes, destinatarios.	Planificación: diseño, técnicas, instrumentos, tiempos, agentes, destinatarios.
Ejecución: recogida, análisis y tratamiento de los datos.	Ejecución: recogida, análisis y tratamiento de los datos.
Informe: descripción y conclusiones.	Informe: descripción y conclusiones.
	Valoración del objeto evaluado.
	Toma de decisiones.

Estas diferencias que hemos reseñado entre la evaluación y la investigación se refieren sobre todo a la confrontación de la evaluación con la investigación básica, no con la investigación aplicada.

En un claro intento de superar esta situación, Patton (1990) situó la investigación evaluativa dentro de un continuo en función del propósito de generalización de los resultados de la investigación (Martínez Mediano, 1996).

En uno de los extremos se situaría la investigación básica (propósito de máxima generalización) y en el otro la investigación-acción (propósito de mínima generalización) (véase figura 1.2).

Figura 1.2. Investigación, evaluación y generalización



En el cuadro 1.6 se establecen los propósitos de cada uno de los cinco procedimientos planteados.

Cuadro 1.6. Propósitos de generalización de la investigación y la evaluación. (Tomado de Patton, 1990, p. 150)

Investigación	Propósito
<i>Básica</i>	Contribuir a la fundamentación del conocimiento y la teoría.
<i>Aplicada</i>	Iluminar los asuntos que conciernen a la sociedad.
<i>Evaluativa sumativa</i> ⁵	Determinar la eficacia de los programas.
<i>Evaluativa formativa</i>	Mejorar un programa.
<i>Investigación-acción</i>	Resolver un problema específico.

Tal y como señala Tejedor (2000), la investigación evaluativa (entendida como evaluación de programas) se lleva a cabo en un proceso de intervención que es lo realmente importante. Por ello, la investigación/evaluación no es prioritaria y debe adaptarse al contexto del programa. Continúa el autor afirmando que:

El proceso de evaluación de programas debe concebirse como una auténtica estrategia de investigación sobre los procesos educativos, en cuyos resultados deberían basarse las pautas sugeridas para orientar los procesos de intervención. Investigación e intervención se

1. Naturaleza de la investigación y evaluación en educación

apoyan mutuamente. La intervención sacará partido de la investigación en la medida en que ésta se extienda a la solución de problemas prácticos. La investigación activa se asimila a la noción de investigación aplicada [...]. Toda situación educativa es susceptible de este tratamiento. Por una parte, existe la necesidad de «conocerla» y de «explicarla». Por otra, es preciso «comprenderla» y «mejorarla».

(pp. 320-321)

El cuadro 1.7 presenta las diferencias entre investigación y evaluación, y aunque es de elaboración propia está adaptado de las propuestas planteadas por Cabrera (1986); Tejedor *et al.* (1994); Jiménez Jiménez (1999b), y Pérez Carbonell (2000).

Volvemos a remarcar que en el caso del cuadro 1.7 se repite la circunstancia anteriormente señalada en el sentido de que tras las características indicadas en la dimensión «Investigación» se esconde la idea de «Investigación básica». En ese caso, las diferencias están muy claras. En el caso de la «Investigación aplicada», las diferencias entre investigación y evaluación en las dimensiones más metodológicas como «control de variables, generalización, metodología, diseño, aleatorización y criterios de validez» no están tan claras llegando en algún caso a no existir. Sin embargo, y a pesar de todo, este cuadro es válido para un primer acercamiento a este tema tan controvertido.

Cuadro 1.7. Diferencias entre investigación y evaluación

Dimensiones	Investigación	Evaluación
<i>Juicios de valor</i>	Se limitan a los implícitos en la selección del problema por el propio investigador.	Se extienden también al desarrollo y aplicación de los procedimientos de estudio.
<i>Hipótesis</i>	Se formulan hipótesis que guiarán la investigación.	Es difícil formular hipótesis y en ocasiones inadecuado.
<i>Replicación</i>	Posible.	No posible.
<i>Recogida de datos</i>	Determinada por el problema y las hipótesis.	Condicionada por la utilidad del proceso.
<i>Manipulación/control de las variables</i>	Posible.	De forma muy superficial cuando es posible.
<i>Interpretación de los datos</i>	Adaptado a la comunidad científica.	Valoración por los responsables del programa.
<i>Decisión</i>	La aceptación/rechazo de la hipótesis es responsabilidad del investigador.	La decisión sobre el futuro del programa depende del patrocinador.
<i>Informe</i>	Adaptado a las normas de la comunidad científica.	Adaptado a las distintas audiencias participantes en la evaluación.

Cuadro 1.7. (continuación)

Dimensiones	Investigación	Evaluación
<i>Finalidad</i>	Buscar nuevos conocimientos.	Ofrecer información para valorar un objeto y tomar decisiones sobre él.
<i>Papel del investigador/ evaluador</i>	El investigador establece y selecciona el problema o las cuestiones a investigar. Autónomo.	El evaluador está a expensas a las demandas de otros. Dependiente del gestor.
<i>Temporalidad</i>	Medio y largo plazo.	Inmediato, corto, medio.
<i>Costo</i>	Medio y alto.	Bajo y medio.
<i>Generalización</i>	Posible.	Limitada.
<i>Metodología</i>	Básicamente cuantitativa.	Diversas.
<i>Diseño</i>	Preestablecido.	Preestablecido, pero flexible.
<i>Aleatorización</i>	Básica.	No procede.
<i>Criterios de validez</i>	Interna, externa y de constructo.	Utilidad, credibilidad.

7. Evaluación y complementariedad metodológica

7.1. Naturaleza de la complementariedad metodológica

Para finalizar este primer capítulo, vamos a hacer una mención especial a la cuestión de la complementariedad metodológica. Es sabido que esta cuestión está relacionada con el debate, paradigmático, pero no es nuestra intención ahondar más en dicho debate, sino centrarnos en los aspectos metodológicos y su aplicación a la evaluación educativa. De todas formas, al hablar de complementariedad metodológica asumimos de alguna manera las ideas de compatibilidad y cooperación frente a las de incompatibilidad y enfrentamiento.

Existen diferentes razones que justifican la necesidad de complementar diferentes métodos. Entre éstas podemos mencionar las siguientes:

- La propia naturaleza compleja de los fenómenos y relaciones educativas objeto de evaluación.
- La evaluación puede tener objetivos variados.
- Hay diferentes modelos de evaluación.
- Cada método realiza sus aportaciones exclusivas que no las puede ofrecer otro distinto.
- Cada método tiene sus limitaciones. Por lo tanto, puede necesitar de las aportaciones de otro método para suplir sus carencias.
- En la evaluación hay una necesidad de recoger diferentes puntos de vista de las partes implicadas.
- Hay una conveniencia de acudir a modelos teóricos múltiples.

Sin embargo, la complementariedad metodológica no es fácil, existen dificultades para llevarla a la práctica. Algunas de esas dificultades son las siguientes (Pérez Juste, 2006):

- Hay limitaciones derivadas de la falta de capacitación de las personas que investigan y evalúan para afrontar con garantías diferentes metodologías. El trabajo en equipo podría solventar esta dificultad.
- Algunas dificultades derivan de las exigencias de tiempo en el sentido de que suele haber necesidad de terminar la evaluación en un plazo determinado que normalmente hace difícil adecuar, añadir o complementar diferentes técnicas y/o métodos.
- Las dificultades derivadas de las limitaciones de recursos económicos.

La complementariedad metodológica puede darse en diferentes fases del proceso de evaluación. Es en las fases de recogida y análisis de datos donde esta complementariedad es más evidente y en las que vamos a encontrar más referencias en la bibliografía, pero no podemos olvidar que esta complementariedad puede darse también en las fases de diseño de la evaluación y en la de acceso al campo o muestreo. En cualquiera de los casos, cualquiera de las formas de complementariedad debería permitir avances o mejoras en aspectos como la representatividad de muestras, profundización en el conocimiento y generalización de resultados.

Cuando hablamos de complementariedad metodológica básicamente hablamos de utilizar diferentes procedimientos en una evaluación. Desde esta concepción caben diferentes posibilidades de combinaciones de métodos. Así, se podrían combinar en una evaluación o investigación dos métodos cuantitativos. Otras opciones serían las derivadas de combinar distintos métodos cualitativos. Por último, cabría hablar de las posibilidades de combinar técnicas y métodos cuantitativos con cualitativos, y viceversa. Es en estas formas de complementariedad, entre métodos cuantitativos y cualitativos, en las que nos centraremos a partir de este momento, ya que es la forma de complementariedad más problemática al responder a paradigmas diferentes.

Otra cuestión que no podemos perder de vista es que una de las condiciones que debe darse para que la complementariedad metodológica sea verdaderamente efectiva es que tenga lugar en unas determinadas condiciones de coherencia, adaptación al objetivo a evaluar y con el mismo rigor que se exige a las evaluaciones que utilizan una sola orientación metodológica.

7.2. Estrategias de complementariedad

A partir de diferentes aportaciones (Greene, Caracelli y Grahan, 1989; Morgan, 1997, y Bericat, 1998) presentamos a continuación algunas de las estrategias de complementariedad de métodos diferentes que pueden aplicarse a estudios evaluativos.

1. *Complementación*. La estrategia denominada complementación («iniciación» para Greene, Caracelli y Grahan) es la pretensión de contar con dos imágenes distintas de la realidad que se pretende evaluar. La estrategia de complementación se da cuando en el marco de una misma evaluación se obtienen dos imágenes: una procedente de métodos de orientación cualitativa y otra de métodos de orientación cuantitativa. Es éste el caso en el que además de realizar una evaluación del impacto de un programa se pretende también una evaluación del propio proceso de implementación del programa. Para la primera de las evaluaciones se utilizaría un método cuantitativo, y para la segunda, cualitativo. En estos casos se suele hablar de estudios simultáneos. Cada perspectiva ilumina una dimensión diferente de la realidad, de manera que no existe o no se pretende el solapamiento de los métodos. El grado de integración metodológica es mínimo. El producto final es normalmente un informe con dos partes diferenciadas en las que se exponen los resultados obtenidos en la aplicación de los respectivos métodos. Si lo que se pretende es la máxima integración de los resultados, se puede llegar a síntesis interpretativas que integren resultados de cada método.

2. *Combinación*. La estrategia de combinación, también denominada «desarrollo» por Greene, Caracelli y Grahan se basa en la idea de que en el marco de una misma evaluación se trata de integrar subsidiariamente un método en el otro método con el objeto de fortalecer la validez de los resultados obtenidos por medio de este último método. En este proceso de combinación, se intentan compensar las posibles debilidades del segundo de los métodos mediante la incorporación de datos o información procedente del primer método. En la estrategia de combinación más que buscar una convergencia de resultados, se pretende una adecuada combinación metodológica, ya que los propósitos de cada uno de los métodos son diferentes. Cuando a partir de los resultados obtenidos en una evaluación en la que se han recogido datos de una muestra representativa se plantea la búsqueda de las interpretaciones de esos resultados a través de entrevistas en profundidad, nos encontramos ante un caso de combinación de métodos. Otra forma de utilizar la combinación es cuando se buscan generalizaciones a partir de estudios de casos en profundidad que han proporcionado información acerca de dimensiones a evaluar de forma cuantitativa. A diferencia de la estrategia de complementación; en este caso, los estudios son sucesivos y no simultáneos.

3. *Triangulación*. La triangulación se distingue de las dos anteriores porque en este caso, los métodos triangulados se encaminan a un mismo objetivo en la evaluación. Greene, Caracelli y Grahan denominan a esta estrategia con el nombre de «expansión». En esta estrategia se pretende obtener una visión más completa de la realidad, no a través de dos miradas, sino utilizando dife-

rentes orientaciones en el estudio de una única dimensión de la realidad. El componente integrador en la estrategia de triangulación se da en la posible divergencia o convergencia de los resultados. La triangulación tiene un amplio campo de aplicación, pudiéndose referir tanto a la aplicación de diferentes métodos como a la de diferentes técnicas de recogida de datos, como la resultante de la utilización de diferentes fuentes suministradoras de datos (alumnado, profesorado, personal directivo, familias), triangulación de diferentes momentos, teorías, etc. La triangulación refuerza la validez y la credibilidad de los resultados obtenidos al incrementarse la confianza en veracidad de la imagen y los resultados obtenidos. La propia naturaleza variada de la triangulación hace que en muchas ocasiones conviva junto a las otras estrategias de complementariedad.

Veamos en qué consisten algunas de las posibles triangulaciones Denzin y Lincoln (1994).

- *Triangulación entre métodos.* En esta triangulación se combinan dos o más métodos diferentes de evaluación aplicados a un mismo objeto. En la triangulación entre métodos hay que considerar algunos principios. He aquí algunos de estos principios básicos:
 - Según cuál es el problema de evaluación y la relevancia de cada uno de los métodos, uno de los métodos puede aparecer como dominante en la combinación y otro como secundario. Para ahondar más en esta cuestión, es muy interesante la reflexión realizada con Brannen (2005) en la que analiza las diferentes posibilidades que surgen en función de diferentes combinaciones.
 - Que la combinación pretenda contrarrestar las debilidades que puedan tener los métodos combinados para garantizar la validez y credibilidad de los resultados en sus diferentes variantes.
 - La evaluación no puede contemplarse de un modo estático, sino que se debe ser lo suficientemente flexible para aprender en el proceso, y, si se ve necesario, llegar a cambiar de método.
- *Triangulación de fuentes de datos.* Además de la utilización de técnicas diferentes para la recogida de datos se trata de ampliar el origen de los datos de la evaluación. Este origen se puede referir a diferentes lugares (triangulación espacial), momentos (triangulación temporal) u orígenes (triangulación de personas).
- *Triangulación de evaluadores.* La participación de diferentes evaluadores que permita contrastar sus diferentes observaciones y sus interpretaciones permite un aumento de la fiabilidad.

A continuación, y a modo de ejemplos de complementariedad metodológica, presentamos tres experiencias en las que han participado los propios autores:

7.3. Ejemplos ilustrativos

A. VITORIA-GASTEIZ COMO CIUDAD EDUCADORA

Evaluación de los Programas del Departamento Municipal de Educación (DME).

Para consultar el informe completo de la evaluación: http://www.sc.ehu.es/plwlumuj/ebaluzioa_pedagogia/praktikak/INFORME-vitoria.pdf

También puede revisarse un resumen en la siguiente publicación:

Lukas, J. F.; Santiago, K.; Lizasoain, L. y Joaristi, L. (2009). «Evaluación del impacto y pertinencia de programas educativos municipales». *Revista Española de Pedagogía*, n.º 243.

Durante el curso 2004-2005 se evaluaron los programas educativos que desarrolla en Departamento Municipal de Educación del Ayuntamiento de Vitoria-Gasteiz (DME):

- *Vitoria-Gasteiz como Espacio Educativo.*
 - Itinerarios Histórico-Artísticos.
 - Informativo Gasteiztxo.
- *Expresión y Arte.*
 - Expresión Musical.
 - Expresión Dramática.
- *Programa Materiales Curriculares.*

En estos programas participaban en aquel momento 30.000 estudiantes de la gran mayoría de los centros escolares de la ciudad.

La razón fundamental por la que se realizó esta evaluación fue la necesidad sentida por el DME de ver hasta qué punto los programas educativos que promueve habían impactado en los centros escolares y junto a ello hacer una reflexión acerca de la pertinencia o no de aquéllos.

El impacto se refiere a si tomar parte en los programas educativos tiene un efecto en el ámbito de los conocimientos, actitudes y valores del alumnado. Igualmente se plantearon cuestiones tales como si tenían efecto en la metodología del profesorado, se integraban en el currículo y organización del centro, se consideran interesantes, útiles, necesarios por los centros y si el profesorado estaba satisfecho con los programas.

Una segunda razón de la evaluación estaba relacionada con los cambios que se estaban dando en los últimos años en la sociedad en general y en la vitoriana en particular. Estos cambios llegan acompañados de nuevas necesidades que quizás deberían ser atendidas por el DME. En esta misma línea se puede

plantear que alguno de los programas haya podido quedar obsoleto y ya no responder a los tiempos actuales. Un objetivo de esta evaluación fue detectar aspectos relacionados con nuevas necesidades surgidas para que se tomaran las decisiones oportunas, y, en su caso, se priorizaran las necesidades a atender.

Objetivos de la evaluación. El objetivo general de esta evaluación fue doble:

1. Obtener evidencia empírica sobre los efectos esperados y no esperados de los programas desarrollados. En concreto, **evaluar el impacto** examinando la eficacia en el logro de los objetivos planteados en dichos programas.
2. Conocer tanto la pertinencia de los programas que se ofrecían como una relación de las **necesidades no cubiertas** por ellos.

Además de los objetivos generales, se plantearon los siguientes objetivos específicos:

- Conocer hasta qué punto el alumnado que participaba en los programas del DME evaluados aprovechaba esa participación con respecto a quienes no participaban, tanto en lo referente a los conocimientos como a las actitudes.
- Ver hasta qué punto los diferentes programas se integraban en el currículo y la metodología docente de los diferentes centros.
- Ver hasta qué punto colectivos como el profesorado, familias, centros de apoyo y recursos (denominados Berritzegunes en el País Vasco), concejales, etc., conocían los programas evaluados.
- Ver hasta qué punto la ciudadanía de Vitoria-Gasteiz, en general, conocía los programas evaluados.
- Conocer los aspectos positivos y negativos de cada uno de los programas a nivel de organización y funcionamiento.
- Ver hasta qué punto el profesorado participaba en los programas evaluados.
- Conocer cómo valoraba el profesorado los programas evaluados.
- Conocer cuáles eran las necesidades no cubiertas según la opinión del profesorado, colectivos ciudadanos, ayuntamiento...
- Conocer la valoración que los diferentes colectivos hacían de la labor del DME en lo relacionado con los programas educativos.
- Conocer cuáles eran las mejoras propuestas por los diferentes colectivos participantes en la evaluación.

Metodología. A fin de lograr los objetivos pretendidos en esta evaluación, se optó por una vía de complementariedad metodológica que permitiera aprovechar las distintas aportaciones tanto de la metodología cuantitativa como de la cualitativa.

En esta evaluación se aplicó la estrategia de *complementación*, que es aquella que se da cuando en el marco de una misma investigación se obtienen dos imágenes: una procedente de métodos cualitativos y otra de métodos cuantitativos. En este caso se aplicaron técnicas cuantitativas con vistas a la evaluación del impacto de los programas y cualitativas para la evaluación de la pertinencia. También podría hablarse de algún tipo de triangulación, ya que la utilización de técnicas diferentes ha permitido dar una visión completa de cada uno de los programas evaluados, al contrastar dos tipos de información diferente, la cualitativa y la cuantitativa en interpretación de resultados.

Con vistas al impacto, los datos recogidos provenían de la aplicación de pruebas objetivas y escalas tipo Likert, mientras que la información cualitativa se recogió, básicamente, en los grupos de discusión y entrevistas realizadas, junto con algunas sesiones de observación, aunque no se dio un seguimiento exhaustivo del proceso de implantación de los programas.

Resultados y conclusiones. Quisiéramos mencionar que aunque, como anteriormente hemos señalado, los datos referentes al impacto de los programas son básicamente cuantitativos, y los referentes a la pertinencia, cualitativos, en su fase de interpretación se hizo un intento de integrar los dos tipos de información en los casos en que fue posible. Así, resultados relacionados con aspectos como la integración, la organización y la valoración de los programas o el conocimiento de los programas por parte de la ciudadanía fueron fruto de ese intento de integración, lo cual indudablemente trajo consigo una visión más completa de la visión estudiada.

B. CONVIVENCIA Y CONFRONTACIÓN ENTRE IGUALES EN LOS CENTROS EDUCATIVOS

Para consultar el informe completo del estudio:

http://www.sc.ehu.es/plwlumuj/ebaluzioa_pedagogia/praktikak/convivencia-ararteko.pdf

También puede consultarse en:

Ararteko (2006): *Convivencia y conflictos en los centros escolares*. Vitoria-Gasteiz: Ararteko.

El estudio que se presenta a continuación tiene algunas diferencias notorias con respecto al ejemplo anterior. En primer lugar, habría que decir que este estudio, más que una evaluación, es una investigación. De todas formas para el objetivo que se pretende, que no es otro que ilustrar un ejemplo de complementariedad metodológica, sigue siendo igualmente válido. Otra diferencia es que en este caso no se hizo una evaluación de algún programa, sino que lo que se pretendió fue analizar el clima escolar en los centros de Se-

cundaria de la Comunidad Autónoma Vasca, con la intención de detectar aquellos factores que posibilitan o entorpecen la convivencia escolar.

Este estudio tuvo dos fases. En la primera de ellas participaron de 3.323 alumnas y alumnos de Educación Secundaria Obligatoria, 2.782 familias, 1.257 docentes y 80 directores de 80 centros elegidos a través de un muestreo aleatorio y representativo. En la segunda fase participaron diez centros seleccionados de los 80 participantes en la primera fase en base a algunos criterios relacionados con la convivencia como: nivel de problemática en el centro, existencia de programas para la convivencia, nivel de éxito ante problemas de convivencia, titularidad del centro y entorno rural/urbano.

Objetivos del estudio. Los objetivos de este estudio fueron los siguientes:

1. Conocer y contrastar los factores que docentes, padres y madres, y estudiantes de Educación Secundaria consideran que influyen en la convivencia escolar, y la importancia que le dan a cada uno.
2. Conocer y contrastar las causas a las que estos colectivos atribuyen el origen de los distintos tipos de conflictos que habitualmente se producen en los centros escolares, así como las soluciones que consideran más adecuadas.
3. Analizar los valores que vertebran las relaciones interpersonales de los alumnos y las alumnas de educación secundaria tanto con sus compañeros como con los adultos con los que conviven.
4. Conocer las actuaciones que los centros escolares de Educación Secundaria están llevando a cabo para la mejora de la convivencia en los centros y los procedimientos mediante los que están resolviendo los conflictos.
5. Conocer las actuaciones que los centros escolares están llevando a cabo en colaboración con los servicios sociales para la mejora de la convivencia y la prevención y resolución de conflictos, así como en su caso, por la vía judicial.
6. Conocer y contrastar la valoración que los distintos colectivos hacen de los programas de intervención que, en su caso, se estén llevando a cabo en el centro, y de las actuaciones de la Administración en este ámbito, e identificar las condiciones que los facilitan o dificultan.
7. Proponer algunas recomendaciones de intervención que puedan derivarse con fundamento de los resultados del estudio.
8. Favorecer el debate de la comunidad educativa sobre la convivencia en los centros, sus condicionantes y las estrategias más adecuadas para favorecer un clima integrador y respetuoso en el ámbito escolar.

Metodología. Como en el caso anterior, y a fin de lograr los objetivos propuestos, en éste también se optó por la complementariedad metodológica a través de la estrategia de **combinación** que, como ya hemos dicho anteriormente, es aquella que se da cuando en el marco de una misma investigación o

evaluación se trata de integrar un método en el otro método con el objeto de profundizar en las aportaciones y resultados obtenidos por medio de este último método. En el caso de dicha investigación, esta combinación se ha dado entre los cuestionarios cumplimentados por alumnado, profesorado y familias y las entrevistas realizadas a dichos agentes. En este caso, cabría añadir que también se ha dado cierto grado de triangulación de fuentes de datos, al contrastar las opiniones de los diferentes agentes participantes en el estudio.

El objetivo del estudio cuantitativo, que se realizó en la primera fase, fue el de conocer y contrastar las opiniones de los diferentes miembros de la comunidad educativa (alumnado, profesorado y familias) acerca de la convivencia en los centros escolares así como la incidencia de los diferentes tipos de conflictos que se producen en ellos, sus causas y sus formas de resolución. En esta fase se utilizaron cuestionarios tipo Likert.

El objetivo del estudio cualitativo, que se llevó a cabo a partir de los resultados del estudio cuantitativo, fue el de comprender con un mayor nivel de precisión y profundidad los factores que subyacen en las relaciones de convivencia en los centros escolares elegidos para este estudio, especialmente los relativos a las relaciones sociales de los alumnos y a sus valores. También se pretendía recoger la opinión del alumnado, profesorado y familias sobre los programas de convivencia que se hubieran podido desarrollar en el centro. En este estudio, las estrategias utilizadas fueron las entrevistas y los grupos de discusión.

Resultados y conclusiones. En cuanto a los resultados y conclusiones se puede decir que en el caso de los diez centros participantes en el estudio cualitativo la integración ha sido total, ya que a través de dos métodos diferentes se recogió información acerca del mismo fenómeno y con las mismas fuentes. Indudablemente, aunque estos resultados no pueden ser estadísticamente generalizables, no cabe ninguna duda de que dan una visión completa de lo que en aquel momento estaba ocurriendo en los centros de ESO de la Comunidad Autónoma Vasca.

C. EVALUACIÓN DEL PROGRAMA IKASYS

Para consultar el informe completo del estudio:

http://www.sc.ehu.es/plwlumuj/ebaluazioa_pedagogia/praktikak/EBALUAZIO%20TXOSTENA.pdf

Desde el curso 2007-2008, la Federación de Ikastolas (red de centros educativos privados concertados del País Vasco) de Gipuzkoa está implementando el denominado Proyecto IKASYS. El Proyecto IKASYS es un sistema

para que el alumnado aprenda con su ordenador. Es un conjunto que une tres elementos: herramientas informáticas (hardware), aplicaciones informáticas (software) y estrategias curriculares (contenidos). Tiene tres funciones: ejercitar al alumnado en diversos procedimientos de aprendizaje, ayudarle a memorizar y desarrollar su autonomía en base a sus capacidades. El objetivo principal del proyecto pasa por ofrecer al estudiante herramientas sólidas adaptadas a sus necesidades, para que entienda y pueda hacer frente a las situaciones-problema que le surgen en el aula y en el día a día, partiendo de situaciones simples (aprendizajes puntuales). A través de las actividades del Proyecto IKASYS, los alumnos se ejercitarán en aprendizajes puntuales que les ayudarán a adquirir distintas competencias y que están definidos como objetivos específicos. Por tanto, mediante estas actividades se quiere conseguir la sistematización de los aprendizajes educativos puntuales dirigidos al desarrollo de las competencias.

Durante el curso 2008-2009 se realizó la evaluación de este proyecto. En este proyecto además de evaluar el impacto del propio programa, se pretendió evaluar también el propio programa y su proceso de aplicación. Este proyecto se experimentó en dieciséis ikastolas del País Vasco en las cuales se ha hecho un seguimiento exhaustivo de su proceso de implementación. La experimentación se realizó en las aulas de 2.º, 4.º y 6.º de Educación Primaria y en las siguientes áreas de aprendizaje: Matemáticas, Conocimiento del Medio, Lengua Vasca, Lengua Castellana e Inglés.

Objetivos de la evaluación. Como ya hemos mencionado en las anteriores líneas, el objetivo de la evaluación fue doble:

1. Obtener evidencia empírica sobre el impacto del programa IKASYS. En concreto, se evaluó el impacto del programa en el aprendizaje de contenidos y competencias relacionados con las cinco áreas mencionadas. Además también se evaluó el impacto en la capacidad de autorregulación del alumnado ante ciertas tareas escolares.
2. Valorar el propio programa y proceso de su aplicación. Este objetivo se dirigió directamente a la mejora del programa y su aplicación.

Aunque indudablemente esta evaluación tuvo una vertiente sumativa en el sentido de que hubo una evaluación de los resultados o del impacto, podemos afirmar que básicamente estamos ante un tipo de «**evaluación formativa**», ya que tuvo por objetivo fundamental la mejora y el perfeccionamiento del programa a evaluar.

Metodología. Como en los casos anteriores, también se aplicó la complementariedad metodológica.

En esta evaluación podemos hablar de **complementación**, ya que se utilizaron métodos diferentes en las mismas situaciones o centros. Para evaluar el impacto del programa se llevó a cabo un «diseño cuasiexperimental con grupo de control no equivalente» en el que se aplicaron pruebas objetivas y cuestionarios cerrados. Sin embargo, con vistas a la evaluación del propio programa y su proceso de aplicación y para llegar a comprender con un mayor nivel de precisión y profundización todos los aspectos relacionados con los componentes del programa (hardware, software, contenidos curriculares) y los procedimientos relacionados con su aplicación se utilizaron estrategias de tipo cualitativo, como entrevistas, grupos de discusión y, sobre todo, observación participante a través de las que se ha pretendido recoger la opinión de alumnado, familias, profesorado y equipos directivos acerca del programa desarrollado en el centro.

Igualmente podemos hablar de **triangulación**, ya que haber utilizado métodos diferentes permitió obtener una visión más completa al contrastar en el caso de algunas dimensiones dos tipos de información diferente, la cualitativa y la cuantitativa. Esto permitió una interpretación más en profundidad de la información recogida a través de las diferencias estrategias e instrumentos utilizados.

Resultados y conclusiones. Si se observan los resultados obtenidos, se puede señalar que hay una serie de conclusiones acerca de algunas variables o dimensiones que como consecuencia de haberse analizado desde las dos perspectivas (cualitativa y cuantitativa) tienen un mayor grado de validez y credibilidad que si hubieran sido analizadas desde una única perspectiva. Algunas de estas dimensiones son las siguientes: formación previa del profesorado, software, motivación del alumnado, integración con el currículo del centro, etcétera. En estos casos, cuando los resultados cuantitativos y cualitativos han convergido, se han confirmado mutuamente y han apoyado las posteriores conclusiones. En otras ocasiones, se ha dado una compensación complementaria de los puntos débiles de cada método individual.

Preguntas y ejercicios

1. ¿Los paradigmas de la investigación educativa son similares a los paradigmas de la evaluación educativa? ¿Cuál es en este sentido la opinión de De Miguel?
2. ¿El concepto de evaluación de programas es sinónimo del concepto de investigación evaluativa?
3. ¿Cuáles son las semejanzas y las diferencias entre la investigación y la evaluación?
4. ¿Cuáles son las diferencias entre la complementación y la combinación?
5. ¿Qué aporta la triangulación y sus distintas variantes a una evaluación?

Notas

1. En el apartado siguiente se señalan de forma pormenorizada las similitudes y la diferencias entre evaluación, evaluación de programas, investigación evaluativa e investigación.
2. De Miguel (2000a, p. 292) define programa como un «plan, proyecto o servicio mediante el cual se diseña, organiza y pone en práctica un conjunto de acciones dirigidas a la consecución de unas metas».
3. Según Bisquerra (1989), los objetos pueden ser los métodos de enseñanza, los materiales del currículum, los programas, las organizaciones, los educadores y los estudiantes.
4. No obstante, cuando definimos el programa como «un plan sistemático diseñado por el educador como medio al servicio de las metas educativas» (Pérez Juste, 2000, p. 268) o con alguna definición similar, nos estamos refiriendo a un campo más específico coincidente casi con el de evaluación del currículum del cual hablaremos más específicamente en el capítulo 4.
5. Los conceptos de evaluación sumativa y formativa son explicados en el apartado 2.1 del capítulo 3.

2. Historia de la evaluación educativa

Desde su aparición y sobre todo durante el recientemente finalizado siglo xx, las transformaciones conceptuales y funcionales que se han producido en la evaluación educativa han sido profundas (Escudero Escorza, 2003).

Hasta hace unas pocas décadas, la evaluación educativa se limitaba a los aprendizajes de los alumnos; sin embargo, en la actualidad se emplea en los distintos objetos educativos (centros, profesores, programas, materiales, etc.). Además, los dirigentes políticos, los administradores y por supuesto las personas involucradas más directamente en la educación piden que se evalúen las distintas dimensiones educativas para conocer su efectividad y para poder optimizar las actuaciones. Como afirma Bordás Alsina (1995, p. 177) «en los países más avanzados, la evaluación de programas es ya una actividad formal, un instrumento político y de la administración pública; es un instrumento base para la búsqueda de la efectividad en el cambio, en la innovación educativa». Aunque los avances en los últimos 100 años han sido muchos, todavía la evaluación educativa tiene muchas lagunas y aspectos por resolver (Bordás Alsina, 1995). En esta misma línea, Worthen y Sanders (1987) afirman que la evaluación tiene grandes carencias en los fundamentos teóricos y un bajo nivel de experimentación en muchos modelos debido a que es un ámbito muy joven. Santos Guerra (1993) aduce que las deficiencias son las siguientes:

- Carencia de una adecuada base teórica.
- Ausencia en general de la evaluación en la evaluación de programas (metaevaluación).

Mateo (2000a, p. 17) va más allá cuando afirma que «no existe en realidad una teoría evaluativa, sino simplemente un conjunto de prácticas más o menos sistematizadas y su base conceptual está formada sustancialmente por un conglomerado de metáforas altamente elaboradas y que aparecen de manera recurrente y superpuesta a lo largo de los distintos intentos explicativos de las acciones evaluativas».

Ante esta situación, en este capítulo vamos a reseñar de manera resumida cuáles han sido los antecedentes de la evaluación educativa. Es decir, vamos a intentar resaltar desde los orígenes hasta la actualidad la evolución y el desarrollo que ha tenido la evaluación. No se trata, pues, de un análisis histórico propiamente dicho.

En la literatura reciente, suelen tacharse como clásicos tres planteamientos en la historia de la evaluación educativa (Escudero Escorza, 2003). En uno de ellos se señalan tres grandes épocas basándose en la figura central de Tyler. Así, a la época anterior se le denomina época de *precedentes* o *antecedentes*; época de *nacimiento* a la de Tyler y época de *desarrollo* a la posterior a Tyler.

Otros autores como Guba y Lincoln (1989) prefieren hablar de generaciones, distinguiendo cuatro. La primera de ellas, que cubre el primer tercio del siglo xx, es la denominada como *medición*. La segunda es la de *descripción*, y la tercera, la del *juicio o valoración*. La cuarta generación, que es en la que estaríamos, se apoya según los autores citados en el enfoque paradigmático constructivista y en las necesidades de los *stakeholders* (demandantes e implicados en la evaluación) como base para determinar la información que se necesita (Escudero Escorza, 2003).

Por último, la mayoría de los autores suelen establecer seis épocas que son las siguientes:

- Antecedentes remotos (2000 a. C.-1900)
- Etapa de la eficiencia y de los tests (1901-1929)
- Etapa de Tyler (1930-1945)
- Etapa de la inocencia (1946-1957)
- Etapa del realismo o de expansión (1958-1972)
- Etapa de la profesionalización (1973-2003)

1. Antecedentes remotos (2000 a. C.-1900)

Cuando se hace referencia a los orígenes de la evaluación educativa, ésta se confunde con la investigación educativa en general y con la medición en las Ciencias Humanas. Es decir, las primeras raíces o hechos históricos relevantes son igualmente válidos para el desarrollo de las distintas disciplinas. Es por ello por lo que algunos de los datos aquí aportados podrían ser igualmente válidos para la historia de la investigación educativa y/o la historia de la medición educativa.

El antecedente más lejano que se conoce con respecto a la evaluación educativa se sitúa alrededor del año 2000 a. C. en China. Al parecer, ya en aquella época se utilizaban los exámenes para acceder a la administración del Estado.

Trasladando el escenario al siglo v a. C. en Grecia, nos encontramos que Sócrates y algunos otros discípulos utilizan los cuestionarios evaluativos como parte de su metodología didáctica.

Muñiz (1998) menciona la publicación en 1599 por parte de los jesuitas de una serie de normas precisas para preparar exámenes escritos. Por otra parte, existen otros dos antecedentes del siglo xvi más cercanos a nuestro contexto. Por un lado, Juan de Huarte de San Juan publica en 1575 el libro *Examen de ingenios para las ciencias*, en donde se reconocen las preguntas clave de la evaluación de sujetos, y, por otro, Luis Vives, quien defiende la evaluación continua, identifica y describe distintos medios de recogida de información para la evaluación de los estudiantes (Perales Montolío, 2002b).

Sin embargo, estos hechos que se han reseñado, inconexos entre sí y sin ningún tipo de desarrollo posterior no son más que meras anécdotas en la evolución de la evaluación educativa. Debemos situarnos en el siglo xix para tener realmente avances en este campo.

La Revolución Industrial producida en esa época introdujo profundos cambios económicos y tecnológicos que a su vez transformaron las estructuras vigentes a nivel social. Entre ellos cabe señalar el incremento en la mejora de los programas educativos y sociales; sobre todo, en Estados Unidos y Gran Bretaña. Tal y como señala Cronbach (1980), los cambios sociales y económicos producidos tras la Revolución Industrial trajeron esta necesidad. Dos de las grandes innovaciones que se produjeron fueron la democratización de la educación y el comienzo de la inspección. Las necesidades industriales en cuanto a mano de obra de cada vez más personas que supieran leer, escribir y contar, junto con los ideales humanistas e ilustrados de la época, abrió la escuela al pueblo. De la misma manera y para controlar las condiciones de la escuela se creó el sistema de inspección.

Al mismo tiempo se dan una serie de avances, la mayoría en el campo de la medición en las Ciencias Humanas, que van a ser relevantes en el devenir de la evaluación educativa:

- Fechner utiliza por primera vez de manera formal la medición en el campo de las Ciencias Humanas en la segunda mitad del siglo xix, tanto en los laboratorios de psicología experimental como también en el estudio de las diferencias individuales, de donde surge la Psicometría (Mateo, 2000a).
- Wundt, junto con otros psicólogos alemanes, abre un laboratorio de psicología experimental en Leipzig y se inician en el desarrollo de la metodología de la observación y el registro de las conductas humanas.

- Galton, interesado en el estudio de las diferencias individuales entre seres humanos emparentados y no emparentados, realizó diversos intentos para medir la inteligencia y fue el primero que utilizó la estadística para analizar datos psicopedagógicos. Para la obtención de sus datos, creó en Londres un laboratorio antropométrico. Aunque algunos autores señalan que fue Galton el primero que utilizó el concepto de test mental (por ejemplo, Mateo, 2000a), la mayoría reconoce su autoría a Cattell en un artículo publicado en 1890. Éste continuó con la línea de los trabajos de Galton en el Columbia College de Estados Unidos realizando pruebas mentales y físicas. Pearson, alumno de Galton, desarrolló métodos estadísticos para el estudio de las diferencias individuales.

Por último, en lo que se refiere al campo de la medición, debemos señalar los trabajos que realizó Binet en Francia. Él logró construir la primera escala métrica para medir la inteligencia (1905) y acuñó el término de edad mental relacionado con la edad cronológica. Estos avances fueron muy importantes para el desarrollo de los tests psicológicos y educativos.

No obstante, a comienzos del siglo xx el desafío de la investigación educativa fue crear pruebas realmente escolares y de comprobación de conocimientos, ya que los tests construidos hasta entonces no respondían a las necesidades educativas (Mateo, 2000a).

Centrándonos en el tema de la evaluación educativa, señala Martínez Mediano (1996) que, aunque todavía estaban lejos de la evaluación educativa propiamente dicha, los inspectores externos de los centros educativos efectuaban una modalidad de evaluación educativa. Este tipo de evaluación del siglo xix era realizado en Estados Unidos y en algunos países de Europa; entre ellos, Gran Bretaña y España. En España, la Constitución de Cádiz de 1812 dio origen a la Dirección General de Estudios que fue la que creó la figura del inspector escolar.

En 1845, Mann realizó un estudio pionero en Estados Unidos. En esta evaluación pretendía comprobar si las escuelas de Boston educaban bien a sus alumnos y para ello aplicó pruebas de rendimiento cuyos datos fueron los indicadores principales para evaluar la efectividad del programa de instrucción. A raíz de este estudio se propuso la sustitución de los exámenes orales por los escritos. Sin embargo, es el trabajo que realizó Rice el que ha sido considerado como la primera evaluación formal. Esta evaluación, llevada a cabo entre 1895 y 1905 en un gran número de escuelas de Estados Unidos (con más de 33.000 alumnos involucrados) utilizó nuevamente las puntuaciones de los tests (en este caso de lectura y aritmética) como criterio de valoración de un programa de instrucción. Rice también contribuyó al desarrollo del enfoque evaluativo adversarial o judicial.

2. Etapa de la eficiencia y de los tests (1901-1929)

Los primeros años del siglo xx se caracterizaron por el desarrollo espectacular de la investigación educativa empírica, produciéndose una integración creciente entre los diseños de investigación y los procedimientos estadísticos (Aliaga, 2000; Perales Montolío, 2002b). Estos mismos autores señalan que a nivel epistemológico se dan dos circunstancias fundamentales:

- Por un lado, la difusión del pensamiento del Círculo de Viena, el neopositivismo, a partir de la década de 1920.
- Por otro, la Escuela de Chicago que plantea y aplica la metodología cualitativa, aunque su influencia en la comunidad científica educativa de la época fue muy limitada.

De la misma manera, la evaluación va a estar condicionada por diversos factores que confluyen en dicho momento (Escudero Escorza, 2003):

- La observación, la experimentación, los datos y los hechos son considerados como fuentes del conocimiento verdadero dentro de las corrientes filosóficas positivistas y empíricas.
- Los movimientos de investigación de las escuelas en Estados Unidos desarrollaron baterías de tests formadas por ítems objetivos que sustituyeron a las preguntas abiertas tipo ensayo como consecuencia de la exigencia del rigor científico y de la objetividad en la medición de la conducta humana y para combatir la subjetividad de los exámenes orales.
- Los trabajos de Darwin, Galton y Cattell también influyeron apoyando la medición de las características de los individuos y las diferencias entre ellos.
- La orientación métrica de la época también fue favorecida por el desarrollo de los métodos estadísticos.
- El desarrollo de la sociedad industrial potenció la necesidad de perfilar procedimientos de acreditación y selección de los estudiantes según sus conocimientos.

En esta etapa, para mejorar los servicios sociales de la comunidad se aboga por evaluaciones realizadas por expertos; es, pues, la etapa de la evaluación eficiente.

De todas maneras, esta época está caracterizada por el avance espectacular que tuvo el desarrollo de los tests psicológicos y por ende los tests escolares. A la primera publicación de la pionera escala métrica para medir la inteligencia le siguieron otras en 1908 y 1911 que se extendieron rápidamente al resto de los países europeos y a Estados Unidos, principalmente. Este primer test que era de aplicación individual pronto fue seguido por la construcción

de otros tests para medir la inteligencia y debido a las necesidades que en Estados Unidos tuvieron en la clasificación de millón y medio de soldados alistados para la Primera Guerra Mundial surgieron los tests colectivos de inteligencia. Pero estos avances continuaron en la medición de otras variables psicológicas. De la misma forma, todo el desarrollo de la medición de variables psicológicas fue trasladado a la construcción de pruebas de conocimientos y aptitudes escolares. Así fueron construyéndose los primeros tests normativos escolares que fueron utilizados como indicadores de calidad de las escuelas. Entre otros, caben citar los trabajos de Thorndike (que ha sido considerado un líder en el desarrollo y uso de los tests educativos), Ayres (tests de escritura y ortografía), Yates y Freeman (escalas de redacción), etc.

Toda la historia, hasta 1930, para algunos ha sido considerada bajo el epígrafe de la etapa pretayloriana (Stufflebeam y Shinkfield, 1987) y otros como Guba y Lincoln (1989) la han denominado como la primera generación o generación de la medida. Estos autores señalan las características de la evaluación educativa desde sus inicios hasta 1930 (Guba y Lincoln, 1982 y 1989; Tejada, 1999; Mateo, 2000a; Escudero Escorza, 2003):

- a) Evaluación y medición son conceptos intercambiables, uno y otro pueden utilizarse como sinónimos. El uso del término «evaluación» era poco usual y siempre estaba unido al de «medición».
- b) La evaluación y la medición están insertos dentro del paradigma cuantitativo característico de las Ciencias Naturales.
- c) La evaluación y la medición se centran en el análisis de las diferencias individuales siguiendo la tradición de Darwin seguida por Galton al principio y por otros después.
- d) La evaluación y la medición se orientan hacia medidas tipificadas tanto en lo que se refiere a las condiciones de aplicación y corrección como en lo que se refiere a la interpretación basada en grupo normativos.
- e) La evaluación y la medición ofrecen información acerca de los sujetos y apenas tenían relación con los programas escolares y con el desarrollo del currículum. No se planteaban la posible mejora del programa escolar.
- f) El papel del evaluador era técnico, como proveedor de instrumentos de medida.

3. Etapa de Tyler (1930-1945)

Previamente a la aparición de Tyler en el escenario americano de la evaluación, en Francia se inició en los años veinte del siglo pasado una corriente independiente conocida como docimología que, como señala Escudero Escorza (2003), supuso un primer acercamiento a la verdadera evaluación educativa.

En este movimiento se criticaba sobre todo la distancia que existía entre lo enseñado y los objetivos educativos y el que la evaluación se dejara en la interpretación subjetiva del profesor. Como señala el mencionado autor, este movimiento proponía como solución:

- I. Elaborar taxonomías para formular objetivos.
- II. Diversificar las fuentes de información.
- III. Unificar los criterios de corrección de los exámenes.
- IV. Revisar los juicios de valoración mediante procedimientos tales como la doble corrección o la media de distintos correctores.

La llegada de Tyler al mundo educativo en general y al campo de la evaluación en particular supuso una revolución en los conceptos establecidos en la época. Es por ello por lo que muchos le consideran el padre de la evaluación educativa. Tyler propuso un nuevo concepto del currículum y de la evaluación. De esta manera señaló que el currículum era «un conjunto de experiencias escolares planificadas, diseñadas e implantadas para ayudar a los estudiantes a lograr respuestas de conducta muy específicas» (Pérez Carbonell, 2000, p. 27). Por tanto, el currículum estará delimitado por las cuatro preguntas siguientes (Escudero Escorza, 2003):

- ¿Qué *objetivos* se desean conseguir?
- ¿Con qué *actividades* se pueden alcanzar?
- ¿Cómo pueden *organizarse* eficazmente estas experiencias?
- ¿Cómo se puede *comprobar* si se alcanzan los objetivos?

Por ello, un programa debía tener unos objetivos claros que serían los que guiarían la evaluación, puesto que ésta consistía en la emisión de un juicio de valor positivo o negativo acerca del logro de los mencionados objetivos tras el periodo de intervención. Por lo tanto, la buena evaluación necesitará (Escudero Escorza, 2003):

1. La propuesta clara de *objetivos*.
2. La determinación de las *situaciones* en las que se deben manifestar las conductas esperadas.
3. La elección de *instrumentos apropiados* de evaluación.
4. La *interpretación* de los resultados de las pruebas.
5. La determinación de la *fiabilidad y objetividad* de las medidas.

Tras la Gran Depresión y la pérdida de optimismo de la sociedad y la falta de recursos en Estados Unidos, surgen las ideas de John Dewey, quien basándose en la filosofía del pragmatismo y la psicología conductista cataliza el Movimiento de la Educación Progresista. Este movimiento hizo que en muchas escuelas se renovaran los currículos en la década de 1930; sin embargo,

las críticas no se hicieron esperar puesto que se argumentaba que los alumnos de estas escuelas estaban menos preparados que el resto de los alumnos al enfrentarse a la Educación Superior. Este problema dio origen al conocido «Estudio de los 8 años» (*Eight-year-Study*, publicado por Smith y Tyler en 1942) entre 1932 y 1940 que fue encargado a Tyler. En este estudio no se siguió la tradición experimental iniciada por Rice en la que comparaba el grupo experimental con el grupo control, sino que se comparaban los resultados obtenidos con los objetivos propuestos previamente. Es decir, el interés no se centra en la comparación de unos sujetos con otros, sino en comprobar el nivel que han alcanzado los sujetos evaluados en relación con las metas propuestas en el programa.

Guba y Lincoln (1989) señalan esta época como el comienzo de la segunda generación de la evaluación, mientras que Mateo (2000a) plantea el comienzo de la concepción dinámica de la evaluación frente a la concepción estática existente previamente. El propio autor señala que la mayor contribución de Tyler fue la insistencia de organizar el currículum en torno a los objetivos que debían lograrse. De esta manera, mediante los objetivos se facilitaba la planificación ya que eran una guía para los profesores. Nació, pues, la evaluación educativa propiamente dicha con una conceptualización dinámica que debía contribuir a una mejora continua del currículum y de la instrucción educacional (Mateo, 2000a).

Las características más importantes de este periodo son las siguientes (Tejada, 1999; Pérez Carbonell, 2000):

- a) La conceptualización estática de la evaluación es sustituida por la conceptualización dinámica, ya que se entiende la evaluación como un proceso para determinar el logro o no de las metas preestablecidas.
- b) Se diferencian claramente la evaluación y la medición como dos conceptos separados. La evaluación no es únicamente la aplicación de un instrumento de medida, sino que además necesita de un juicio de valor sobre la información recogida.
- c) La emisión del juicio de valor se basa en los objetivos específicos operativizados preestablecidos, convirtiéndose consecuentemente en los criterios de evaluación.
- d) Los objetivos definidos operativamente son la base en los que se asientan los procesos instruccional y evaluativo.
- e) El objetivo de la evaluación se traslada a constatar el cambio experimentado en los alumnos.

No obstante, el desarrollo de la evaluación durante las décadas de 1930 y 1940 se truncó debido a la Segunda Guerra Mundial, y no fue hasta la década de 1960 cuando vuelven a ser recogidas las ideas de Tyler sobre el currículum, la medición y la evaluación (Martínez Mediano, 1996).

4. Etapa de la inocencia (1946-1957)

Este periodo de posguerra se caracteriza por el aumento de la oferta educativa en Estados Unidos, aunque sin ninguna previsión sobre los recursos y sin ningún interés por comprobar la responsabilidad de los profesionales (Pérez Carbonell, 2000).

Sin embargo, el desarrollo de los tests estandarizados, las técnicas de medida y la fundamentación en el diseño de los tests fue muy grande. Paralelamente también se incrementó el uso de dichos tests (Martínez Mediano, 1996). Es en esta época (1947) cuando se crea el Educational Testing Service (ETS). También son de esta época las primeras recomendaciones técnicas para los tests psicológicos estandarizados (American Psychological Association, APA, 1954).

En esta época, Tyler publica varios libros sobre la evaluación educativa, aunque no es el único, y es su modelo el que domina el panorama educativo centrándose en las necesidades locales y en la consecución de las metas (Martínez Mediano, 1996).

Es también en este periodo cuando se extiende de la mano de Lindquist la aplicación del diseño experimental en la evaluación y surgen las taxonomías de los objetivos educativos; entre otras, las de Bloom y Krathwohl.

Las características propias de la época son las siguientes (Stufflebeam y Shinkfield, 1987; Tejada, 1999):

- a) Aunque la educación se extiende, el interés por la evaluación se retrae.
- b) Se da un retraso en el desarrollo de los aspectos técnicos de la evaluación.
- c) Gran desarrollo de tests estandarizados y de taxonomías de objetivos para su utilización a nivel nacional.
- d) Las evaluaciones dependen de las directrices escolares locales.

La etapa de la inocencia es, pues, una etapa que carece de una visión concienciada y renovada de la sociedad y, por tanto, de la educación, pero en la que se dan importantes avances técnicos tales como los conceptos de medición educativa, el desarrollo de los tests y de las taxonomías de objetivos por un lado, y, por otro, se da un amplio uso de los modelos de evaluación experimental y de Tyler (Perales Montolío, 2002b).

5. Etapa de la expansión (1958-1972)

Tal y como señala Mateo (2000a), probablemente la irrupción de una nueva conceptualización de la evaluación educativa fue provocada por las críticas desatadas a finales de la década de 1950 contra el sistema educativo de Estados Unidos. Si cuando en épocas anteriores el progreso económico y so-

cial era debido en gran medida al sistema educativo, también el fracaso debía ser achacado al mismo sistema. Este fracaso no es otro que el lanzamiento al espacio en 1957 por parte de la Unión Soviética del primer satélite (*Sputnik I*).

[...] llegó desde el espacio exterior el sonido de una alarma proveniente de un artefacto que los Estados Unidos no habían colocado allí. Había un satélite girando en torno a la tierra y no había sido construido por los Estados Unidos. Era evidente que algo no marchaba. Otra nación había sido capaz de realizar algo que los Estados Unidos no habían podido lograr. Algo o alguien debía ser culpable. A los ojos de muchos ciudadanos, la culpa era, sin lugar a dudas, del sistema escolar. Las escuelas no proporcionaban un número suficiente de científicos geniales. Por lo tanto, las escuelas tenían una falla.

(Popham, 1983, pp. 21-22)

A raíz de este suceso, el Gobierno Federal de los Estados Unidos decidió participar de manera más activa en el sistema educativo contrariamente a lo que había sucedido hasta entonces en donde cada estado era más o menos autónomo. En este sentido se aprueban dos leyes¹ a través de las cuales el Gobierno Federal decide invertir ingentes cantidades de dinero en el sistema educativo para reformar los currículos existentes. Paralelamente surgió un movimiento que exigía la petición de responsabilidades (*accountability*) a los docentes y a los directores tanto para el uso conveniente de los fondos federales como en la consecución de las metas educativas establecidas (Mateo, 2000a).

Como consecuencia de la ley de 1958 se asignaron también fondos para evaluar los programas renovados de matemáticas y ciencias. Los métodos que se utilizaron para evaluar estos programas fueron los siguientes (Stufflebeam y Shinkfield, 1987):

- a) El método de Tyler.
- b) Los nuevos tests estandarizados a nivel nacional.
- c) El método del criterio profesional se usó para valorar las propuestas y comprobar periódicamente los resultados.
- d) Muchos evaluadores evaluaron el currículum a través de pruebas concretas.

Sin embargo, muy pronto se dieron cuenta de que estas evaluaciones no ayudaban a los que debían elaborar los currículos y tampoco respondían sobre su efectividad. Por ello, se inició un periodo de reflexión y se publicaron ensayos teóricos que criticaban las prácticas evaluativas tradicionales y planteaban nuevas formas de evaluar.

De entre los artículos escritos en la época destaca el ya considerado como clásico de Cronbach de 1963 «Course improvement through evaluation». En este trabajo, Cronbach propone lo siguiente:

- Que no se hagan evaluaciones «*post hoc*» basadas en la comparación de puntuaciones medias, obtenidas mediante tests normativos, entre grupo experimental y control. La evaluación debe centrarse más en el análisis de las características estructurales del propio programa.
- Que se amplíe el ámbito conceptual de la evaluación para que pueda resultar de utilidad para los encargados de elaborar los currículos. De tal forma que su actividad se debe centrar alrededor de la toma de decisiones respecto a: perfeccionamiento del programa, decisiones sobre los estudiantes y regulación administrativa.
- Que deben abandonarse los tests estandarizados como únicos instrumentos de recogida de información incorporándose otras técnicas como los cuestionarios, las entrevistas y la observación.
- Que la evaluación debe comenzar mientras el programa está en marcha (evaluación del proceso) y no sólo cuando ha concluido (evaluación del producto). De esta forma se remarca el carácter regulador y optimizante de la evaluación.
- Que en la evaluación se incluyan:
 - Estudios de proceso, es decir, hechos que tienen lugar en el aula.
 - Medidas de rendimiento y actitudes para valorar los cambios producidos en los alumnos.
 - Estudios de seguimiento; es decir, el camino posterior seguido por los estudiantes que han participado en el programa.

Otro artículo que marcó el nuevo rumbo de la evaluación educativa es el que publicó Scriven en 1967 «The methodology of evaluation». En él pueden destacarse las ideas siguientes:

- Se señalan por primera vez las funciones que puede adoptar la evaluación: evaluación formativa (cuyo propósito es mejorar un programa cuando aún está en fase de desarrollo) y evaluación sumativa (cuyo propósito es comprobar la eficacia de un programa una vez que ya ha terminado).
- Critica el modelo de evaluación por objetivos de Tyler, puesto que antes que evaluar la posible consecución de las metas deberían analizarse si éstas son pertinentes, adecuadas o carentes de valor.
- No es suficiente señalar si un objetivo ha sido conseguido o no, habrá que señalar en qué medida ha sido logrado.
- En contraposición a Cronbach, defiende el carácter comparativo de la evaluación.
- Distingue entre evaluación intrínseca (valoración del objeto en sí mismo) y evaluación extrínseca (valoración de los efectos que produce el objeto).

También en esta época aparece otro artículo, que debido a su carácter más técnico, no ha tenido quizás tanta repercusión. Nos referimos al que publicó Glaser en 1963 «Instructional technology and the measurement of learning outcomes: Some questions». En este breve artículo el autor estableció las diferencias entre la medición normativa y la medición criterial (Mateo, 2000a).

A partir de esta época se van sucediendo propuestas evaluativas de diferentes autores basándose en los trabajos de Cronbach y Scriven. Comienza la evaluación en su acepción moderna y el inicio de lo que Guba y Lincoln (1989) han denominado tercera generación evaluativa. Durante este periodo, al que Mateo (2000a, p. 30) ha denominado eclosión de los modelos evaluativos, surgen la mayoría de los modelos de evaluación.

A pesar de las críticas recibidas, algunos de estos nuevos modelos se alinean en la evaluación por objetivos.

Por otro lado, los modelos que siguieron la línea de Cronbach intentaron asociar el proceso evaluativo con el de la toma de decisiones. Entre otros, los modelos CIPP (Contexto, Input, Proceso, Producto) de Stufflebeam, el CSE (*California State Evaluation*) de Alkin y el UTO (Unidades, Tratamientos, Operaciones) de Cronbach estarían dentro de este enfoque (Mateo, 2000a).

Este mismo autor señala que también en esta época surgen el modelo de análisis de sistemas cuyos máximos representantes son Rivlin, Rossi, Freeman, Cook y Wight.

Por último, Campbell y Stanley (1966) señalan el uso de los diseños que facilitan las decisiones que apoyan las mejoras sociales superando las limitaciones de los diseños experimentales.

No obstante, todos estos modelos de evaluación han sido muy criticados debido a que (Mateo 2000a, p. 31):

- Sólo se tienen en cuenta los propósitos del grupo en el poder.
- Los métodos utilizados son monistas y dogmáticos.
- Los comportamientos observables son reducidos a indicadores.
- Plantean una visión teórica de la realidad.

También son de esta época muchas de las asociaciones de evaluación aún vigentes en la actualidad. Así por ejemplo, el Phi Delta Kappa crea el National Study Committee on Evaluation; en 1967 se crea en la Universidad de California el Centro para el Estudio de la Evaluación (CSE); en 1972, el Instituto Nacional de Educación (NIE) centrado en investigaciones de programas de evaluación educativa; en 1964, la National Assessment of Educational Progress (NAEP).

Martínez Mediano (1996) señala que las características de esta época son las siguientes:

- a) Se evidenciaron los problemas originados por el uso de los métodos de evaluación basados en los objetivos.

- b) Se evidenció la necesidad de evaluar tanto los logros esperados como los no esperados.
- c) Se evidenció la importancia de emitir juicios sobre el mérito y el valor, destacándose los valores y las normas.
- d) Se desarrolló un vocabulario y una literatura evaluativas.
- e) Se contabilizaron más de 40 modelos de evaluación más o menos formalizados.

6. Etapa de la profesionalización (1973-actualidad)

Tal y como señala Pérez Carbonell (2000), las evaluaciones realizadas en las etapas anteriores fueron un fracaso debido en parte a la no existencia de medios que propiciasen el intercambio entre los evaluadores, a la falta de literatura sobre evaluación² y a la ausencia de medios para una adecuada formación en evaluación.

A partir de mediados de la década de 1970, la evaluación emergió como una profesión distinta de la investigación educativa. A ello contribuyeron diferentes aspectos:

1. La proliferación de manuales, tesauros, libros de texto, ensayos, etc. sobre temas de evaluación.
2. La consolidación de la evaluación como un ámbito específico de la investigación: la investigación evaluativa (Mateo, 2000a).
3. La aparición de revistas especializadas tales como *Educational Evaluation and Policy Analysis*, *Evaluation and Program Planning*, *Evaluation News*, *Evaluation and Policy Analysis*, *Evaluation Quarterly*, *Evaluation Review*, *New Direction for Program Evaluation*, *Studies in Evaluation*, *Assessment in Education*. También recopilaciones anuales como *Evaluation Studies Review Annuals*.
4. Las grandes universidades de Estados Unidos comienzan a impartir cursos de evaluación como especialización diferenciada de la metodología de la investigación (Universidades de Illinois, Stanford, Boston College, UCLA, Minnesota y la de Western Michigan, por ejemplo). (Stufflebeam y Shinkfield, 1987; Pérez Carbonell, 2000).
5. La evaluación centrada en los objetivos fue evolucionando hacia una evaluación orientada a la toma de decisiones.
6. Se crean asociaciones profesionales en torno al campo de la evaluación:

— El Joint Committee on Standards for Educational Evaluation en 1975, que tras agrupar a los profesionales más representativos de Estados Unidos publicó en 1981 las Normas para la evaluación de programas, proyectos y materiales educativos.

- En 1974 se crea The International Association for Educational Assessment (IAEA).
 - En 1976 se crea la Evaluation Research Society.
 - En 1986 se crea la American Evaluation Association.
7. La aparición de nuevos modelos de evaluación a raíz de la crisis entre paradigmas vivida en la investigación educativa. Estos nuevos modelos de corte cualitativo utilizan procedimientos antropológicos, periodísticos, etnometodológicos y naturalistas. Entre otros, podemos reseñar el modelo de evaluación iluminativa Parlett y Hamilton (1972), el modelo libre de objetivos de Scriven (1973), la evaluación respondente de Stake (1975), la evaluación democrática de McDonald (1977) y el modelo de crítica artística de Eisner (1985).

Siguiendo a Alvira Martín (1991, pp. 9-10) podemos afirmar que esta época se caracteriza por:

- a) Una expansión de la evaluación de programas a los países occidentales.
- b) La aparición de pluralidad de perspectivas, modelos y métodos.
- c) Un interés creciente en lograr que las evaluaciones sean útiles informando no sólo del logro de los objetivos, sino también de las partes del programa que funcionan y que no funcionan.
- d) Responder a muchas más preguntas superando la mera evaluación de resultados.

Guba y Lincoln (1989) pensaron hace casi dos décadas que el paradigma que resolvería la pluralidad existente sería el constructivista-respondente al integrar el paradigma constructivista con el modelo respondente de Stake. Sin embargo, todavía su profecía no se ha cumplido. Esos mismos autores denominaron a esta época la cuarta generación evaluativa cuyas características más sobresalientes son: sociopolítica; conjunto de colaboración; de enseñanza-aprendizaje; continua, recursiva y muy divergente; emergente; con resultados impredecibles; que crea realidad (Mateo, 2000a).

No obstante, Stufflebeam y Shinkfield (1987) y Pérez Carbonell (2000, p. 45) señalan algunos de los peligros que tiene la evaluación educativa en el momento actual:

- 1. La proliferación de modelos de evaluación ha traído un enriquecimiento léxico que también ha contribuido a que haya una mayor confusión.
- 2. Se da una mayor comunicación entre los evaluadores, pero también ha habido un aumento de «cháchara».
- 3. Surge el miedo a que los evaluadores funcionen como un club exclusivo y restringido debido al aumento de la profesionalización.

4. La fragilidad en la cooperación entre las organizaciones profesionales que se interesan por la evaluación educativa.
5. La polarización, fundamentalmente ideológica, entre la perspectiva positivista y la interpretativa.

7. Historia de la evaluación educativa en España

Como hemos visto, la historia social, económica, política y cultural, la historia de la ciencia y la historia de la evaluación se desarrollan paralelamente y lo mismo sucede en España (Perales Montolío, 2002b).

La necesidad de evaluar el rendimiento de los alumnos es tan antigua como el comienzo de la educación. No obstante, los únicos antecedentes específicos que podemos mencionar se reducen a las ya citadas normas para redactar exámenes escritos publicadas por los jesuitas en 1599 y las aportaciones de Huarte de San Juan y Luis Vives.

Aparte de los antecedentes que acabamos de mencionar, hasta 1970 el único precedente constatable en la evaluación educativa es la creación de la figura del inspector escolar en el siglo XIX. La función principal del inspector de aquella época y hasta hace pocas décadas ha sido fundamentalmente la de controlar y fiscalizar la actividad escolar.

7.1. La Ley General de Educación de 1970

Hasta la publicación de la Ley General de Educación de Villar Palasí de 1970, no aparece el término evaluación en el quehacer educativo del Estado. En esta ley, se recogen a nivel teórico los avances y el conocimiento acumulado en otros países (sobre todo, en Estados Unidos). Así se habla de evaluación del rendimiento de los alumnos, de evaluación continua y orientadora, de evaluación sumativa y formativa, etc. No obstante, como señala Tejada (1999), la mayoría de sus potencialidades no se explotaron en los 20 años de vigencia de la ley.

Este tipo de evaluación centrada exclusivamente en los alumnos, dentro de un programa curricular fuertemente centralizado y prescriptivo, y un profesorado en la enseñanza pública rigurosamente ideológico y de actitudes funcionariales hizo que los conceptos novedosos se convirtieran en los términos de una retórica oficial asumida, pero nunca utilizada en su profundidad.

(Sáez Brezmes y Carretero, 1995, pp. 132-133)

Como señalan los autores, la evaluación educativa fue recogida de Estados Unidos en su doble significado por el sistema educativo español:

- La inspección lo adapta a su función de control y supervisión.
- Los profesores la utilizan para calificar y promocionar a los alumnos.

Las críticas que se hicieron a esa ley en el ámbito de la evaluación han sido las siguientes:

- a) No se tienen en cuenta otros objetos susceptibles de ser evaluados (los profesores, los centros, los programas).
- b) El modelo propuesto es el de Tyler, con lo que se da excesiva importancia a los resultados finales olvidándose del proceso. Por ello, aunque en la ley se habla de la función formativa de la evaluación con vistas a utilizar ésta como fuente de mejora, no es posible ya que se evalúa sólo al final.
- c) Falta de preparación/formación del profesorado para llevar a cabo una evaluación de los alumnos formativa y continua. La evaluación continua fue entendida como una mayor aplicación de exámenes a lo largo del curso y la mejora-recuperación consistía en volver a aplicar otro examen a aquellos alumnos que en su día no lo superaron.

7.2. La reforma educativa de 1990

Con la llegada al poder del PSOE se inició la reforma de la educación. En este proceso se articularon debates en torno a diferentes documentos en los que de una manera más o menos extensa se incluía también la evaluación. Martínez Mediano (1996) señala los siguientes documentos: *Proyecto para la reforma educativa. Propuesta para debate* (1987); *Libro Blanco para la Reforma del Sistema Educativo* (1989); *Diseños Curriculares Base* (1989); *Ley Orgánica de Ordenación General del Sistema Educativo* (1990); *Órdenes Ministeriales de Educación y Ciencia* (27 de abril de 1992 y 12 de noviembre de 1992) que regulan los principios de evaluación del Sistema Educativo Español.

Al final, todo este proceso desembocó en la aprobación en 1990 de la Ley de Ordenación General del Sistema Educativo (LOGSE) y en 1995 de la Ley de Participación, Evaluación y Gobierno de los Centros Docentes (LOPEG). Esta ley, en su artículo 27, regula la realización de evaluaciones sistemáticas tanto de los alumnos como de los procesos educativos, profesores, centros y administración.

En el cuadro 2.1 se señalan las diferencias entre las características de la evaluación en la Ley General de Educación (LGE) y la LOGSE.

No obstante, en una reciente investigación (Lukas *et al.*, 2001) acerca de cómo han adoptado los profesores de Educación Secundaria las directrices de la LOGSE en relación con la evaluación de sus alumnos se llegó a las siguientes conclusiones:

- La mayoría del profesorado realiza una evaluación inicial/diagnóstica de los conocimientos del alumnado para ajustarlos a su programación; de la misma manera afirman que realizan una evaluación proce-

Cuadro 2.1. Análisis comparativo entre la LGE y la LOGSE
(Blanco, 1994, p. 23)

LGE (4-8-1970)	Aspecto	LOGSE (3-10-1990)
Sancionadora	Aplicación	Formativa
Orientador	Carácter	Regulador
Equipo de educadores	Agente	Profesores Alumnos Departamento de orientación Comunidad educativa
Alumno Centro	Objeto	Alumno Profesor Procesos Unidades didácticas Programaciones Proyecto curricular Funcionamiento del centro
Normativa	Marco	Criterial Individualizado
Conceptos Comportamientos	Contenidos	Conceptos Procedimientos Actitudes
Rendimientos Objetivos Saberes	Atención	Procesos Procedimientos Capacidades
Boletín de notas	Información	Informe de evaluación
Continua	Momento	Continua Integradora
Pruebas Intercambios orales	Instrumentos	Pruebas escritas Intercambios orales Observación Análisis de tareas Autoevaluación
	Documentación Planificación	Diseño de evaluación

sual/formativa en aras de mejorar el propio proceso y el rendimiento de los alumnos. Sin embargo, todavía dan muchísima importancia a las pruebas o exámenes del final de cada tema.

- En cuanto al normotipo utilizado para evaluar a los alumnos, aunque todavía algunos tienen en cuenta el nivel medio del grupo (evaluación normativa), cada vez son más los que tienen en cuenta el nivel de partida del alumno (evaluación idiográfica) y sobre todo se basan en los objetivos preestablecidos (evaluación criterial).
- En cuanto al agente evaluador, si bien en la LOGSE se señala la necesidad de utilizar la coevaluación y la autoevaluación, éstas todavía están muy lejos de ser empleadas. Sigue prevaleciendo la evaluación del alumno que realiza el profesor (heteroevaluación).
- En cuanto a la importancia que dan a los distintos tipos de contenidos, apreciaron que los conceptuales y procedimentales constituyen los que son tenidos en cuenta en mayor medida, mientras que las actitudes y valores y sobre todo los temas transversales no son prácticamente evaluados.

A partir de la LOGSE, se crea el Instituto Nacional de Evaluación y Calidad (INCE), cuyas funciones son las siguientes (Mayor Ruiz, 1998, p. 84; Pérez Carbonell, 2000, p. 51):

- I. Evaluar el grado mínimo de enseñanzas que se han adquirido.
- II. Evaluar las reformas generales.
- III. Elaborar un sistema de indicadores tendentes al estudio de la eficacia y de la eficiencia del sistema educativo.
- IV. Confeccionar sistemas de evaluación para las diferentes enseñanzas reguladas en la LOGSE.
- V. Coordinar la participación española en estudios de ámbito internacional.
- VI. Cooperar con todas las administraciones que tengan algún tipo de responsabilidad en materia educativa y evaluativa.
- VII. Proporcionar información tanto a las administraciones como a la sociedad en general del funcionamiento y logros del sistema educativo para facilitar la toma de decisiones.
- VIII. Publicar y difundir la información obtenida.

Otro hecho relevante fue la organización del Estado en comunidades autónomas con competencias en materia educativa y por ende en materia de la evaluación del sistema educativo. El gobierno vasco, entre otros gobiernos autonómicos como el canario y el valenciano por ejemplo, creó en 2001 el Instituto Vasco de Evaluación e Investigación Educativa (IVEI/ISEI) que sería el paralelo al INCE a nivel nacional. El Instituto Vasco de Evaluación e Investigación Educativa no universitaria tiene como finalidad realizar la eva-

luación general del sistema educativo no universitario, promover la investigación educativa en el ámbito no universitario y mantener un servicio de documentación y de recursos relacionados con el ámbito educativo. En el campo de la evaluación las funciones del IVEI/ISEI son las siguientes:

1. Diseñar y desarrollar los proyectos de evaluación del sistema educativo de la Comunidad Autónoma del País Vasco que se le encomienden.
2. Elaborar los indicadores de calidad propios del sistema educativo vasco.
3. Elaborar los instrumentos de evaluación de las necesidades educativas promovidos por la Viceconsejería.
4. Evaluar los procesos de experimentación de educación.
5. Diseñar en colaboración con el Servicio de Inspección educativa los procesos de evaluación que se lleven a cabo tanto desde el Instituto como desde la Inspección.
6. Elaborar los informes en los que se recoja la evaluación realizada y las propuestas de mejora que se deriven de ella para presentarlos al Consejo Rector.
7. Informar al Departamento de Educación, Universidades e Investigación y a la comunidad escolar, de acuerdo a las determinaciones del Consejo Rector, de los resultados de la evaluación.
8. Colaborar, a propuesta de la Dirección de Innovación Educativa, en la formación del profesorado.
9. Participar como servicio técnico de la Comunidad Autónoma del País Vasco en el Instituto Nacional de Calidad y Evaluación y colaborar con otras instituciones en los ámbitos de su competencia.
10. Colaborar con las Direcciones y Servicios del Departamento de Educación, Universidades e Investigación en relación con la evaluación en los ámbitos de su competencia.

Fuera de las administraciones públicas se ha creado la primera institución evaluativa. Se trata del Instituto de Evaluación y Asesoramiento Educativo (IDEA). Este instituto privado sin ánimo de lucro vinculado a la Fundación Santa María y con sede en Madrid tiene como objetivo la evaluación de los centros educativos no universitarios. Aunque este hecho resulte todavía anecdótico, puede que sea el origen de la creación de otros entes tal y como sucedió en Estados Unidos y Gran Bretaña.

7.3. La Ley de Calidad de 2002

En el año 2002 se aprobó la Ley de Calidad (LOCE) en la que se define la evaluación como la identificación de los errores y de los aciertos como un factor básico de calidad que constituye un instrumento para hacer políticas

educativas inteligentes y para incrementar su oportunidad y su adecuación a los cambios. Esta ley quiere introducir una serie de cambios en lo que se refiere a los propósitos de la evaluación y la utilización de ésta de una manera diferente a la propuesta en la LOGSE. En el segundo eje de medidas que propone la ley se declara claramente «en orientar más abiertamente el sistema educativo hacia los resultados, pues la consolidación de la cultura del esfuerzo y la mejora de la calidad están vinculadas a la intensificación de los procesos de evaluación de los alumnos, de los profesores, de los centros y del sistema en su conjunto, de modo que unos y otros puedan orientar convenientemente los procesos de mejora». Es decir, si en la LOGSE se acentuaba la importancia de los procesos, en esta ley se acentúa la importancia de los resultados. En este sentido, la evaluación idiográfica o personalizada que se aconsejaba en la LOGSE desaparece por completo dado que el único criterio utilizado para evaluar a los alumnos tanto en Educación Primaria como en Secundaria son los objetivos específicos establecidos en el currículo.

Los alumnos accederán al ciclo siguiente si han alcanzado los objetivos correspondientes establecidos en el currículo. Cuando un alumno no haya alcanzado los objetivos, podrá permanecer un curso más en el mismo ciclo.

Educación Primaria

La evaluación del aprendizaje de los alumnos en la Educación Secundaria Obligatoria será continua y diferenciada según las distintas asignaturas del currículo. [...] cuando el número de asignaturas no aprobadas sea superior a dos, el alumno deberá permanecer otro año en el mismo curso.

Educación Secundaria

En cuanto a la evaluación de otros ámbitos de la educación, en el Título VI se señala que la evaluación se extenderá a todo el ámbito educativo aplicándose tanto a los aprendizajes de los alumnos, los procesos educativos, la actividad del profesorado, los centros docentes, la inspección de educación y la propia administración educativa.

No obstante, el cambio de gobierno en 2004 supuso, primero, la paralización en parte de la aplicación de esta ley y, segundo, su derogación con la aprobación de la Ley Orgánica de Educación en 2006 que se presenta a continuación.

7.4. La Ley Orgánica de Educación de 2006

La Ley Orgánica 2/2006 de Educación (LOE) aprobada más recientemente tiene tres principios fundamentales sobre los que se basa.

El primero de ellos insiste en la necesidad de proporcionar una educación de calidad en todos los niveles educativos a todos los ciudadanos. Esto supo-

ne la mejora de los resultados generales y la reducción de las tasas de abandono y fracasos escolares, por un lado, y, por otro, la garantía de la igualdad efectiva de oportunidades prestando para ello los apoyos necesarios. Todo ello compatibilizando la calidad de la educación con la equidad en su reparto.

El segundo principio se basa en la necesidad de que todas las personas involucradas en la comunidad educativa ayuden en la consecución del primero de los principios; es decir, se apela al principio del esfuerzo compartido. En este sentido se alude tanto a la responsabilidad del alumnado como a la de las familias, los centros y el profesorado y la propia administración educativa.

El tercer principio implica la aceptación de los objetivos educativos compartidos por la Unión Europea. Estos objetivos hacen referencia a tres aspectos:

- Mejorar la calidad y la eficacia de los sistemas educativos y formativos.
- Facilitar el acceso generalizado a los sistemas de educación y de formación, facilitando para ello la igualdad de oportunidades y la cohesión social.
- Reforzar los lazos de la educación con la vida laboral, con la investigación y con la sociedad en general.

En definitiva, de acuerdo con los principios señalados, la LOE remarca que la educación se concibe como un aprendizaje permanente que se desarrolla a lo largo de toda la vida. Por ello, todas las personas han de tener la oportunidad de formarse tanto en el sistema educativo como fuera de él con el objetivo de adquirir, actualizar, completar y ampliar sus capacidades, conocimientos, habilidades, aptitudes y competencias para su desarrollo personal y profesional.

Ciñéndonos al tema de la evaluación, la LOE en su Título VI recoge sus fines, ámbitos y procedimientos. Establece una doble finalidad a la evaluación: la considera, por una parte, un elemento fundamental para la mejora de la educación y, por otra, como instrumento de aumento de transparencia del sistema educativo. Dicho en otros términos, señala las funciones formativa y sumativa de la evaluación que se desarrollarán más adelante.

En cuanto a los ámbitos en los que debe aplicarse se señalan los siguientes:

- Los procesos de aprendizaje de los alumnos.
- La actividad del profesorado.
- La función directiva.
- El funcionamiento de los centros docentes.
- La inspección.
- Las administraciones educativas.

Una vez más, la evaluación del sistema educativo es responsabilidad del Instituto de Evaluación (que a partir de la aprobación de la LOE ya no se denomina ni INCE como en la LOGSE ni INECSE como en la LOCE) que de-

berá trabajar de manera coordinada con el resto de institutos creados en las diferentes comunidades autónomas.

La rendición de cuentas del sistema educativo debe hacerse mediante la presentación de un informe anual al Parlamento. Este informe deberá sintetizar los resultados obtenidos en las evaluaciones de diagnóstico así como en el resto de pruebas generales que se realicen y los principales indicadores de educación.

En cuanto a la evaluación del alumnado se refiere, la LOE propone un marco diferenciado para cada etapa.

En Educación Primaria, la evaluación de los procesos de aprendizaje del alumnado será continua y global y tendrá en cuenta su progreso en el conjunto de las áreas.

En Educación Secundaria Obligatoria, la evaluación será también continua, pero diferenciada según las distintas materias del currículo. En el caso del Bachillerato, se menciona lo mismo, pero además, se incide en que el profesor de cada materia decidirá, al término del curso, si el alumno ha superado los objetivos de la misma. Al final del Bachillerato se mantiene la Prueba de Acceso a la Universidad.

7.4.1. La evaluación diagnóstica

El apartado más novedoso de la LOE en lo que a la evaluación se refiere es, quizás, el relativo a la evaluación diagnóstica.

Este tipo de evaluación no realizado hasta la aprobación de la Ley pretende obtener datos representativos tanto del alumnado y de los centros de las comunidades autónomas como del conjunto del Estado español. Esta evaluación, responsabilidad de la administración educativa, tiene como finalidad comprobar la adquisición de las competencias básicas del currículo.

Esta evaluación, que deberán realizarla todos los centros en 4.º de Educación Primaria y en 2.º de la Educación Secundaria Obligatoria, tiene un carácter formativo y orientador para los centros e informativo para las familias y para el conjunto de la comunidad educativa.

Para su realización se aplican de manera estandarizada un conjunto de pruebas y cuestionarios de contexto comunes para todo el alumnado de los mencionados cursos.

La elección de esos cursos, situados hacia la mitad de las etapas obligatorias, posibilita que tras los resultados obtenidos, cada centro detecte los problemas, adapte los currículos, adopte medidas de apoyo para los alumnos que lo precisen, etc.

La LOE no obliga a evaluar anualmente todas las competencias básicas, sino que cada comunidad autónoma debe establecer su propio plan de evaluación. En el cuadro 2.2, a modo de ejemplo, aparece el plan de evaluación aprobado por la Consejería de Educación del Gobierno Vasco.

Cuadro 2.2. Plan de Evaluación diagnóstica de la Consejería de Educación del Gobierno Vasco

Parte fija	Parte variable	
	Curso	Competencia básica
<i>Competencia en comunicación lingüística en euskara.</i>	2008-2009	• Competencia científica, tecnológica y de la salud.
	2009-2010	• Competencia social y ciudadana.
<i>Competencia en comunicación lingüística en lengua castellana.</i>	2010-2011	• Competencia en el tratamiento de la información y competencia digital.
<i>Competencia matemática</i>	2011-2012	• Competencia para aprender a aprender.
	2012-2013	• Competencia en cultura humanística y artística.
	2013-2014	• Competencia para la autonomía e iniciativa personal.
	2014-2015	• Competencia científica, tecnológica y de la salud.
	2015-2016	• Competencia social y ciudadana.

Como puede observarse, hay tres competencias que van a evaluarse todos los años mientras que el resto de las competencias irán evaluándose paulatinamente.

Para terminar con el capítulo, en el cuadro 2.3 se presentan los hechos más relevantes acaecidos con relación a la historia de la evaluación educativa.

Preguntas y ejercicios

1. Realiza un esquema en donde aparezcan las distintas etapas de la historia de la evaluación educativa, las características principales y los autores relevantes de cada una de ellas.
2. ¿Por qué en la etapa de Tyler la evaluación pasó de una conceptualización estática a una dinámica?
3. ¿Cuáles son los métodos utilizados para evaluar programas en la etapa de la expansión?
4. ¿Cuáles son las ideas principales de los artículos publicados por Cronbach (1963), Scriven (1967) y Glaser (1963)?

Cuadro 2.3. Ventajas y limitaciones de la evaluación interna y externa. (Tomado de Lukas *et al.*, 2000b, p. 44)

	Ventajas	Desventajas
<i>Evaluación interna</i>	<ul style="list-style-type: none"> • Legitimidad conferida al proceso. • Mejor comunicación. • Mayor capacidad de dinamizar la vida del centro. • Estrecha conexión con las decisiones de mejora que puedan adoptarse. • Mejor conocimiento del contexto. • Menos amenazador para los que van a ser evaluados. 	<ul style="list-style-type: none"> • Falta de tiempo. • La falta de distancia emocional y el compromiso de los agentes puede convertir el proceso evaluador en un factor de autojustificación. • Menor objetividad y credibilidad. • Ausencia de contraste externo.
<i>Evaluación externa</i>	<ul style="list-style-type: none"> • Mayor independencia. • Más tiempo dedicado a la tarea. • Mayor rigor debido a la especialización técnica. • Mayor credibilidad de sus conclusiones. • Imparcialidad del juicio. • Inclusión de una referencia externa en la valoración. 	<ul style="list-style-type: none"> • Dificultad de adentrarse en la vida del centro. • Mayor rechazo de la comunidad escolar. • Sesgo informativo.

5. ¿Cuáles son las características principales de la etapa de la profesionalización?
6. Teniendo en cuenta la LGE, la LOGSE, la LOCE y la LOE, señala las diferencias que se han dado en la evaluación educativa.
7. ¿Cuáles son los objetivos de la evaluación diagnóstica propuesta en la LOE?

Notas

1. En 1958, la *National Defense Education Act*, y en 1965, la *Elementary and Secondary Act*. Cuando se aprobó esta segunda ley, se acordó también un fondo para evaluar los programas; es decir, fue el primer mandato federal para una evaluación educativa que supuso a la postre la generalización para futuras acciones.
2. Hasta 1973 no aparece el término «evaluación de programas» indizado en la revista *Psychological Abstract*. El término «investigación evaluativa» aparece por primera vez en 1974 en el *Educational Index*.

3. Concepto, componentes y fases de la evaluación

1. Concepto

Si nos atenemos a la raíz etimológica del término «evaluación», podemos comprobar que éste procede del latín *valere*, que significa la acción de tasar, valorar, justipreciar, es decir, en definitiva el atribuir cierto valor a una cosa. En el castellano actual, este vocablo ha evolucionado hacia los términos «evaluación» y «valoración» que hoy en día son considerados como sinónimos (Perales Montolío, 2002b). No obstante, si nos atenemos al uso técnico de los términos, podemos observar que su origen es anglosajón.

El concepto de evaluación educativa ha estado ligado a las circunstancias históricas que ha sufrido la propia educación. Es por ello por lo que a lo largo de los últimos años el concepto de evaluación educativa ha variado en función de las distintas posiciones filosóficas, epistemológicas y metodológicas que han ido predominando (Tejada, 1999). Este mismo autor clasifica la variedad de definiciones de evaluación que nos podemos encontrar en cuatro categorías. Nosotros hemos mantenido esa misma clasificación, aunque hemos añadido una quinta categoría (definiciones que resaltan la importancia de la metodología) y hemos cambiado algunas definiciones de categoría:

1. *Definiciones que contienen especificaciones sobre los logros de los alumnos después de un proceso de instrucción*

Estas definiciones se corresponden con la conceptualización de Tyler sobre la evaluación educativa, es decir, una evaluación sumativa, basada en los ob-

jetivos preestablecidos y que sólo tiene en cuenta los logros de los alumnos, olvidándose los demás ámbitos educativos. Se subrayan los procesos de medición y normalmente hacen mención a un solo estudiante. Entre otras tenemos las definiciones siguientes¹:

- Tyler (1950, p. 69; 1970): «Proceso para determinar en qué medida los objetivos educativos han sido alcanzados».
- Lafourcade (1977, p. 21): «Etapa del proceso educacional que tiene por fin controlar de modo sistemático en qué medida se han logrado los resultados previstos en los objetivos que se hubieran especificado con antelación».
- Bloom *et al.* (1975, p. 23): «Reunión sistemática de evidencias a fin de determinar si en realidad se producen ciertos cambios en los alumnos y establecer el grado de cambio en cada estudiante».
- Provus (1971): «Comparación de las ejecuciones o comportamientos de los alumnos con ciertas normas para determinar si se debe llevar a cabo, se continúa o se concluye en el proceso de enseñanza».
- Mager (1975): «Acto de comparar una medida con un estándar y emitir un juicio basado en la comparación».
- Gronlund (1985, p. 5): «Proceso sistemático de recopilar, analizar e interpretar información para determinar el grado en que los alumnos han logrado los objetivos instructivos».

2. *Definiciones que hacen referencia a la determinación del mérito o valor*

Las definiciones pertenecientes a esta categoría son más recientes que las anteriores y no se reducen a la mera medición de los logros de los alumnos. Implica la emisión de un juicio valorativo sobre el mérito (capacidad del programa de lograr lo que pretende) y el valor (capacidad del programa para responder a las necesidades) del objeto evaluado. Además, el objeto evaluado no se ciñe únicamente a los logros de los alumnos, caben otros ámbitos distintos tales como un programa, un centro, etc.

- Suchman (1967): «Proceso de emitir juicios de valor».
- Scriven (1967): «Proceso por el que se determina el mérito o valor de alguna cosa».
- Stufflebeam y Shinkfield (1987): «Proceso sistemático planificado, dirigido y realizado con el fin de ayudar a un grupo de clientes a juzgar y/o perfeccionar el valor y/o mérito de algún objeto».
- Joint Committee (1988, p. 37; 1998, p. 31): «Investigación sistemática del valor o mérito de algún objeto».
- Mateo (1998): «Proceso de recogida de información orientado a la emisión de juicios de mérito o de valor respecto de algún sujeto, objeto o intervención con relevancia educativa» (citado en Mateo, 2000, pp. 21-22).

3. *Definiciones de evaluación que se basan en la información para la toma de decisiones*

Esta nueva conceptualización de la evaluación educativa surge en 1963 a partir de las ideas de Cronbach. Sin embargo, sus orígenes podemos buscarlos en el mundo de la empresa norteamericana utilizado con fines comerciales (López Mojarro, 1999). Epistemológicamente, el fundamento de la evaluación con fines empresariales y comerciales era claro. A saber, cuanto mejor se conozcan las circunstancias que rodean la fabricación de un producto, así como sus características, más eficaces serán las medidas que se puedan adoptar para mejorar la calidad y el costo de producción. Este mismo autor señala que se dieron dos avances importantes en este sentido:

- a) El conocimiento es fundamental para tomar las decisiones de mejora.
- b) El conocimiento debe basarse en una recogida y análisis científico de la información.

Estas mismas ideas fueron trasladadas a la evaluación educativa, primeramente a Estados Unidos pasando más tarde a Europa a través de Inglaterra. En España, el término aparece por vez primera de un modo generalizado con la promulgación de la Ley General de Educación de 1970 (Monedero Moya, 1998). A partir de ahí, en las definiciones de evaluación educativa pueden apreciarse las características siguientes (Tejada, 1999):

- Se incluyen múltiples objetos de evaluación.
- Incorporan nuevas funciones y momentos evaluativos (evaluación diagnóstica, formativa y sumativa).
- La evaluación es utilizada para mejorar la práctica.
- Se incluyen aspectos relativos al agente evaluador (evaluación interna-externa).

He aquí algunas definiciones que siguen esta línea:

- Cronbach (1963, p. 672): «Proceso de recopilación y utilización de la información para tomar decisiones».
- Tenbrinck (1984, p. 19): «Proceso de obtener información y usarla para formar juicios que a su vez se utilizarán en la toma de decisiones».
- Stufflebeam y Shinkfield (1987, p. 183): «Proceso de identificar, obtener y proporcionar información útil y descriptiva acerca del valor y el mérito de las metas, la planificación, la realización y el impacto de un objeto determinado, con el fin de servir de guía para la toma de decisiones, solucionar los problemas de responsabilidad y promover la comprensión de los fenómenos implicados».

- De Miguel (2000a, p. 290): «Proceso sistemático de búsqueda de evidencias para formular juicios de valor que orienten la toma de decisiones».

4. *Definiciones que subrayan la importancia de la metodología*

- Berk (1981, p. 4): «Es la aplicación del método científico para recoger la información fiable y válida necesaria para tomar decisiones sobre un programa educativo».
- Ruttman (1984): «Es la utilización del método científico con la intención de tomar decisiones sobre la implantación de un programa y medir sus resultados».
- García Llamas (1995, p. 46): «Proceso de identificación, recolección y tratamiento de datos para obtener una información que justifique una determinada decisión».

5. *Definiciones que intentan sintetizar y aglutinar los aspectos anteriores*

Estas últimas definiciones, más completas que las anteriores, intentan recoger la mayoría de los aspectos tratados en las secciones precedentes.

- Pérez Juste (1986, p. 31; 1995b, p. 23): «Acto de valorar una realidad, formando parte de un proceso cuyos momentos previos son los de fijación de las características de la realidad a valorar, y de recogida de información sobre las mismas, y cuyas etapas posteriores son la información y la toma de decisiones en función del juicio de valor emitido».
- Casanova (1995a, p. 55): «Recogida de información rigurosa y sistemática para obtener datos válidos y fiables acerca de una situación con objeto de formar y emitir un juicio de valor con respecto a ella. Estas valoraciones permitirán tomar las decisiones consecuentes en orden a corregir o mejorar la situación evaluada».
- Pérez Juste (1995a, p. 82; 2000, p. 269; 2002, p. 61): «La valoración, a partir de criterios y referencias preespecificados, de la información técnicamente diseñada y sistemáticamente recogida y organizada, sobre cuantos factores relevantes integran los procesos educativos para facilitar la toma de decisiones de mejora».
- Jiménez Jiménez (1999b, p. 21): «Proceso ordenado continuo y sistemático de recogida de información cuantitativa y cualitativa, que responde a ciertas exigencias (válida, creíble, dependiente, fiable, útil...) obtenida a través de ciertas técnicas e instrumentos, que tras ser cotejada o comparada con criterios establecidos nos permite emitir juicios de valor fundamentados que facilitan la toma de decisiones que afectan al objeto evaluado».

- Bordas Alsina (1999, p. 238): «Detectar cómo es una realidad educativa —alumno, grupo clase, centro formativo o educativo, planificación curricular, sistema educativo, etc.— sea desde una perspectiva cualitativa, cuantitativa o mixta, con el fin de tomar decisiones».

Como se ha podido comprobar en las definiciones precedentes, la evaluación es un término polisémico, de tal manera que a lo largo de la historia (e incluso actualmente) se ha definido (o se define) de maneras tan diferentes como medición de logros realizada mediante pruebas estandarizadas, como auditorías, como aplicación de pruebas objetivas, como análisis de procesos, como investigación en la acción, etc. (Santos Guerra, 1993). Desde un punto de vista metodológico hemos visto que aparecen aspectos propios de la construcción de instrumentos de medida, de medición y recogida de información, de investigación y de análisis de datos, pero, tal y como señala Mateo (2000a), la evaluación traspasa todas y cada una de esas actividades ya que constituye un universo en sí misma.

En términos precisos, debe entenderse que evaluar con intención formativa no es igual a medir ni a calificar, ni tan siquiera a corregir. Evaluar tampoco es clasificar ni es examinar ni aplicar tests. Paradójicamente, la evaluación tiene que ver con actividades de calificar, medir, corregir, clasificar, certificar, examinar, pasar tests, pero no se confunde con ellas. Comparten un campo semántico, pero se diferencian por los recursos que utilizan y los usos y fines a los que sirven. Son actividades que desempeñan un papel funcional e instrumental. De estas actividades artificiales no se aprende. Respecto a ellas, la evaluación las trasciende. Justo donde ellas no alcanzan, empieza la evaluación educativa.

(Álvarez Méndez, 2001, pp. 11-12)

Otro aspecto relevante en la evolución del concepto de evaluación es la relativa a la toma de decisiones ya que ésta no pertenece formalmente a la propia evaluación (Pérez Juste, 1995a). Sin embargo, hoy en día prácticamente todos los autores aceptan como necesario este aspecto para que la evaluación tenga una función pedagógica; es decir, tenga una función de mejora del objeto evaluado. Tal y como señala Mateo (2000a) lo que substantiva la evaluación (asignación de valor) no es lo que la legitima (mejora del objeto evaluado). De tal forma que una evaluación deberá asignar valoraciones del objeto evaluado (así asegura la validez de la evaluación) y también deberá derivar acciones para mejorar el objeto evaluado (así asegura la credibilidad).

En el cuadro 3.1 podemos observar cuáles eran las características de la evaluación en sus primeras conceptualizaciones y cuáles son las que se le asignan y se le exigen en la actualidad.

A partir de la revisión realizada, podríamos proponer nuestra propia definición de evaluación educativa, basada en las definiciones precedentes:

La evaluación es el proceso de identificación, recogida y análisis de información relevante —que podrá ser cuantitativa o cualitativa—, de manera sistemática, rigurosa, planificada,

Cuadro 3.1. Características de las conceptualizaciones tradicional y actual de la evaluación

Características de las conceptualizaciones anteriores	Criterios de diferenciación	Características de la conceptualización actual
Centrada en el alumno	Objeto	Holística
Control	Función	Perfeccionamiento, mejora y aprendizaje
Puntual	Momento	A lo largo del proceso
Cuantitativa	Metodología	Múltiple: cuantitativa y cualitativa
Informal	Sistematicidad	Formal y sistemática
Monolítica	Concepción	Pluralista
Estática	Concepción	Dinámica
Aséptica	Utilización de los valores	Valorativa
Aislada	Relación con el proceso	Integrada
Certificar	Objetivo	Instrumento de mejora
Trabajo técnico	Tipo de responsabilidad	Responsabilidad social y política
Evaluación profesional	Responsabilidad	Evaluación participativa
Dirección/administración	Poder	Participantes

dirigida, objetiva, creíble, fiable y válida para emitir juicios de valor basados en criterios y referencias preestablecidos para determinar el valor y el mérito del objeto educativo en cuestión a fin de tomar decisiones que ayuden a optimizar el mencionado objeto.

Vayamos por partes:

Cuando se señala que es un *proceso de identificación, recogida y análisis de información*, se hace referencia a la propia metodología de la investigación educativa; es decir, se alude a la necesidad de utilizar los procedimientos propios del método científico. En este sentido, la información ha de ser *relevante* (no toda información susceptible de ser recogida y analizada es necesaria y pertinente; ello dependerá de los objetivos y del diseño de la propia evaluación). La alusión al *objeto educativo* propone la posibilidad de poder evaluar cualquier hecho educativo.

Cuando señalamos que la información debe ser recogida y analizada de manera *sistemática, rigurosa, planificada y dirigida* hacemos referencia a que se trata de un proceso no improvisado, ni espontáneo; bien al contrario, será organizado, en donde se han temporalizado sus fases, se han previsto los recursos necesarios y se han construido o seleccionado los instrumentos. Se trata, pues, de analizar de una manera consciente, organizada y con una intencionalidad un aspecto de la realidad (Aguilar y Ander-Egg, 1994).

Además, se alude a una información recogida y analizada de forma *objetiva* (es decir, evitando subjetividades que sesguen la información o su posterior análisis) y que sea *creíble* (sobre todo en el caso de información de tipo cualitativo) además de *fiable* y *válida*. Hay que tener en cuenta que la calidad de la evaluación nunca podrá superar la calidad de los datos recogidos y analizados.

Al señalar específicamente *que podrá ser cuantitativa o cualitativa* se hace referencia a la complementariedad de los métodos tanto para la recogida como para el análisis de los datos.

«Emitir un juicio de valor» supone valorar la información recogida y analizada explicando su *mérito y/o valor* (bondad). Para ello, dispondremos de *criterios* establecidos de antemano que podrán ser idiográficos, normativos o criteriosales.

Por último, al añadir *a fin de tomar decisiones que ayuden a optimizar el objeto evaluado*, se subraya el carácter instrumental y no final de la evaluación, es decir, de utilidad para tomar decisiones que ayuden a promover procesos de cambio y mejora del objeto evaluado. De esta manera, se cierra el círculo que legitima la evaluación educativa.

2. Componentes

En la evaluación educativa se pueden distinguir una serie de componentes, dimensiones o aspectos en los que el evaluador deberá fijar su atención. Los más importantes son los siguientes:

2.1. La finalidad

Las funciones que se le asignan a la evaluación son diversas. En este sentido Ketele y Roegiers (1995) hablan de diferentes funciones de la evaluación en relación a distintos criterios. De esta manera emplean funciones tales como de regulación, predictiva, descriptiva, formativa, etc., como funciones relativas a las personas; preventiva, prospectiva, etc., como funciones relativas a un sistema; y, de desarrollo, de verificación, etc., como funciones relativas al conocimiento.

Sin embargo, aunque hacer diferenciaciones sutiles entre las distintas funciones es posible, aquí nos vamos a centrar en las dos funciones más importantes de la evaluación educativa. Hace ya algunos años, Scriven (1967) acuñó los términos de evaluación formativa y evaluación sumativa² para referirse a las dos funciones que puede desempeñar la evaluación educativa.

La función formativa de la evaluación se refiere a la utilización de ésta como medio para mejorar el objeto que está siendo evaluado. Esto significa que la evaluación debe llevarse a cabo de manera paralela a la actividad que se está evaluando. Tal y como se ha visto en el punto anterior, ésta es la fun-

ción por excelencia de la evaluación educativa. Hoy en día es prácticamente inconcebible hablar de evaluación educativa sin tener presente que su función es optimizar lo que se está evaluando.

Por el contrario, la función sumativa tiene un carácter de control, de certificación sobre los resultados o los productos obtenidos, dejándose la acción de mejora para posteriores intervenciones. Resulta, pues, adecuada para valorar productos o procesos que han finalizado, determinando el valor positivo o negativo de ellos.

Cuando Scriven planteó las dos funciones de la evaluación, consiguió ampliar conceptualmente el sentido y el significado de la evaluación ya que, previamente, únicamente la función sumativa de la evaluación era tomada en consideración. Hoy en día, sin embargo, hay autores que plantean estos dos tipos de evaluación como enfrentados cuando en realidad la evaluación es una acción única que dependiendo del momento o de la necesidad puede tener distintas funciones (Mateo, 2000a). Este mismo autor señala que lo que le confiere calidad no es el calificativo sino:

- La oportunidad del momento de aplicación.
- La adecuación con respecto al objetivo.
- La capacidad de utilizarlas complementariamente.

Aunque la evaluación formativa debe ser la dominante en toda acción evaluativa de carácter educativo, hay que tener en cuenta que desde un punto de vista holístico son las dos caras de la misma moneda (Casanova, 1995a; Mateo, 2000a). Además, es un error no tomar en consideración el valor formativo que puede tener la evaluación sumativa o la aportación de la evaluación formativa en la evaluación sumativa. Es por ello por lo que será difícil encontrar acciones evaluativas puramente formativas o puramente sumativas. En el cuadro 3.2 aparecen resumidas las características de la evaluación formativa y sumativa.

Cuadro 3.2. Características de la evaluación formativa y sumativa.
(Tomado de Casanova, 1995a, p. 66)

Evaluación Formativa	Evaluación Sumativa
Es aplicable a la evaluación de procesos.	Es aplicable a la evaluación de productos terminados.
Se debe incorporar al mismo proceso de funcionamiento como un elemento integrante de él.	Se sitúa puntualmente al final de un proceso, cuando éste se considera acabado.
Su finalidad es la mejora del proceso evaluado.	Su finalidad es determinar el grado en que se han alcanzado los objetivos previstos y valorar positiva o negativamente el producto evaluado.
Permite tomar medidas de carácter inmediato.	Permite tomar medidas a medio y largo plazo.

2.2. El objeto

Cualquier componente del sistema educativo es susceptible de ser evaluado, aunque habrá que contextualizar y operativizar el sistema, identificando sus aspectos más relevantes (Mateo, 2000a). Tradicionalmente, los objetos evaluativos son el alumnado, el profesorado, los programas, el centro educativo y el sistema educativo. En este sentido, el cuadro 3.3 pretende estructurar la realidad a evaluar en torno a diferentes unidades de análisis.

Cuadro 3.3. Unidades de análisis de los procesos de evaluación educativa. (Adaptado de Jornet, 2001, p. 64)

Categoría	Unidades evaluadas	
<i>Actores de la Educación</i>	Alumnado	
	Profesorado	
<i>Organizaciones</i>	Dentro del Sistema Educativo (Educación Formal)	Instituciones educativas: <ul style="list-style-type: none"> • Centros de Infantil, Primaria y Secundaria • Universidades
	Fuera del Sistema Educativo (Educación no Formal)	
<i>Elementos formales de materialización de la intervención educativa</i>	Diseños curriculares y proyectos educativos	
	Programas	Programas Educativos
		Programas de Educación para el trabajo
		Programas Socioeducativos
Materiales educativos		

Lógicamente ésta es una primera aproximación que debería seguir operativizándose en función de los objetivos de la evaluación. Es decir, cuando hablamos de la evaluación del alumnado, por ejemplo, deberemos especificar qué es lo que vamos a evaluar: ¿su rendimiento?, ¿su satisfacción?, etc. Si elegimos el rendimiento todavía deberemos especificar más: su rendimiento general, en alguna asignatura en concreto, qué tipos de contenidos interesan (conceptuales, actitudinales o procedimentales), etc.

2.3. La información

La información es la base de la evaluación y tal y como señala Pérez Juste (2000 y 2002); nunca ésta será mejor que aquélla. Por ello, la recogida de la información deberá ser técnicamente diseñada y sistemáticamente recogida y organizada.

En este punto debemos distinguir entre las fuentes de información, la instrumentalización y el tratamiento de la información.

Las fuentes de información serán las personas, documentos o entidades que van a proporcionar los datos necesarios para llevar a cabo la acción evaluativa.

La instrumentalización se refiere a las estrategias, técnicas y/o instrumentos que serán necesarios para recoger la información necesaria. Debido a la amplitud de este tema y a la gran cantidad de instrumentos, técnicas y estrategias, en el capítulo 7 («Procedimientos para la recogida de información») se trata en profundidad.

Por último, el tratamiento de la información hace referencia al análisis de la información recogida que es tratado en el capítulo 8 («Procedimientos para el análisis de la información»).

2.4. El agente

El agente evaluador se refiere a la persona o personas que van a llevar a cabo la evaluación. En este aspecto, la confusión terminológica es reseñable.

Casanova (1995a) cuando habla de la evaluación del alumnado distingue tres tipos de evaluación en función del agente evaluador: heteroevaluación, autoevaluación y coevaluación.

La heteroevaluación se da cuando una persona evalúa el trabajo realizado por otra persona. Es la situación típica que se da en todos los niveles educativos desde la Educación Infantil hasta la Universidad cuando es el profesor el que evalúa a su alumnado.

La autoevaluación ocurre cuando el evaluador y el objeto evaluado son coincidentes. Cuando el propio alumno, por ejemplo, evalúa el trabajo realizado por él mismo.

Por último, la coevaluación consiste en la evaluación mutua de un trabajo realizado entre varios. Por ejemplo, el profesor y los alumnos evalúan una unidad didáctica de forma conjunta.

Sin embargo, al referirnos a la evaluación de programas o de centros suele utilizarse otra terminología que a veces no es coincidente. Suelen distinguirse la evaluación externa, interna, autoevaluación y mixta, aunque no todos los autores los definen de la misma manera.

La evaluación externa se da cuando el agente o agentes evaluadores son externos al programa o centro evaluado. En esta definición coinciden la totalidad de los autores, aunque no sucede lo mismo con la evaluación interna. Para la mayoría sería sinónimo de autoevaluación, y la definen diciendo que consiste en que son los propios evaluados los que realizan la evaluación. Es decir, coincide el agente evaluador con el objeto de evaluación. Sin embargo, algunos autores, sobre todo en el ámbito de la evaluación de programas sociales, distinguen entre la evaluación interna y la autoevaluación. La autoevaluación

3. Concepto, componentes y fases de la evaluación

la definen tal y como hemos señalado, pero la evaluación interna sería cuando agentes de la propia entidad, pero no directamente involucrados en el programa en cuestión, son los encargados de llevar a cabo la evaluación. Por último, la evaluación mixta se refiere a una evaluación en la que participan de forma conjunta tanto agentes internos como agentes externos. Hoy en día, es la evaluación que se propone por parte de una gran mayoría de los autores.

En el cuadro 3.4 se presentan las ventajas y los inconvenientes de la evaluación interna y la evaluación externa (Lukas *et al.*, 2000b, p. 44).

Cuadro 3.4. Ventajas e inconvenientes de la evaluación interna y externa

	Ventajas	Inconvenientes
<i>Evaluación interna</i>	<ul style="list-style-type: none">• Legitimidad conferida al proceso.• Mejor comunicación.• Mayor capacidad de dinamizar la vida del centro.• Relación más estrecha con las decisiones de mejora que se puedan adoptar.• Mayor conocimiento y comprensión del objeto que se evalúa.• Mayor conocimiento del contexto.• Menor amenaza para los que van a ser evaluados y por lo tanto mayor aceptación.	<ul style="list-style-type: none">• Falta de tiempo.• La falta de distancia emocional y el compromiso de los agentes puede hacer que el proceso evaluador se convierta en una autojustificación.• Menor objetividad y credibilidad.• Falta de contraste externo.
<i>Evaluación externa</i>	<ul style="list-style-type: none">• Mayor tiempo de dedicación a la evaluación.• Mayor rigurosidad debida a la mayor especialización técnica.• Mayor credibilidad de los resultados obtenidos.• Imparcialidad de las opiniones.• Establecer una referencia externa en las valoraciones.• Mayor independencia.	<ul style="list-style-type: none">• Dificultades para entrar en la vida del centro.• Mayor rechazo por parte de la comunidad escolar.• Sesgo informativo.

2.5. El momento

El periodo o periodos en los que se llevará a cabo la acción evaluativa viene condicionado por el objeto evaluado y el propósito de la evaluación (Mateo, 2000a). Suelen distinguirse tres momentos: inicial, procesual y final.

La evaluación inicial es la que se aplica antes del comienzo de una acción evaluativa. En el caso de hablar del rendimiento de los alumnos y de los procesos de enseñanza-aprendizaje consistiría en detectar la situación inicial de los alumnos para poder proseguir en consecuencia con el proceso de enseñanza-aprendizaje.

Por otra parte, cuando nos referimos a evaluación de programas, la evaluación inicial o *ex ante*³ (también denominada de predecisión, de factibilidad o de pertinencia) se realiza antes de tomar la decisión de poner en marcha un programa, es decir, se hace en la fase de diseño del programa (Nirenberg, Brawerman y Ruiz, 2000). Esta evaluación comporta tres aspectos (Aguilar y Ander-Egg, 1994, p. 28):

1. *Pertinencia del programa con la realidad*: trata de establecer la coherencia entre las soluciones que se proponen y la realidad que se pretende modificar.
2. *Coherencia y congruencia internas*: es un análisis que atiende fundamentalmente a los aspectos formales. Interesa básicamente evaluar la organización/unidad del programa. Para ello, hay que examinar la viabilidad y coherencia de las relaciones entre:
 - Medios, recursos y objetivos.
 - Objetivos generales, objetivos específicos y metas.
 - Diagnóstico y proposiciones.
3. *Rentabilidad económica* de las diferentes acciones para alcanzar los objetivos propuestos.

Cuando se hace referencia a los procesos de enseñanza-aprendizaje, la evaluación procesual se concibe como la valoración continua del aprendizaje del alumno y la enseñanza del profesor mientras se está dando ese proceso. Necesariamente es de carácter formativo debido a que los datos que se van recogiendo durante el proceso sirven para tomar decisiones sobre la propia marcha. Al hablar de evaluación de programas, este tipo de evaluación adquiere distintas denominaciones tales como evaluación durante la ejecución del programa, de progreso o proceso, de gestión, concurrente, concomitante, intermedia, de operación, etc. (Aguilar y Ander-Egg, 1994; Nirenberg, Brawerman y Ruiz, 2000).

En la evaluación procesual, el objetivo es detectar en qué medida se va cumpliendo el programa en función de la propuesta inicial. La información que se va recogiendo de forma periódica sobre los procesos, las actividades y los productos es comparada con las metas preestablecidas, de tal manera que se van obteniendo los puntos fuertes y débiles del programa lo que va a contribuir a tomar las decisiones pertinentes en aras a reforzar, reorientar, corregir o subsanar el propio proceso, incrementando de esta manera las posibilidades de éxito.

Por último, la evaluación final es aquella que se realiza al finalizar un proceso, bien de enseñanza-aprendizaje, o bien al finalizar con un programa. En este caso, se pretende establecer la congruencia entre los objetivos planteados y los logros alcanzados, aunque también deben tenerse en cuenta los efectos secundarios o los resultados no esperados. También se deberá tratar de señalar la relación entre los resultados obtenidos y la influencia del programa en la consecución de dichos resultados.

En este tipo de evaluación se subrayan por lo tanto las cuestiones relativas a la eficacia, eficiencia y efectividad del programa. Es decir, se quiere llegar a conclusiones para decidir la continuación o no del programa, y en caso afirmativo señalar los aspectos que pueden ser optimizados en futuras implementaciones.

La evaluación final también suele denominarse como evaluación *ex post*, de fin de proyecto, a posteriori y terminal. La mayoría de los autores coincide en que este tipo de evaluación debe realizarse inmediatamente después de finalizar el programa y utilizan otro tipo de denominación para señalar la evaluación que se realiza al cabo de un tiempo (quizás meses o incluso años) cuando el programa ha alcanzado su pleno desarrollo. En estos casos se habla de evaluación del impacto o evaluación *post decisión*.

2.6. La valoración

La valoración de la información recogida supone la emisión de un juicio de valor, pero para ello necesitamos de algún referente con el que comparar. Este referente es conocido como *criterio de evaluación*. Carrión Carranza (2001, p. 138) define criterio de evaluación como «la expresión de algún atributo que debe tener el objeto de análisis para ser considerado de calidad, ya sea un programa o currículum, un proyecto institucional o la institución escolar misma, los subsistemas o las formas de relación de la comunidad escolar». Por su parte, Husen y Posthewaite (1985, p. 1735) definen criterio como «un aspecto o dimensión de la calidad a ser evaluada, que es ponderada con un estándar arbitrario o nivel de esta calidad, como una base para evaluarla». Por lo tanto, se requieren al menos tres componentes (Mayor Ruiz, 1998; Cabrera, 2000a):

1. *La selección de criterios*. Estos criterios serán la base en la que nos apoyaremos para decidir el éxito o no del programa. Lógicamente, los criterios seleccionados variarán en función de los objetivos de la evaluación.
2. *El desarrollo de los indicadores*. Los indicadores son realizaciones observables de los criterios, muestran, pues, los aspectos específicos de la realidad sobre los que habrá que recoger información.
3. *El establecimiento de estándares*. Los estándares servirán como términos de comparación que delimitarán la consecución o no del éxito y el grado de consecución en su caso.

Sin el establecimiento de criterios y referencias o estándares de evaluación, ésta no se produce. Es decir, el mero hecho de recoger información no es suficiente para que haya una evaluación. Podremos conocer, describir o indagar una situación, pero no evaluarla. Para que se dé la evaluación se deben emitir juicios de valor sobre el objeto evaluado. Para ello, la información recogida sobre los criterios preestablecidos será valorada teniendo en cuenta los referentes señalados como indicadores de éxito. De todos es conocido que ésta es una tarea ardua⁴; sin embargo, nunca debe ser obviada.

2.6.1. Los criterios

Autores como Le Boterf (1985), Pain (1992, 1993) y Bramley (1996) han presentado listados de criterios aplicables a situaciones de evaluación de programas de formación. Basados en esos listados, Cabrera (2000a, pp. 31-32) publicó su propia propuesta también aplicada al campo de la formación. Esta propuesta puede ser válida con alguna modificación para la evaluación educativa en general. He aquí, pues, el listado de criterios específicos más o menos aceptados en el campo de la evaluación educativa:

1. *La idoneidad o suficiencia*

Alude a la potencialidad que tiene un programa o, más específicamente, sus componentes para contribuir a alcanzar los objetivos y metas programados. De esta manera, se hace referencia por ejemplo a la idoneidad de los materiales utilizados, la suficiencia de los recursos humanos, etc.

2. *La pertinencia o propiedad*

Alude a la adecuación y capacidad de un programa, o alguno de sus componentes, para responder a las necesidades y las demandas o para resolver la situación. Dicho de otra forma, son criterios que valoran la relevancia y utilidad del programa para responder a las necesidades que lo originaron.

3. *La eficiencia o rendimiento*

Alude a la relación entre los productos o resultados esperados de un programa y los costos, esfuerzos o insumos empleados. Se establece hasta qué punto los gastos son justificables con los resultados obtenidos.

4. *La eficacia y la efectividad*

Los dos criterios se refieren a la capacidad de un programa para alcanzar los resultados esperados; es decir, objetivos programados. Hay autores que no hacen distinciones entre los dos términos; sin embargo, es cada vez más aceptada la no utilización de ambos como sinónimos y establecer una clara diferencia entre ellos. De esta manera, Pineault y Daveluy (1987) hablan de eficacia potencial y eficacia real o efectividad como dos conceptos diferenciados. Eficacia potencial

sería la capacidad de un programa para lograr los objetivos y metas planteados en condiciones ideales. Por otro lado, la eficacia real o efectividad sería la comparación de los resultados obtenidos con los resultados esperados; es decir, alude a las circunstancias reales de aplicación del programa.

5. *La coherencia o congruencia*

Alude a la coherencia de los propios elementos internos del programa que se está evaluando (adecuación de los objetivos, coherencia de los métodos utilizados para alcanzar dichos objetivos, adecuación de las actividades programadas para alcanzar los objetivos, etc.) y elementos externos de la organización (recursos disponibles, etc.).

6. *La conformidad o legalidad*

En este campo, los criterios se usan para determinar hasta qué punto se ha cumplido la normativa vigente, tanto externa (normativa legal, por ejemplo) como interna (reglamento de régimen interior o acuerdos establecidos, por ejemplo).

7. *La vigencia o actualidad*

Los criterios relativos a esta área aluden a la actualidad científica o humanística; es decir, a la presencia de los conocimientos, procesos instrumentales o tecnología vigentes en los contenidos y los procesos de enseñanza-aprendizaje.

8. *La oportunidad o sincronía*

Los criterios relativos a la oportunidad hacen referencia a la adecuación de las acciones emprendidas a lo largo del tiempo. Es decir, el criterio de calidad hará alusión a que las acciones no se han ejecutado ni prematura, ni tardíamente y además se han llevado a cabo con el ritmo adecuado.

9. *La aceptación o satisfacción*

Estos criterios hacen referencia al nivel de agrado, aceptación y/o satisfacción por parte de los usuarios tanto directos (por ejemplo, el alumnado) como indirectos (por ejemplo, las familias) del programa o institución en cuestión.

10. *El impacto*

Los criterios relativos al impacto aluden a los resultados que produce el programa a largo plazo.

2.6.2. Los indicadores

Mateo (2000a, p. 169) define los indicadores como «datos (cuantitativos o cualitativos) recogidos de manera sistemática, que informan respecto a los re-

cursos impartidos, los procesos realizados o las metas conseguidas, en relación con los objetivos de calidad que se había propuesto...».

Lázaro señala las condiciones exigibles a un indicador en el cuadro 3.5.

Cuadro 3.5. Condiciones exigibles a un indicador (Lázaro, 1991)

Requisitos del CONSTRUCTO

1. El indicador recoge y hace referencia a alguna propiedad del objeto (significatividad).
 2. La propiedad referida es de tipo esencial (relevancia).
 3. Los datos recogidos en el indicador están en congruencia con los fines y objetivos de la evaluación pretendida (teleologicidad).
 4. Los datos recogidos por el indicador tienden a ser permanentes (caducidad).
 5. Con los datos recogidos por el indicador se pueden realizar influencias y predicciones (utilidad).
 6. El indicador ofrece datos integrables en un sistema de indicadores (coherencia).
-

Requisitos de MEDIDA

7. Los datos que reclama el indicador son observables (observación).
 8. Los datos que reclama el indicador pueden ser observados por diferentes evaluadores simultáneamente o en situaciones distintas (replicación).
 9. Los medios que se utilizan para recoger los datos que reclama el indicador son adecuados (aplicabilidad).
 10. Las fuentes que facilitan la información solicitada por el indicador son accesibles (accesibilidad).
 11. Los indicadores poseen una forma de anotar y valorar la información recogida (codificación).
 12. Los indicadores tienen un peso específico dentro de un sistema de indicadores (ponderación).
 13. Los datos obtenidos por los indicadores pueden interpretarse numéricamente (expresión cuantitativa).
 14. Los datos obtenidos por los indicadores pueden relacionarse numéricamente con los de otros indicadores (obtención de índices).
-

Requisitos FORMALES

15. Cada indicador solicita información de una sola cosa (universidad).
 16. La información solicitada se puede recoger en una situación definida (manifestación de precisión).
 17. El indicador está expresado con claridad semántica sin ambigüedad (brevedad expresiva).
 18. La información reclamada se recoge en un mismo tipo de manifestación (concreción).
 19. El enunciado será directo y descriptivo (claridad).
 20. El indicador se relaciona causalmente con otros (dependencia).
 21. Se puede desglosar y especificar en ítems (formalización).
-

2.6.3. Los estándares

Una vez establecidos los criterios, son necesarios los estándares para discriminar las realizaciones aceptables de las no aceptables. Los referentes, normas, estándares o normotipos son términos utilizados por diferentes autores como sinónimos.

Tradicionalmente se han señalado tres tipos de estándares en función de los cuales se suele hablar de tres tipos de evaluación, aunque como en ocasiones anteriores, los distintos autores no se ponen de acuerdo con la terminología utilizada. De esta forma, Casanova (1995a) al referirse a la evaluación de los alumnos señala que dependiendo de si el normotipo es interno o externo al sujeto evaluado podemos distinguir dos tipos de evaluación:

1. *Evaluación nomotética*, si el referente es externo al propio sujeto. Ahora bien, este normotipo externo puede ser definido de dos formas diferentes:
 - a) Teniendo en cuenta la realización de la población o grupo al que pertenece el sujeto. En este caso, Casanova (1995a) habla de *evaluación nomotética normativa*. De tal manera que teniendo en cuenta este estándar, un sujeto será evaluado positiva o negativamente en función de la posición relativa que ocupa dentro de su grupo de referencia.
 - b) Teniendo en cuenta la realización absoluta del sujeto, es decir, se establecen de antemano cuáles son los estándares que debe alcanzar el sujeto. En este caso, habla de *evaluación nomotética criterial*.
2. *Evaluación idiográfica*, si el referente es interno, es decir, se toma al propio sujeto como referencia. Sus anteriores logros son los referentes que se tomarán en cuenta a la hora de evaluar los logros actuales. Algunos autores la denominan *evaluación personalizada*.

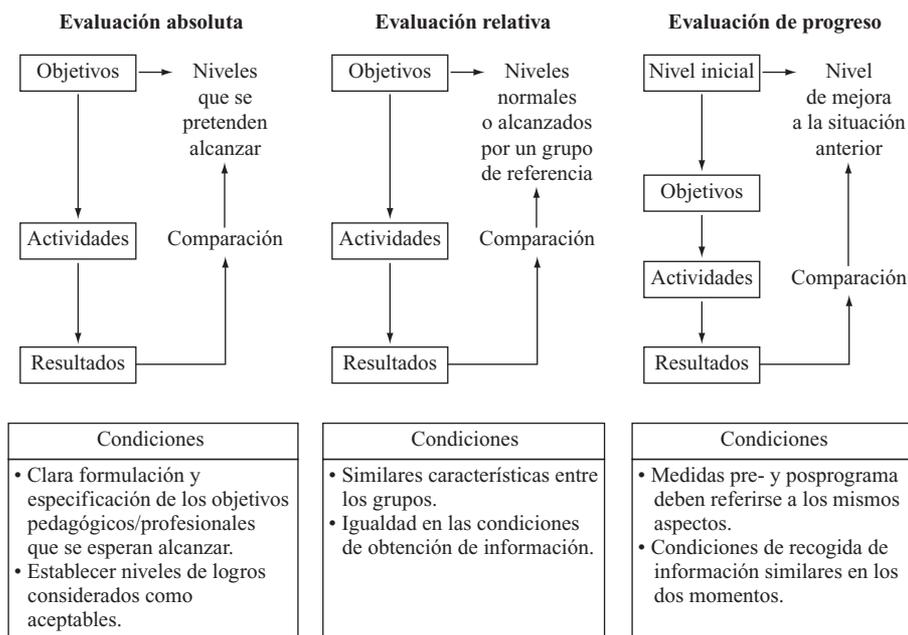
Al generalizar esta situación a cualquier objeto susceptible de ser evaluado, nos volvemos a encontrar con los mismos tres referentes, aunque con distintos nombres según los diferentes autores. A saber:

1. Evaluación relativa (Cabrera, 2000a) o evaluación normativa, en la que establecemos la posición relativa del objeto evaluado en función de su grupo de referencia.
2. Evaluación criterial, también denominada evaluación absoluta (Cabrera, 2000a) o evaluación basada en la autoridad (Jiménez Jiménez, 1999b), en la que el objeto evaluado se compara con los estándares de excelencia previamente establecidos.

- Evaluación personalizada, también denominada evaluación idiográfica o idiosincrásica (Casanova, 1995 y Pérez Juste, 2000 y 2002, respectivamente) o evaluación de progreso (Cabrera, 2000a), en la que la comparación se establece con el propio objeto evaluado. Es decir, se compara la situación inicial del objeto con la situación actual del mismo.

En la figura 3.1 se presenta un resumen de los tipos de evaluación según el referente.

Figura 3.1. Tipos de evaluación según referentes (Cabrera, 2000a, p. 34)



2.7. El informe

El informe es un componente relevante de la evaluación educativa puesto que si se quiere que las conclusiones obtenidas sean utilizadas, éstas deben ser comunicadas de forma adecuada. Su propósito es describir los resultados favorables y desfavorables del objeto evaluado a fin de que éste pueda ser mejorado. Para ello, es necesario que se establezca una buena comunicación entre los evaluadores y los interesados en la evaluación de manera que se asegure su impacto (Mateo, 2000a). Esto exige tener en cuenta las diferentes audiencias a las que habrá que presentar los informes de evaluación. Tal y como señala Santos Guerra (1990, p. 137) que «un mal informe invalida un buen proceso de evaluación».

El objetivo de un informe proviene de la función que pretende cumplir. En este sentido, Rebollo Catalán (1993) citando a Brinkerhoff *et al.* (1983) señala nueve propósitos: *demostrar rentabilidad, convencer, educar, explorar e investigar, documentar, implicar, prestar apoyo, promover la comprensión y promover relaciones públicas*. Sin embargo, y aunque la finalidad pueda ser diferente, la misma autora aprecia que «el propósito central, por excelencia, que guía la elaboración de un informe evaluativo es comunicar a las audiencias los hallazgos y las conclusiones que resultan de la recogida, análisis e interpretación de los datos de la evaluación» (Rebollo Catalán, 1993, p. 173).

Diversos autores han planteado a lo largo de los últimos años una serie de recomendaciones o prescripciones para la redacción de informes evaluativos. A partir de estas recomendaciones y consejos se ha elaborado una guía para la redacción de informes de evaluación:

1. Realizar un informe completo que incluya:
 - a) Todos los elementos implicados en el proceso.
 - b) Una relación exhaustiva de todos los resultados obtenidos.
2. Contextualizar la información que se da sobre los resultados (describir y explicar).
3. Facilitar la comprensión utilizando un lenguaje sencillo, claro y conciso y evitando el uso de vocabulario científico muy específico.
4. Aportar recomendaciones aconsejando las posibles alternativas de decisión y estableciendo pautas para su elección y estrategias de seguimiento y posterior evaluación de la opción escogida.
5. Aportar una síntesis de la información obtenida.
6. Elaborar informes específicos para cada una de las audiencias implicadas.
7. Responder a las necesidades e interrogantes que se plantearon al inicio.
8. Formular claramente los juicios valorativos respecto a los contenidos evaluados.
9. Redactar un informe equilibrado en los juicios, señalándose los aspectos de mala calidad y las virtudes que deben mantenerse.
10. Entregar los informes en el tiempo acordado, en el momento adecuado y de la forma convenida, antes de que los resultados de la evaluación pierdan interés.
11. Asegurar la independencia del evaluador en la redacción del informe.
12. Negociar con las audiencias implicadas antes de entregar el informe final/definitivo.
13. Utilizar ejemplos, anécdotas e ilustraciones para exponer las ideas y los conceptos.

14. Empezar el informe en general y cada apartado con la información más relevante e importante que se quiere comunicar.

Ahora bien, para maximizar el potencial uso de los informes de evaluación deberíamos responder a las siguientes cuatro cuestiones (De Miguel, 2000a, p. 302):

1. ¿Qué clases o tipos de usos pueden tener los informes?
 - Si el propósito de la evaluación es la búsqueda de observaciones y explicaciones generalizables mediante la utilización de procedimientos científicos, el informe será similar al de una investigación.
 - Si el propósito de la evaluación es ofrecer información para mejorar los procesos de toma de decisión, nos encontramos con que cada evaluación es singular y habrá que adecuar el informe a las demandas y usos más frecuentes que, siguiendo a De Miguel (2000a), son: para tomar decisiones, identificar problemas, inducir acciones y justificar decisiones.
2. ¿Quiénes son los usuarios potenciales de los informes?
 - En la evaluación de cualquier objeto educativo nos encontraremos con diferentes grupos de personas implicadas. Sin embargo, el grado de implicación de los distintos grupos no será el mismo (algunos serán meros receptores del programa, por ejemplo, mientras que otros serán los encargados de tomar decisiones, otros tendrán incluso una implicación indirecta, etc. Para asegurar que las recomendaciones de la evaluación (más todo el trabajo realizado) lleguen a todos los grupos, deberemos, en primer lugar, identificar las distintas audiencias y, en segundo lugar, asegurarnos de que cada una de ellas recibe el informe adecuado, de esta manera incrementaremos las posibilidades de que la evaluación tenga consecuencias.
3. ¿Qué papel puede jugar el informe para los usuarios?
 - Se deberá clarificar el uso que tendrá el informe en el proceso de toma de decisiones. En la fase de negociación de la evaluación debería quedar claro cuál va a ser el uso que se le va a dar al informe. De Miguel (2000a) citando a Shadish, Cook y Leviton (1991) señala tres usos del informe en este proceso:
 - a) *Uso instrumental*, cuando los resultados obtenidos se utilizan directamente para tomar decisiones.

- b) *Uso conceptual*, cuando las valoraciones no son utilizadas como base para decidir cambios, pero indirectamente influyen sobre cómo valoran y perciben las personas las intervenciones sociales.
 - c) *Uso persuasivo*, cuando son utilizadas para convencer a las audiencias que poseen estimaciones apriorísticas sobre los programas de que la evaluación es el proceso correcto para valorar cualquier estrategia de intervención.
4. ¿Cuáles son los factores políticos y contextuales que pueden dificultar la utilización de los resultados?
- Ésta es una problemática sobre todo de orden político en donde caben los dos extremos. Es decir, cuando los que encargan la evaluación se adueñan del informe y sólo lo divulgan en la medida en que responde a sus intereses sería un extremo. El opuesto ocurre cuando independientemente del nivel de las audiencias todos los informes deben ser difundidos para que todos tengan acceso al proceso de toma de decisiones.

Teniendo en cuenta todo lo dicho hasta ahora, en el cuadro 3.6 se ofrece una posibilidad de índice para un informe de evaluación estándar. Obviamente este índice es general y habrá que modificarlo y adecuarlo a la situación del objeto evaluado. Es decir, en función de la finalidad de la evaluación, del diseño utilizado y del objeto evaluado, este esquema será alterado, bien cambiando el orden de los apartados o bien suprimiendo algunos de ellos por no tener sentido en esa situación.

3. Fases del proceso evaluativo

La evaluación se lleva a cabo a través de un plan previamente establecido en el que pueden identificarse una serie de fases. No se puede hablar de un único proceso, más bien deberíamos señalar que se trata de un conjunto de procesos interrelacionados entre sí. Lejos queda la concepción tradicional en la que la evaluación se asociaba al control de los resultados esperados de los alumnos y se realizaba siempre al final del proceso. Hoy en día, las distintas fases del proceso evaluativo son de naturaleza multidimensional. En cada fase del proceso hay diferentes alternativas entre las que el evaluador tendrá que seleccionar. Es decir, podemos identificar múltiples propósitos de la evaluación (no sólo el control como ocurría antaño), variados objetos (no sólo el rendimiento de los alumnos), instrumentos de recogida de información diversos (no sólo los tests estandarizados), distintos momentos evaluativos (no sólo al final), etc. Esta diversidad ha complicado el proceso evaluativo en sí mismo, pero como contrapartida ha convertido la evaluación en un instrumento de gran potencial educativo.

Cuadro 3.6. Índice para un informe estándar de evaluación

1. Introducción

- 1.1. Origen, motivos y finalidad de la evaluación.
 - 1.2. Audiencias y utilidad de la evaluación.
 - 1.3. Contexto y limitaciones de la evaluación.
 - 1.4. Estructura del informe.
-

2. Objeto y finalidad de la evaluación

- 2.1. Definición y descripción del objeto de evaluación.
 - 2.2. Criterios, parámetros y estándares.
-

3. Diseño del modelo evaluativo

- 3.1. Proceso de evaluación seguido.
 - 3.2. Descripción del modelo evaluativo.
 - 3.3. Presentación de los instrumentos, técnicas y estrategias utilizadas para la recogida de información.
 - 3.4. Información recogida.
-

4. Resultados

- 4.1. Resultados obtenidos.
 - 4.2. Interpretación de los resultados obtenidos.
-

5. Conclusiones

- 5.1. Respuesta a las preguntas de la evaluación.
 - 5.2. Virtudes del objeto evaluado.
 - 5.2. Puntos débiles.
-

6. Recomendaciones

- 6.1. Consejos y sugerencias para mejorar el objeto evaluado.
-

7. Anexos

- 7.1. Resumen (síntesis de los resultados y conclusiones de la evaluación, así como de las acciones a emprender para utilizar éstos).
 - 7.2. Instrumentos de recogida de información utilizados.
 - 7.3. Muestras de observaciones.
 - 7.4. Listados de datos brutos.
 - 7.5. Otros...
-

La proliferación de los trabajos evaluativos en las últimas décadas ha contribuido a que los distintos autores propongan diferentes fases en el proceso de evaluación.

Fernández-Ballesteros (1995), basándose en diferentes trabajos, plantea el proceso de evaluación en seis fases. A saber:

FASE 1: Planteamiento de la evaluación.

FASE 2: Selección de las operaciones a observar.

FASE 3: Selección del diseño de evaluación.

FASE 4: Recogida de información.

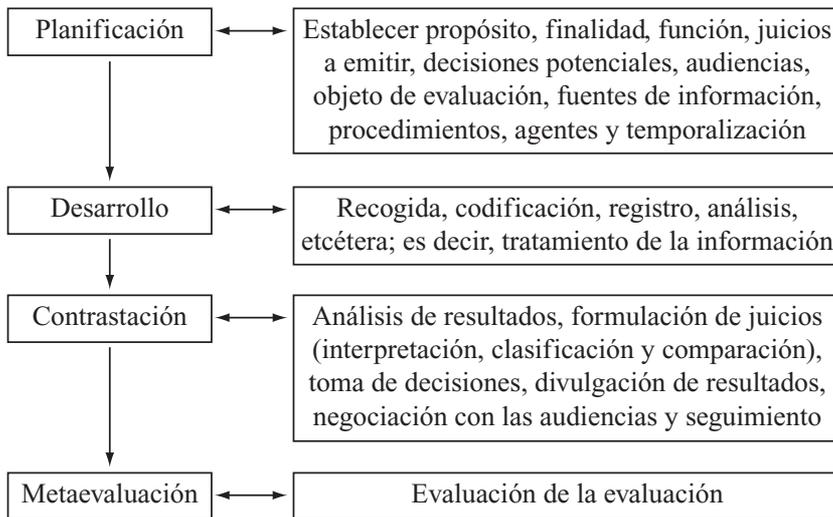
FASE 5: Análisis de datos.

FASE 6: Informe.

Para cada una de las fases, la autora plantea una serie de preguntas que el evaluador deberá ir contestando.

De la misma manera, Mateo sintetiza en la figura 3.2 las cuatro fases identificadas por él.

Figura 3.2. Fases características de un proceso de evaluación
(Mateo, 2000a, p. 37)



Como se puede apreciar al comparar las propuestas de Fernández-Ballesteros y Mateo comprobamos que las tres primeras fases de esta segunda propuesta abarcan las seis fases planteadas por Fernández-Ballesteros. Sin embargo, la propuesta de Mateo añade una fase más, la metaevaluación.

La propuesta de Nirenberg, Brawerman y Ruiz (2000) también recoge la fase de metaevaluación. En el planteamiento de estas autoras aparecen únicamente tres fases, aunque en cada una de ellas distinguen una serie de subfases (pp. 48-49):

FASE 1: PROGRAMACIÓN DE LA EVALUACIÓN

- Definición del *sujeto-objeto* de evaluación.
- Establecimiento de la *finalidad*.
- Identificación de las *audiencias*.
- Determinación del nivel de *análisis*.
- Clasificación de las *dimensiones* de análisis.
- Definición de las *variables*.
- Selección de los *indicadores*.
- Determinación de las *técnicas e instrumentos* para la recolección de información.
- Determinación de la *secuencia* de actividades.
- Identificación de *responsables y actores participantes*.
- Establecimiento de la *duración, frecuencia y periodicidad*.
- Estimación de los *recursos necesarios (humanos y materiales)*.
- *Presupuestación*.

FASE 2: EJECUCIÓN DE LA EVALUACIÓN

- *Recogida de información* mediante el desarrollo de las actividades y la aplicación de las técnicas e instrumentos.
- *Sistematización, procesamiento y análisis de la información*.
- Elaboración de *informes* (incluyendo conclusiones y recomendaciones).
- *Devolución a los actores y diseminación de conclusiones*.

FASE 3: EVALUACIÓN DE LA EVALUACIÓN (META-EVALUACIÓN)

- Análisis de la *utilidad* de los hallazgos.
- Apreciación sobre la *viabilidad* del proceso evaluativo y del uso de sus resultados.
- Juicio acerca de la *pertinencia* metodológica y ética.
- Análisis de la *confiabilidad* de los procesos y hallazgos.

Por último, Cabrera (2000a) plantea también cuatro fases como en la propuesta de Mateo, pero no incluye la metaevaluación. Para cada una de las fases plantea los objetivos a conseguir y las preguntas a las que hay que contestar. La propuesta que planteamos (en los cuadros 3.7 a 3.11) está basada inicialmente en las fases señaladas por Cabrera (2000a, pp. 40-55), aunque se ha intentado completar teniendo en cuenta las otras propuestas analizadas.

Cuadro 3.7. Fase de delimitación de la finalidad y el alcance de la evaluación

FASE 1: Delimitar la finalidad y el alcance de la evaluación	
Objetivos	Aspectos a resolver o aclarar
1. Clarificar el origen y la necesidad de la evaluación.	<ul style="list-style-type: none"> • Necesidad de la evaluación. • Uso que se hará de los resultados. • Identificación de las audiencias. • Quiénes utilizarán los resultados. • Quién debe ser informado. • Identificación de las personas que han encargado la evaluación.
2. Valorar la viabilidad de la evaluación.	<ul style="list-style-type: none"> • Identificación de la metodología y de las técnicas evaluativas susceptibles de ser utilizadas. • Calidad del programa en cuanto a poder ser evaluado. • Obstáculos que pueden producirse durante la evaluación. • Estimación de los recursos necesarios (materiales y humanos) para llevar a cabo la evaluación. • Delimitación del presupuesto.
3. Formar el equipo de evaluación.	<ul style="list-style-type: none"> • Partes interesadas que estarán integradas en el equipo de evaluación. • Credibilidad del equipo evaluador.
4. Determinar la finalidad y el objeto de evaluación.	<ul style="list-style-type: none"> • Propósito de la evaluación. • Definición del objeto de evaluación.
5. Decidir la naturaleza de la evaluación.	<ul style="list-style-type: none"> • Evaluación formativa o sumativa. • Evaluación cualitativa, cuantitativa o mixta. • Selección del modelo de evaluación. • Diseño de la evaluación.

Cuadro 3.8. Fase de planificación de la evaluación

FASE 2: Planificación de la evaluación	
Objetivos	Aspectos a resolver o aclarar
1. Especificar los criterios de evaluación y las cuestiones específicas a las que la evaluación debe responder.	<ul style="list-style-type: none"> • Selección de las preguntas que debe responder la evaluación. • Selección de las personas que participarán en la elaboración de las preguntas de evaluación. • Selección de los criterios y referentes que se utilizarán para responder a las preguntas de la evaluación. • Decisiones potenciales.
2. Planificar la recogida de la información.	<ul style="list-style-type: none"> • Identificación de la información necesaria. • Elección del diseño de la evaluación. • Elección de los procedimientos que se utilizarán y la persona que los aplicará. • Identificación de las fuentes de información. • Identificar los potenciales sesgos que puedan producirse durante la recogida de información. • Temporalización de las tareas evaluativas.

Cuadro 3.9. Fase de tratamiento de la información

FASE 3: Tratamiento de la información	
Objetivos	Aspectos a resolver o aclarar
1. Seleccionar o en su caso elaborar los instrumentos de recogida de datos.	<ul style="list-style-type: none"> • Selección de los instrumentos necesarios. • Elaboración de los instrumentos necesarios. • Selección de la muestra. • Fiabilidad y validez de los instrumentos. • Aplicación de los instrumentos.
2. Analizar la información cuantitativa y cualitativa.	<ul style="list-style-type: none"> • Identificación de las formas de almacenamiento de la información. • Identificar los criterios que guiarán los análisis. • Identificación de las técnicas estadísticas. • Análisis de la información cualitativa.
3. Interpretar la información.	<ul style="list-style-type: none"> • Identificación de los patrones con los que se compararán los resultados. • Identificación de los niveles que deben presentar los resultados para emitir un juicio positivo. • Formulación de juicios (interpretación, clasificación y comparación).

Cuadro 3.10. Fase de elaboración del informe y difusión de los resultados

FASE 4: Elaboración del informe y difusión de los resultados	
Objetivos	Aspectos a resolver o aclarar
1. Devolver y contrastar la información.	<ul style="list-style-type: none"> • Identificar las personas con las que se negociarán las interpretaciones. • Contrastación de las interpretaciones.
2. Elaborar el informe.	<ul style="list-style-type: none"> • Contenido del (de los) informe(s). • Información a las audiencias.
3. Difundir los resultados.	<ul style="list-style-type: none"> • Exposición/divulgación de los resultados. • Publicación del informe. • Toma de decisiones. • Seguimiento.

Cuadro 3.11. Fase de metaevaluación: evaluar la calidad de la evaluación efectuada

FASE 5: Metaevaluación: Evaluar la calidad de la evaluación efectuada ⁵	
Objetivos	Aspectos a resolver o aclarar
1. La <i>utilidad</i> de la evaluación; es decir, evaluar si se satisfacen las necesidades prácticas de información de las audiencias.	<ul style="list-style-type: none"> • Identificación de la audiencia o de los interesados. • La confianza/credibilidad del evaluador. • La selección y alcance de la información. • La interpretación valorativa e identificación de valores. • La claridad, difusión y oportunidad del informe. • La trascendencia e impacto de la evaluación.
2. La <i>factibilidad o viabilidad</i> de la evaluación; es decir, analizar si la evaluación realizada ha sido realista, prudente, diplomática y económica.	<ul style="list-style-type: none"> • Los procedimientos prácticos. • La viabilidad política. • La relación coste/producto ajustada o eficacia de costos.

Cuadro 3.11. (continuación)**FASE 5: Metaevaluación: Evaluar la calidad de la evaluación efectuada⁵**

Objetivos	Aspectos a resolver o aclarar
3. La <i>legitimidad o ética</i> de la evaluación, es decir, asegurar que la evaluación se ha realizado de forma legal, con la ética y el respeto necesarios, por el bienestar de los participantes en la evaluación, así como de los afectados por los resultados.	<ul style="list-style-type: none"> • La obligación formal. • El conflicto de intereses. • La exposición total y franca. • El derecho a la información pública. • Los derechos del sujeto. • Las interacciones entre personas. • Revelación de los resultados y equilibrio del informe. • La responsabilidad fiscal y económica.
4. La <i>precisión o rigurosidad</i> de la evaluación; es decir, asegurar que la evaluación ha proporcionado información adecuada sobre las características del objeto de estudio para determinar su mérito o su valor.	<ul style="list-style-type: none"> • La identificación del objeto. • El análisis del contexto. • Los propósitos y procedimientos descritos. • Las fuentes de información confiables. • Las mediciones válidas. • Las mediciones fiables. • El control sistemático de los datos. • El análisis de la información cuantitativa. • El análisis de la información cualitativa. • Las conclusiones fundamentales. • El informe objetivo. • Metaevaluación.

Conviene no olvidar una serie de aspectos relacionados con estas fases evaluativas, de tal manera que su forma de presentación y secuenciación no lleve a extraer conclusiones erróneas. Cabrera (2000a) hace hincapié en los aspectos siguientes:

- Tradicionalmente, la evaluación se ha centrado excesivamente y casi exclusivamente en la tercera fase. Ahora, se concibe como una fase dependiente de las dos anteriores, aunque se siga reconociendo e insistiendo en la rigurosidad de la recogida y análisis de la información.
- Las fases planteadas no deben ser consideradas como una secuencia lineal rígida e inflexible. Puede haber fases previas que podrían ser clarificadas al avanzar la evaluación a fases posteriores. De la misma manera, aunque la metaevaluación la hemos incluido en la quinta fase, se supone que ésta está presente a lo largo de todo el proceso de evaluación.

- Concepciones como *proceso abierto, flexible, emergente* son las que deben caracterizar la evaluación. Este carácter, sin embargo, no es excluyente con el establecimiento de un plan de acción sistemático.
- En la actualidad, se intenta favorecer la participación de las distintas personas relacionadas con la evaluación en el propio proceso de evaluación. Lógicamente, esta participación variará de una fase a otra y dependerá del modelo de evaluación adoptado.

4. La evaluación educativa y otras disciplinas

Para terminar con el capítulo dedicado a la aproximación conceptual a la evaluación educativa, vamos a intentar disipar las diferencias y las similitudes que tiene ésta con algunas disciplinas con las que está relacionada y con las que a veces se funde o se confunde. Éstas no son otras que la medición, el *assessment* y el diagnóstico.

4.1. Medición y evaluación

Estos dos conceptos tuvieron un origen común en la historia de la educación y han compartido durante tiempo incluso la definición, puesto que a principios del siglo XX eran tomados como sinónimos y no es hasta la llegada de Tyler en los años treinta del pasado siglo y su nueva conceptualización de la evaluación que diferencia ésta de la medición.

La definición de medición en el ámbito de las Ciencias Sociales en general y en el educativo en particular no plantea los problemas ya señalados a la hora de definir la evaluación. Existe un acuerdo más o menos unánime basado en la ya clásica definición de Stevens (1951) de proponer que la medición «es el acto de asignar numerales a objetos o hechos empíricos de acuerdo con ciertas reglas». Los numerales⁶ pueden o no ser números, y las reglas⁷ serán más o menos rigurosas. En función de esta rigurosidad, la precisión de los números variará (Pérez Juste, 1995b).

De esta manera, en las siguientes definiciones podemos comprobar que el concepto de medición en educación es unívoco y, por lo tanto, no conduce a errores:

Lord y Novick (1968, p. 17) definen la medida como

un procedimiento para la asignación de números (puntuaciones, medidas) a propiedades especificadas de unidades experimentales, de modo que las caractericen y preserven las relaciones especificadas en el dominio comportamental.

Por su parte, Santisteban (1990, p. 19) señala que

medir es asignar números a las características de los objetos o de los individuos mediante un procedimiento sistemático que nos dé una representación de las propiedades concernientes a esos objetos o a esas características o rasgos.

Por último, Nunnally y Bernstein (1995, p. 1) apuntan que

la medición consiste en reglas para asignar símbolos a objetos de manera que 1) representen numéricamente cantidades de atributos (escala de medición) o 2) definan si los objetos caen en las mismas categorías o en diferentes con respecto a un atributo determinado (clasificación).

En este sentido, Serrano, Tójar, Galante y Elósegui (1995, p. 130) aducen que

todavía estamos lejos de tener instrumentación tecnológica para medir en educación. El metro o el litro no se ha inventado en educación. Se mide con distintos tests y cuestionarios y se tienen los mismos inconvenientes que cuando se medía a palmos. Eran medidas groseras y no universales. Así, en cada caso, se mide de una forma y con una unidad distinta.

Las palabras de Batanaz describen cuál es la situación actual:

La práctica de la medición en aspectos relacionados con la educación ha dado origen a un cuerpo disciplinar que, con la denominación de Medición en Educación, tendría como objetivo el estudio sistemático de las diferentes aproximaciones científicas sobre la medida aplicada a la educación. Conceptos tan usuales como validez, fiabilidad, análisis de ítems, por citar sólo algunos, proceden de este campo, y aunque su aplicación rigurosa corresponde a quienes se ocupan de elaborar los conocidos como instrumentos de medida, deben ser, así mismo, familiares a quienes aplican los resultados de estos instrumentos a procesos de evaluación o diagnóstico.

(Batanaz, 1996, p. 69)

Sin embargo, este acto de precisar «el grado en que un individuo posee un determinado atributo no valora la calidad o el mérito de dicha medida» (Popham, 1988, p. 9). Es decir, no hemos llegado más que a establecer una relación comparativa entre dos variables, una de las cuales es tomada como referente y el resultado se expresa en un número (Wolf, 1990b; Batanaz, 1996). Este número no es capaz por sí mismo de ser interpretado. Y he aquí lo que diferencia la medición de la evaluación. Para que se dé esta última, es necesario que se emita un juicio de valor sobre el objeto medido. Es decir, a través de la evaluación se determinará el grado en que ese objeto posee un valor (Livas, 1980). Dicho de otra forma, la medición ofrece cierto tipo de información cuantitativa, mientras que la evaluación implica valorar esa información con el propósito de emitir un juicio sobre el objeto evaluado (Aguilar y Ander-Egg, 1994).

Además, mientras que la medida tiene como finalidad la descripción y comparación de individuos, la evaluación no comporta la necesidad de la comparación, ni de la medida (Batanaz, 1996). Es decir, nos vamos a encontrar con tres situaciones diferentes:

- a) *Utilización de la medida sin ánimo de evaluar.* Hay situaciones, en cierto tipo de investigación educativa por ejemplo, en las que la medi-

da juega un papel importante; sin embargo, en estos casos no está unida a la evaluación.

- b) *Evaluar sin necesidad de la medida.* Si la recogida y análisis de la información se reduce a técnicas y procedimientos cualitativos, se podrá hacer evaluación (emitir juicios de valor y tomar decisiones) sin necesidad de medir.
- c) *Utilización de la medición como parte de la evaluación.* En este sentido, la medición adquiere carácter instrumental, es una parte de la evaluación que proporciona la forma de recabar información valiosa y métricamente bien construida.

Esta tercera situación es la que ha contribuido a que, en diversas ocasiones, ambos términos se confundan; no obstante, las diferencias entre ellos son claras:

- El resultado de la medición es una medida, es decir, un número (o numeral para ser más precisos), mientras que el resultado de la evaluación es un juicio de valor.
- La medición (en esta tercera situación) es parte de la evaluación que es un proceso mucho más amplio.
- Siguiendo con el razonamiento anterior, la medición se convierte en un instrumento de la evaluación.
- La medición tiene siempre carácter objetivo, mientras que la evaluación necesita de juicios de valor.
- La medición tiene carácter pretérito, mientras que la evaluación, además de tener carácter pretérito, tiene carácter prospectivo (toma de decisiones y mejora).

4.2. *Assessment* y evaluación

Como ya se apuntó en el apartado dedicado a la historia de la evaluación educativa, la mayoría de los primeros avances se dieron en el mundo anglosajón. En él fueron creándose o, mejor dicho, adoptándose una serie de vocablos para representar distintos conceptos relacionados con la evaluación. En nuestro contexto, hemos tenido que ir traduciendo en algunos casos y adoptando neologismos en otros para seguir el tren de los avances en este campo. Estas traducciones y adopciones han sido acertadas en muchos casos y hoy funcionan como términos usuales entre nosotros. Sin embargo, en otros, las traducciones han creado problemas todavía no resueltos. Éste es el caso de los términos ingleses *assessment* y *evaluation*. Ellos, como veremos más adelante, aunque admiten que son términos sinónimos los distinguen entre sí. Sin embargo, en el mundo hispano los dos términos sajones han sido traducidos por el mismo término español: evaluación. Esto ha supuesto que nosotros utilicemos indistintamente el término evaluación como traducción de ambos con la confusión que ello crea en muchas ocasiones.

En esta línea, Mateo (1986, p. 48) señala que «quedando claro, por tanto, que *evaluation* y *assessment* implican conceptos diferentes y que si bien literalmente son sinónimos, no pueden ser utilizados en el vocabulario especializado de forma intercambiable ya que implican objetos diferentes y no son semánticamente idénticos».

Veamos primeramente cómo distinguen estos términos para, posteriormente, señalar algunas contribuciones que han intentado resolver el problema.

Doval (1995) sostiene que el *assessment* definido como valoración, tasación, medida o estimación (o incluso comprobación o apreciación para Batanaz, 1996) nació vinculado al asesoramiento (o *counseling*) y a la toma de decisiones clínicas de asesoramiento y tutoría a conductas humanas.

La mayoría de los autores señalan que el término *assessment* debe utilizarse para aplicarlo a personas (alumnos, profesores, etc.) que puede incluir el uso de procedimientos de tests, exámenes, certificación o actividades de clasificación con algún criterio especificado para posicionarle dentro de una escala. El término *assessment* abarca desde el *testing* y *measurement* hasta el diagnóstico y la evaluación continua.

Wolf (1990c) y Batanaz (1996) señalan que el *assessment* forma también parte de las prácticas habituales de los profesores para adaptar el proceso de enseñanza-aprendizaje a las características coyunturales que presentan los alumnos. Choppin (1990) apunta estas características del *assessment*:

1. Es un término reservado para aplicarlo a sujetos considerados individualmente.
2. Su realización debe llevarse a cabo por medio de procesos formales (tests, exámenes, etc.) e informales (observación espontánea, entrevista informal, etc.).
3. Sus efectos se traducen en algún tipo de expresión calificativa del sujeto con el fin de posicionarlo en un determinado lugar de una escala.

El mismo autor señala las situaciones en las que se puede utilizar el *assessment*:

- Cuando se analiza el rendimiento académico de un alumno.
- Cuando se describen las aptitudes de una persona para una determinada actividad laboral.
- Cuando se describe la competencia de un profesional para el desempeño de su tarea.

Lázaro (1990, p. 8) mantiene las diferencias apuntadas por los autores previamente citados. De esta manera, define *assessment* como los «procedimientos y procesos utilizados para recoger medidas sobre la conducta»,

3. Concepto, componentes y fases de la evaluación

mientras que señala que la evaluación se da «cuando se comparan los datos obtenidos respecto a otras variables relevantes, formándose en torno a ellos una opinión, un juicio adecuado, adoptándose un juicio valorativo sobre las variables medidas».

Por otra parte, Doval (1995) alude a que para los anglosajones el término evaluación se utiliza para los programas, el currículum y las situaciones organizativas y/o políticas. Es decir, cuando se aplica a entidades abstractas (Choppin, 1990; Lázaro, 1990).

La evaluación, pues, «es una actividad más vinculada a procedimientos de investigación aplicados con rigor científico, para la que se necesita una determinada pericia y orientada a la mejora de sistemas complejos» (Batanaz, 1996, p. 70).

Ante esta disyuntiva, hace ya más de 15 años, algunos autores intentaron resolver esta situación presentando nuevas propuestas para designar ambas actividades (véase cuadro 3.12). En este sentido, Fernández-Ballesteros (1985) propuso utilizar el término evaluación como traducción de *assessment* y valoración como equivalente a *evaluation*. Lejos de que esta propuesta cristalizara, un año después Mateo (1986) expresó su desacuerdo señalando que en vez de aclarar, la propuesta de Fernández-Ballesteros complicaba aún más la situación (téngase en cuenta que en su propuesta se utiliza valoración para *evaluation*, y evaluación, para *assessment*, dándose una traducción fonéticamente cruzada). Mateo (1986) propone utilizar el término evaluación para *assessment* y para *evaluation* propone utilizar *investigación evaluativa*. Tampoco parece que esta propuesta haya calado. Todavía es usual en la actualidad utilizar el término evaluación como traducción tanto de *evaluation* como de *assessment* (tal y como propuso por ejemplo Silva, 1989). Massons (1997), por su parte, sugiere la utilización de evaluación para la traducción de *evaluation*, y análisis o diagnóstico, para la traducción de *assessment*. Ahora bien, hay algunos autores que emplean la palabra inglesa *assessment* sin traducción, y evaluación, para *evaluation* (Lukas *et al.*, 2000a; Santiago *et al.*, 2001).

Cuadro 3.12. Propuestas de traducción de *evaluation* y *assessment*

Vocablo Inglés	Fernández-Ballesteros (1985)	Mateo (1986)	Silva (1989)	Massons (1997)	Alternativa
<i>ASSESSMENT</i>	Evaluación	Evaluación	Evaluación	Diagnóstico o análisis	<i>Assessment</i>
<i>EVALUATION</i>	Valoración	Investigación evaluativa	Evaluación	Evaluación	Evaluación

4.3. Diagnóstico y evaluación

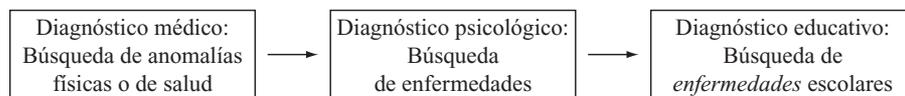
Un debate todavía quizás no superado es el relativo a la(s) diferencias(s) y similitud(es) entre el diagnóstico educativo y la evaluación educativa. Estos dos campos, cuyo origen es diferente, están abocados en la actualidad a entenderse, a complementarse e incluso a unirse.

Como ya se ha visto previamente, en sus orígenes la evaluación educativa estaba centrada en el alumno, se llevaba a cabo al final del proceso de enseñanza-aprendizaje y tenía una función sancionadora. Con el paso del tiempo, esta concepción fue evolucionando hacia posiciones en las que se admitían múltiples objetos para ser evaluados, la evaluación debía llevarse a cabo a lo largo del proceso y su función era claramente formativa.

Por su parte, el diagnóstico proviene originariamente del campo de la medicina en la que es definida como un «conjunto de signos que sirven para fijar el carácter de una enfermedad» (Padilla Carmona, 2002, p. 19). Posteriormente, Rorschach utilizó a comienzos de la década de 1920 «el término psicodiagnóstico para referirse al campo específico de las enfermedades mentales» (Padilla Carmona, 2002, p. 19).

Más tarde surge el diagnóstico educativo, siguiendo la tradición médica y psicológica, como una concepción cuyo objetivo es detectar las disfunciones o anomalías que se dan entre los alumnos en el proceso escolar. El proceso seguido está reflejado en la figura 3.3.

Figura 3.3. Evolución del concepto de diagnóstico



Este enfoque tradicional y ya superado del diagnóstico educativo tenía las características siguientes (Padilla Carmona, 2002, p. 19):

- Es una actividad independiente y no relacionada con la intervención educativa.
- Es utilizada con aquellos alumnos que presentan dificultades o problemas de aprendizaje.
- No está contextualizada en el proceso de aprendizaje en donde se producen las dificultades.
- Es una actividad clínica.

Desde esta perspectiva del diagnóstico, las diferencias entre éste y la evaluación son claras (Padilla Carmona, 2002):

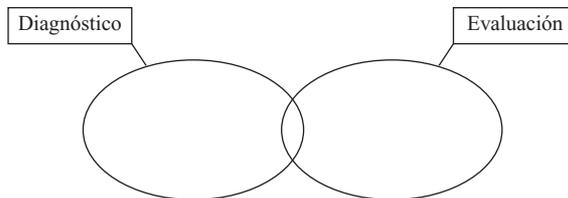
- a) El diagnóstico estaría muy cerca del término inglés *assessment* que se utiliza para indicar el carácter descriptivo y aplicado siempre a suje-

tos, mientras que la evaluación (*evaluation*) sería utilizada con entidades abstractas, como se ha mencionado previamente, y orientada a emitir juicios de valor y a la toma de decisiones.

- b) El diagnóstico está vinculado a variables psicobiológicas (aptitudes, intereses, personalidad), mientras que la evaluación está más relacionada con el rendimiento académico (sobre todo de tipo cognoscitivo).
- c) El diagnóstico se aplica a sujetos con dificultades o problemas, por contra la evaluación se aplica a todos los alumnos (Pérez Juste, 1994b; Batanaz, 1996).
- d) El diagnóstico se concentra en descubrir e interpretar signos, mientras que la evaluación se concentra en valorar, justipreciar algo (Pérez Juste, 1994b).
- e) El diagnóstico antecede a la actuación de los profesores, ya que se concibe como orientadora y guía del tratamiento, mientras que la evaluación trata de comprobar el progreso y los resultados del tratamiento, por lo que es posterior (Pérez Juste, 1994b; Batanaz, 1996).
- f) El diagnóstico es realizado por profesionales especializados (en muchos casos externos a la propia institución escolar), mientras que la evaluación es realizada por los profesores.

En este sentido, si utilizáramos la teoría de conjuntos para expresar la relación entre ambos, observaríamos que apenas existe relación (figura 3.4), reduciéndose ésta a la utilización de algunas técnicas de recogida de información comunes.

Figura 3.4. Relación entre la evaluación y el diagnóstico



Sin embargo, al igual que ha sucedido con la evaluación educativa, el diagnóstico ha ido ampliando su radio de acción y lejos de limitarse a acciones terapéuticas y correctivas ha desviado su atención hacia dimensiones más preventivas y de desarrollo de potencialidades (Pérez Juste, 1994b). En la actualidad, las características que definen el diagnóstico son las siguientes:

- El diagnóstico no se aplica únicamente a los alumnos con dificultades, sino a todos los alumnos, ya que todos son considerados como especiales, y, por tanto, la educación debe adecuarse a todos ellos.

- La función del diagnóstico no es la clasificación de los sujetos, sino que adopta un carácter formativo en relación con la toma de decisiones y mejora educativa en general. Se ha pasado de pensar en términos médicos y psicológicos a pensar en términos educativos.
- De tener un carácter correctivo y terapéutico ha pasado a ser preventivo y de desarrollo de potencialidades.
- Ha de realizarse en contextos en los que interactúa el alumno, pasando a ser estos contextos susceptibles de ser diagnosticados.
- Como consecuencia de la característica anterior, ya no se diagnostica solamente al alumno, sino que también los programas, el profesorado, los centros, etc.
- El diagnóstico no tiene sentido en sí mismo, si no es dentro de una propuesta de intervención educativa de mejora de una situación.
- Pasar de estar centrado únicamente en el sujeto a estar en el sujeto y en su entorno ha contribuido a que en el diagnóstico participen todos los agentes educativos implicados (profesores, padres, alumnos, etc.).

En las mencionadas características se ha utilizado intencionadamente el término diagnóstico en todas las ocasiones. No obstante, si intercambiáramos ese término por el de evaluación nos encontraríamos con que nada ha cambiado; es decir, se podría admitir que los dos campos han llegado a confluir en uno mismo.

Hay autores, quizás los menos, que señalan diferencias etimológicas y conceptuales. Castillo Arredondo (1994) postula que el diagnóstico se da en todo momento inicial del proceso didáctico y está dirigido a describir, explorar y descubrir dificultades que pueden interferir en dicho proceso. Por contra, la evaluación supone el seguimiento crítico del proceso y está dirigida a emitir juicios de valor con intenciones formativas y sumativas. Es decir, el proceso didáctico comienza con el diagnóstico y termina con la evaluación. De todas formas, Castillo Arredondo (1994) admite la complementariedad de ambos. De la misma manera Alfaro (1994, p. 578) señala que «evaluar es asignar un valor a una cosa objetivamente dada y de conformidad con un criterio [...] la evaluación es valoración, ponderación (*pondus*). Hablo de evaluación cuando se busca el cuánto [...]. Entiendo por diagnóstico la búsqueda de qué. La búsqueda de cuáles son las causas, síntomas, indicios, etc., tanto de un ajuste como un desajuste». La figura 3.4 podría valer para expresar la relación de la evaluación y el diagnóstico desde este punto de vista.

Otros autores como Lázaro (1994) señalan que el diagnóstico abarca a la evaluación, ya que el diagnóstico no termina con la valoración emitida sino que conlleva un proceso permanente de seguimiento y control de la intervención:

Con muchas matizaciones conceptuales, entendemos que el hecho diagnóstico implica la delimitación dinámica de situaciones educativas, para valorarlas y tomar decisiones de in-

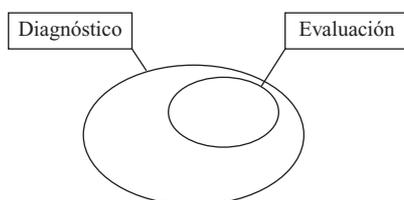
3. Concepto, componentes y fases de la evaluación

tervención que afecten a alumnos, profesores, programas, procesos, productos, organizaciones, instituciones o sistemas. Y, por tanto, al diagnosticar se describe y se evalúa; y, por supuesto, se evalúa porque se está diagnosticando.

(Lázaro, 1994, p. 594)

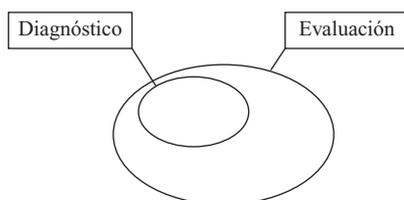
Martínez González (1994) coincide con Lázaro al afirmar que el diagnóstico es un concepto más amplio que el de evaluación ya que aunque los dos pertenecen al mismo proceso, la evaluación es anterior a la interpretación diagnóstica por lo que, de alguna manera, ésta la abarca (Batanaz, 1996). Según la autora, no puede haber diagnóstico si no hay valoración o evaluación. Desde esta perspectiva, la figura 3.5 representa esta situación.

Figura 3.5. Relación entre la evaluación y el diagnóstico



Álvarez y Bisquerra (1996, p. 40), no obstante, señalan lo contrario al afirmar que «la evaluación se encuentra dentro de las funciones orientadoras, pero que es una función más amplia que la de diagnóstico, siendo el diagnóstico una fase de la evaluación». «Para evaluar hay que diagnosticar, pero el diagnosticar no implica necesariamente evaluar» (Mayor Ruiz, 1998, p. 41). En la misma línea se presenta en un primer momento Pérez Juste (1989) cuando alude a que el diagnóstico es la fase inicial del proceso evaluador (figura 3.6).

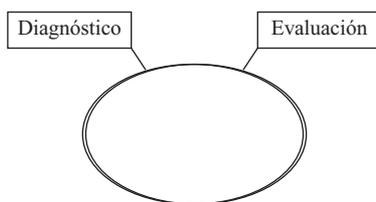
Figura 3.6. Relación entre la evaluación y el diagnóstico



Aunque en 1994 este mismo autor señala que apenas se aprecian diferencias entre ambos conceptos excepto las relativas al profesional que las realiza.

En la actualidad, parece que la mayoría de los autores aceptan que las semejanzas entre el diagnóstico y la evaluación son tales que podrían intercambiarse prácticamente los términos tal y como aparece en la figura 3.7 (García Jiménez, 1994a, Padilla Carmona, 2002). Frente a los esfuerzos de algunos

Figura 3.7. Relación entre la evaluación y el diagnóstico



autores por defender el espacio propio y exclusivo de cada uno de los campos (Batanaz, 1996), García Jiménez (1994a) defiende la idea de que ambos son conceptualmente similares y las actividades derivadas cada vez menos diferentes (González Such, 2000; Jornet, 2001).

De hecho hay autores que proponen una misma definición para evaluación y para diagnóstico:

Recogida sistemática de información, a través de técnicas que garanticen una cierta objetividad y precisión en la misma. Así mismo, la interpretación y valoración de la información recogida con la finalidad de facilitar la toma de decisiones son actividades que aparecen en las definiciones de ambos conceptos.

(Martínez González, 1994, citado por Padilla Carmona, 2002, p. 23)

Proceso o resultado de la indagación de una realidad educativa a examen, que permite describirla, compararla con un estándar y emitir juicios ordenados a la toma de decisiones.

(Doval, 1995, p. 211)

Proceso de búsqueda de información a través de métodos científicos orientado a la valoración de una determinada situación educativa con vistas a la toma de decisiones que repercutan en su mejora.

(Batanaz, 1996, p. 63)

Desde el punto de vista de estos autores, éstas serían a grandes rasgos las similitudes entre ambos:

- Uso de las mismas técnicas, estrategias y procedimientos de recogida de información.
- Misma finalidad, es decir, están orientados a la toma de decisiones y a la mejora del objeto evaluado.

Según Batanaz (1996) va ganando la hipótesis según la cual las coincidencias superan por mucho a las diferencias tanto por lo que se refiere al contenido como a los aspectos metodológicos.

Por ello, algunos autores insisten en la necesidad de suprimir el término diagnóstico por sus connotaciones médicas y psicológicas (Verdugo, 1994).

Martínez González (1994) no obstante señala que una de las diferencias entre ambos sería la relativa a las funciones, dado que el diagnóstico tiene carácter únicamente formativo, mientras que la evaluación puede, además, tener carácter sumativo. De tal forma que, como señala Parra (1996), en el diagnóstico la toma de decisiones es siempre de carácter pedagógico, mientras que en la evaluación puede tener también carácter político.

Para terminar con este apartado, recogemos las palabras de Padilla Carmona (2002, p. 23):

En nuestra opinión, la cuestión de las semejanzas y diferencias entre diagnóstico y evaluación no está zanjada en la actualidad, quizá como consecuencia de que estamos ante una disciplina que, tras años de adaptar los modelos y enfoques psicológicos, sólo recientemente se ha incorporado a la actividad educativa y está, por tanto, en plena fase de expansión.

Preguntas y ejercicios

1. Explica lo que significa la definición de evaluación que se ha presentado en el capítulo.
2. Teniendo en cuenta las diferentes etapas educativas (Infantil, Primaria, Secundaria, Universidad, etc.), ¿qué espacio tienen las funciones formativa y sumativa de la evaluación en cada una de ellas?
3. Teniendo en cuenta el evaluador, ¿qué tipos de evaluación se distinguen?
4. ¿En qué casos convendría realizar una evaluación externa de un centro y en cuáles una evaluación interna? Señala las razones de tu elección.
5. ¿Qué funciones puede desempeñar la evaluación si evaluamos al comienzo del proceso de enseñanza-aprendizaje, durante el proceso y a su finalización?
6. ¿Qué tipos de estándar se distinguen?
7. Teniendo en cuenta las diferentes etapas educativas (Infantil, Primaria, Secundaria, Universidad, etc.), ¿qué espacio tienen los distintos estándares en cada una de ellas?
8. Señala las recomendaciones más relevantes para la redacción de un informe de evaluación.
9. Señala las fases más relevantes de un proceso de evaluación.
10. ¿Cuáles son las diferencias entre la medición, la evaluación y el *assessment*?

Notas

1. En el Anexo 1 se pueden revisar más definiciones de evaluación educativa, tanto de esta categoría como de las otras.
2. Hay autores (Nirenberg, Brawerman y Ruiz, 2000, p. 62) que señalan que la evaluación sumativa debería llamarse evaluación de resumen. Argumentan que la traducción lite-

ral de *summative* por sumativa no es adecuada, pues en castellano connota la suma o acumulación y, sin embargo, tal y como la usó Scriven en su origen «connota una recapitulación de los efectos de un conjunto de acciones sobre sus destinatarios». De todas formas, en este libro se seguirá utilizando el término sumativo debido a su aceptación y extensión.

3. La terminología *ex ante* y *ex post* así como su distinción provienen de la tradición económica y ha sido trasladada a las Ciencias Sociales.
4. Cabrera (2000a, p. 29) llega a decir que la determinación de los criterios para poder emitir juicios de valor es el «talón de Aquiles» de la evaluación.
5. Como en toda evaluación, al realizar la metaevaluación necesitamos de unos referentes. El Joint Committee on Standards for Educational Evaluation (un comité internacional con más de 200 miembros que se ocupa de la ética profesional de los evaluadores) estableció en 1981 una serie de estándares y normas para garantizar la evaluación de la propia evaluación. Estas normas fueron revisadas y nuevamente publicadas en 1994. Las dos publicaciones están traducidas al español y publicadas en 1988 y 1998, respectivamente. Las normas aquí expuestas son, pues, las propuestas por el mencionado comité.
6. En el caso de las escalas nominales, los números más bien son meras etiquetas que únicamente nos diferencian un conjunto de otro. Es decir, no tienen valor numérico.
7. En función de estas reglas, Stevens (1951) señaló cuatro tipos de escalas (nominal, ordinal, de intervalos y de razón), en función de las cuales los números obtenidos en cada una de ellas tienen un nivel de precisión diferente y, por lo tanto, admiten la utilización de técnicas estadísticas más o menos sofisticadas.

4. Evaluación de programas

1. Concepto

Antes de definir el concepto de evaluación de programas, vamos a aclarar lo que se entiende por programa. De la Orden (1985b) afirma que un programa es «cualquier curso sistemático de acción para el logro de un objetivo o un conjunto de objetivos». Pérez Juste ofrece dos definiciones similares a la anterior al señalar que un programa es «un documento intencional y técnicamente elaborado consistente en un plan de actuación al servicio de metas pedagógicas valiosas» (1995a, p. 84) o un «plan sistemático diseñado por el educador como medio al servicio de las metas educativas» (2000, p. 268). Este documento necesita el acople de aspectos científicos y técnicos con otros psicopedagógicos y sociales. En un programa deben integrarse componentes tan diversos como los contenidos, el enfoque de estudio ligado a las metas y a los objetivos, los medios y recursos para lograr el aprendizaje (motivación, metodología, recursos didácticos, las actividades y los materiales, etc.) y la evaluación (Pérez Juste, 1995a, 2000 y 2002).

No obstante, un programa también puede ser definido fuera del ámbito educativo, de tal manera que tenemos definiciones más generales. Valga como ejemplo la definición de Bordás Alsina (1995, p. 178) que señala que un programa es «una actividad dirigida a solucionar un determinado problema, considerando como tal la carencia de algún aspecto, la necesidad de introducción de alguna modificación para la mejora, o la realización de alguna

innovación». O esta otra de Stake (1976) que considera que un programa es toda actividad organizada:

- Que se prolonga en el tiempo.
- Que persigue unos objetivos.
- Que cuenta con un sistema de gestión y financiación.
- Que se dirige a un grupo de individuos.
- Que despierta el interés de muchos otros.

Por otra parte, Mateo (2000a) señala que el programa como unidad de análisis (o ente aislado) no ha aportado suficiente significado a la interpretación de la realidad educativa y por ello prefiere hablar de currículum, ya que éste ofrece una perspectiva más holística. De tal forma que el programa no es más que una subunidad dentro del currículum. Sin embargo, cuando define currículum se encuentra que las distintas aproximaciones están dentro de un continuo entre los extremos siguientes:

- a) Un extremo sería cuando definimos el currículum como «un plan para la acción, o un documento escrito que incluye un conjunto de estrategias para alcanzar los objetivos o fines deseados» (p. 126). Esta definición coincidiría con las de programa que hemos señalado previamente.
- b) El otro extremo sería cuando se define el currículum relacionándolo con las experiencias del alumno (Mateo, 2000a, p. 127): «se considera cualquier elemento de la escuela e incluso de fuera de ella (en la medida en que está planificado), como parte integrante del currículum».

Entre estos dos extremos cabrían casi todas las definiciones de currículum.

De todas maneras, no todos los autores coinciden con la precisión realizada por Mateo. Por ejemplo, Mayor Ruiz (1998) señala que el término currículum tiene connotaciones específicamente enfocadas hacia la escuela, mientras que el término programa tiene acepciones más amplias relacionadas con otros ámbitos sociales. Desde su punto de vista, el currículum no es ni más ni menos que un programa educativo con sus matizaciones y singularidades.

Centrándonos en la metodología evaluativa, todos los autores coinciden en que su aplicación tanto a la evaluación de programas como a la evaluación del currículum no difiere significativamente; es por ello por lo que muchas veces se utilizan indistintamente (Mateo, 2000a).

Veamos ahora cómo se define la evaluación de programas. Las definiciones seleccionadas, como se verá, son más o menos coincidentes:

- Bordas Alsina (1999, p. 239): «Es un conjunto de actividades que se realizan para detectar y valorar la utilidad del programa en el contexto donde se proyecta y se toman decisiones, si son necesarias, para optimizar el programa sea en su totalidad sea en aspectos específicos».

- De la Orden (2000, p. 383): «Es el proceso sistemático de recogida, análisis e interpretación de información relevante y fiable para describir un programa educativo, o una faceta significativa del mismo, y formular un juicio sobre su adecuación a un criterio o patrón, que representa un valor aceptado, como base para la toma de decisiones sobre el programa o faceta programática».
- De Miguel (2000a, p. 289): «Es un conjunto de principios, estrategias y procesos que fundamentan la evaluación de toda acción o conjunto de acciones desarrolladas de forma sistemática en un determinado contexto con el fin de tomar las decisiones pertinentes que contribuyan a mejorar las estrategias de intervención social».
- Nirenberg, Brawerman y Ruiz (2000, p. 32): «Es una actividad programada de reflexión sobre la acción, basada en procedimientos sistemáticos de recolección, análisis e interpretación de información, con la finalidad de emitir juicios valorativos fundamentados y comunicables sobre las actividades, resultados e impactos de esos programas, y formula recomendaciones para tomar decisiones que permitan ajustar la acción presente y mejorar la acción futura».

Por último, quisiéramos resaltar la definición que Pérez Juste ha ido defendiendo a lo largo de sus escritos acerca del tema de la evaluación de programas:

Proceso sistemático, diseñado intencional y técnicamente, de recogida de información rigurosa —valiosa, válida y fiable—, orientado a valorar la calidad y los logros de un programa, como base para una posterior toma de decisiones de mejora tanto del programa como del personal implicado y, de modo indirecto, del cuerpo social en que se desarrolla.

(Pérez Juste, 2002, pp. 63-64)

En todas estas definiciones se distinguen claramente tres elementos o componentes básicos que caracterizan la evaluación y la hacen peculiar con respecto a otros tipos de estudios (Pérez Juste, 2000 y 2002; De Miguel, 2000a):

- a) La recogida de información (evidencias en términos de De Miguel) sobre el programa en cuestión.
- b) La valoración a partir de criterios y referencias especificados de antemano (emisión de juicios de valor).
- c) La toma de decisiones para promover las mejoras necesarias (utilización de los resultados).

De esta manera, mediante la evaluación de programas se determinará si el programa tal como ha sido diseñado, desarrollado e implementado, ha,

está o va a generar los efectos esperados. La evaluación contribuirá a señalar los aspectos fuertes y débiles del programa en tres momentos (Mateo, 2000a):

- Antes de su **implementación**.
- Durante su **desarrollo** como fuente de mejora continua.
- Después de su **finalización** para analizar su efectividad.

Sin embargo, señala Mateo (2000a) citando a Talmage (1985) que en muchas ocasiones los prácticos de la educación se han sentido frustrados por no conocer lo que pueden obtener de la evaluación. Los autores citados señalan cinco cuestiones:

- I. *El valor intrínseco*, es decir, la bondad y adecuación del programa una vez que éste ya ha sido diseñado. Los juicios de valor acerca de esta cualidad serán realizados por expertos en la materia basándose en las exigencias científicas de los contenidos curriculares y en las necesidades educativas asociadas.
- II. *El valor instrumental*, es decir, la adecuación del programa diseñado en función de los objetivos y fines que se ha establecido.
- III. *El valor comparativo*, es decir, la comparación entre programas rivales. Esta situación puede darse cuando el programa utilizado hasta el momento ha quedado obsoleto y se piensa en renovarlo completamente o sustituirlo por otro ya existente. En ese caso se plantea la cuestión de si el nuevo programa es superior al anterior.
- IV. *El valor de idealización*, es decir, la recolección de datos durante el desarrollo del programa para su mejora continua.
- V. *El valor de decisión*, es decir, la toma de decisiones basada en los cuatro aspectos anteriores, de tal manera que si éstos han sido logrados, las decisiones serán de calidad.

2. Modelos de evaluación de programas

2.1. Concepto de modelo

A partir de las décadas de 1960 y 1970 se produjo, como se ha visto anteriormente, la eclosión de los llamados modelos de evaluación de programas. Sin embargo, no todos los autores coinciden en aceptar la mayoría de esos modelos como tales. Stake (1981), por ejemplo, señala que más bien son «persuasiones», House (1986) aduce que hay que enfatizar el término de «metáfora» y Pérez Carbonell (2000) prefiere hablar de «enfoques». Nevo (1986) señala que la mayoría de los llamados modelos no tienen el grado de complejidad y de globalidad que corresponde a un modelo.

De Miguel (2000b) define modelo desde un punto de vista científico afirmando que es una

representación en pequeño del fenómeno a investigar. Esencialmente constituye una expresión y/o representación formalizada de una teoría que consideramos adecuada para investigar un problema (p. 194).

Para que un modelo sirva de guía para una investigación debe (De Miguel 2000b):

1. Estar fundamentado en una teoría y explicitar las relaciones entre los fenómenos intervinientes en él.
2. Especificar las variables operacionalmente.
3. Precisar la validez empírica del modelo.
4. Validar y generalizar las aplicaciones del modelo.

De esta forma, un modelo hace referencia a la forma de abordar y representar la conceptualización teórica del fenómeno o problema que se quiere analizar y consecuentemente dependerá de la perspectiva o paradigma adoptado para su comprensión/explicación (De Miguel, 2000b).

2.2. Criterios para la clasificación de los modelos

Hemos creído más conveniente en este apartado especificar cuáles son los distintos criterios que pueden emplearse para clasificar los diferentes modelos de evaluación más que hacer una presentación más o menos exhaustiva y más o menos pormenorizada de los más de 40 modelos (o enfoque en muchos casos) planteados hasta el momento en la evaluación de programas. En el apartado siguiente sí que se expondrán de forma resumida algunos modelos como representativos de otros que pueden ser clasificados en el mismo ámbito según el criterio elegido.

Para clasificar un modelo de evaluación hemos de tener en cuenta el criterio en el cual nos apoyamos para tal efecto. En este sentido vamos a seguir los tres criterios señalados por De Miguel (2000b):

2.2.1. La finalidad de la evaluación

Conocer la finalidad de la evaluación supone responder a la cuestión de qué quiero (o quiere el cliente) saber con relación al programa y qué pretendo (o pretende el cliente) decidir con respecto al programa en cuestión. El marco teórico de la evaluación vendrá definido por estas preguntas. De esta forma

se distinguen cuatro tipos de modelos evaluativos según cuál sea el propósito de la evaluación.

1. Los *basados en la evaluación de resultados*, cuyo objetivo es comprobar si los objetivos y metas preestablecidos han sido logrados. Es decir, verificar la eficacia del programa.
2. Los *orientados a las audiencias implicadas*, cuyo objetivo es recolectar opiniones fundamentadas sobre la valoración que hacen del programa las audiencias implicadas (*Stakeholders*). Es decir, estimar la valoración de los implicados.
3. Los *enfocados a las decisiones*, cuyo objetivo es obtener información cualificada según la cual puedan tomarse decisiones sobre el futuro del programa. Es decir, tomar decisiones que incidan en la mejora.

Se pueden agrupar los modelos evaluativos en tres bloques paralelos a los tres objetivos básicos de la investigación en las ciencias sociales: verificar, comprender y transformar. De ahí que toda aproximación y/o clasificación de los modelos evaluativos nos remite a los clásicos paradigmas —positivistas, naturalista y crítico— de la investigación científica [...]

(De Miguel, 2000b, p. 198)

4. Los *basados en costos-resultados*, cuyo objetivo es estimar la relación entre los recursos utilizados y los resultados logrados tanto inmediatos como a un plazo más largo.

2.2.2. Los criterios de valoración

Desde este punto de vista supone que respondemos a la cuestión de qué criterios vamos a utilizar para emitir los juicios de valor acerca del programa en cuestión. En función de este criterio de clasificación se distinguen cinco tipos de modelos de evaluación:

1. Los centrados sobre la *eficacia*, en los que los criterios para valorar el programa se fundamentan en la cantidad y calidad de los objetivos logrados (esto puede hacerse evaluando el nivel de consecución de las metas establecidas, evaluando también los efectos colaterales o utilizando un proceso abierto o libre de objetivos).
2. Los fundamentados sobre la *eficiencia*, en los que el criterio para valorar el programa es la relación de los beneficios (resultados) obtenidos en función de los recursos (medios) utilizados.
3. Los *basados en los participantes*, en los que el criterio para valorar el programa es la respuesta que da a las necesidades, intereses y expectativas de los que participan en él (efectividad). Dentro de este tipo de

modelos se distinguen dos grupos: los centrados sobre los destinatarios (consumidor, cliente) y los que utilizan como criterio las expectativas de todas las audiencias implicadas.

4. Los *profesionales*, en los que los criterios para emitir los juicios de valor los establecen los propios evaluadores, normalmente suele ser debido a la falta de concreción del programa.
5. Los orientados a la *calidad total*, en los que previamente debe definirse lo que se entiende por calidad de un programa tanto en lo que se refiere a los procesos como a los resultados y en consecuencia establecer los criterios para emitir juicios de valor sobre la calidad de los productos (satisfacción de los clientes, resultados, impactos, etcétera) como los sistemas acordados para garantizar la calidad de los procesos.

2.2.3. El rol de los evaluadores

Desde este punto de vista se debe responder a la cuestión de cuál es el papel o rol que va a desempeñar el evaluador en el proceso evaluativo. Hay tres factores que van a limitar su papel en el proceso:

1. *Los modelos de evaluación ex ante/ex post*. El momento (antes o después de comenzar) en el que se encuentra el desarrollo del programa es uno de los condicionantes más importantes con respecto a la elección del modelo.
2. *De evaluación interna/externa*. La situación del evaluador con respecto al programa (pertenece al programa o no) va a determinar también el modelo de evaluación.
3. *Orientados hacia la evaluación sumativa/formativa*. Si el encargo se orienta hacia la evaluación sumativa el evaluador toma el cargo de técnico recolector de datos fiables que ayuden a tomar decisiones sobre el programa. Por contra, si se orienta hacia la evaluación formativa, el evaluador tomará el cargo de animador/facilitador/conductor de la reflexión conjunta de todos los implicados para mejorar el programa.

Los cambios paradigmáticos experimentados en la evaluación de programas, al igual que en la investigación educativa, ha hecho evolucionar de una posición cuantitativa, en la que la evaluación de un programa se reducía a la medición de sus efectos, a que sea necesario tomar en cuenta la perspectiva de los directamente implicados. De hecho, la confrontación entre la orientación nomotética o idiográfica de la evaluación de programas ha estado presente en estos últimos años, aunque cada vez son más quienes optan por posturas intermedias alejándose de los extremos.

2.3. Algunos modelos representativos de la evaluación de programas

En función de los tres tipos de criterios que hemos mencionado, De Miguel (2000b, p. 210) basándose en otro trabajo de Stecher y Davis (1988) presenta el cuadro 4.1, en el que aparecen la finalidad, el enfoque y el rol del evaluador relacionados¹.

Cuadro 4.1. Correspondencia entre modelos y diseños de evaluación

MODELO			DISEÑO		
Finalidad	Enfoque	Rol del evaluador	Plan	Metodología	Sistema de control
Evaluar resultados	Positivista/ Economicista	Experimentador	Preordenado	Experimental	<i>Comparativos:</i> • Aleatorios
Eficacia/ Eficiencia	Orientado a metas	Especialista en medida		Cuasi-experimental	• Construidos • Estadísticos
Tomar decisiones para mejorar el programa	Centrado sobre las decisiones	Experto en toma de decisiones			<i>No Comparativos:</i>
Recabar las valoraciones de los implicados	Orientado al usuario Naturalista/ Interpretativo	Colaborador Facilitador/ Asesor	Respondente/ Emergente	Selectiva Observacional	• Genéricos • Reflexivos • En sombra

Como puede observarse, en la segunda de las columnas aparecen los cinco enfoques diferentes que pueden adoptar los distintos modelos de evaluación que se han propuesto. Ya hemos mencionado previamente que no se van a presentar de forma exhaustiva ni pormenorizada los modelos de evaluación de programas; sin embargo, sí creemos conveniente presentar al menos uno, de forma resumida, por cada uno de los cinco enfoques planteados.

2.3.1. Modelos de evaluación de costos-resultados

En este primer ejemplo, en vez de mostrar un modelo, vamos a señalar cuáles son las características generales de los modelos encuadrados en este apartado. Como señala Carballo (1990a) en un principio este tipo de modelos fue muy utilizado por administradores y legisladores sociales para determinar la

inversión económica que supone la puesta en marcha de un programa. Estos gastos pueden ser diversos:

- Costos para el programa:*
1. Periódicos: recursos materiales y humanos.
 2. De capital: edificio, mobiliario, equipamiento...

- Costos para el usuario:*
3. Directos: matrícula, material didáctico, transporte.
 4. Indirectos: pérdida de oportunidades...

A partir de la evaluación de distintos programas se podía decidir aquel que exigía un menor costo, sin embargo, pronto se dieron cuenta que la relación de los costos con los efectos educativos producidos también era importante. Consecuentemente surgió la necesidad de valorar los costos en relación con los beneficios que reporta. Siguiendo a Carballo (1990a), se pueden distinguir cuatro tipos de estudios dentro de este enfoque:

1. Aquellos que han relacionado los costos con los beneficios en el que tanto unos como otros se traducen a términos monetarios para determinar su ratio.
2. Aquellos que han relacionado los costos con la efectividad cuya finalidad es determinar el costo económico que supone conseguir el producto en cuestión.
3. Aquellos que han relacionado los costos con la utilidad son parecidos al anterior, en donde los criterios de eficacia se asocian a medidas de utilidad del programa.
4. Aquellos que han relacionado los costos con la viabilidad en los que se pretende especificar la posibilidad/viabilidad presupuestaria de uno o varios programas.

En definitiva, todos estos modelos pretenden valorar la adecuación y la rentabilidad de las inversiones realizadas.

2.3.2. El modelo de consecución de metas de Tyler

Como ya se ha mencionado en los capítulos anteriores, Tyler fue el precursor de la evaluación educativa y el primero que propuso un modelo de evaluación propiamente dicho. Tyler definía la evaluación como «el proceso de determinar hasta qué punto los objetivos educativos han sido alcanzados mediante los programas de currículum y enseñanza» (Tyler, 1950, p. 69). Por lo tanto, los objetivos se convierten en los criterios en los que fundamentar la evaluación.

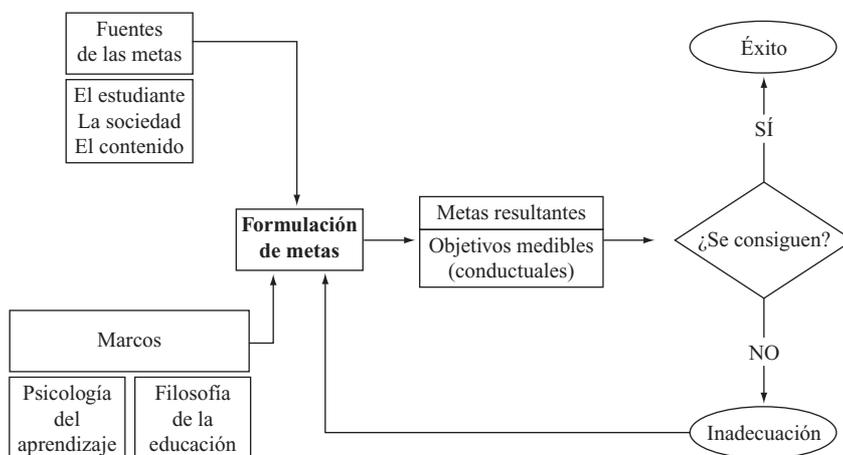
Este modelo tiene su origen dentro de una concepción tecnológica de la enseñanza en la que hay una base lógica organizada a partir del desarrollo de

un programa. Se considera la evaluación como una etapa fundamental en el desarrollo (Tejada, 1999). El procedimiento para evaluar un programa sería el siguiente (Stufflebeam y Shinkfield, 1987, p. 93):

1. Establecer las metas u objetivos. Para ello, las fuentes serían el alumno, la sociedad y la naturaleza de los contenidos, por un lado, y de la filosofía y la psicología que las engloba, por otro.
2. Ordenar lo objetivos en amplias clasificaciones.
3. Definir los objetivos en términos de comportamiento, es decir, como objetivos operativos.
4. Establecer situaciones y condiciones según las cuales puede ser demostrada la consecución de los objetivos.
5. Explicar los propósitos de la estrategia al personal más importante en las situaciones más adecuadas.
6. Escoger o desarrollar las medidas técnicas apropiadas.
7. Recopilar los datos de trabajo (en el caso de los programas educativos, deben referirse al trabajo de los estudiantes).
8. Comparar los datos con los objetivos de comportamiento.

La figura 4.1, tomada de Monedero Moya (1998, p. 75) esquematiza el proceso de evaluación propuesto por Tyler.

Figura 4.1. Proceso de evaluación de Tyler



Para Tejada (1999), las ventajas que aporta este modelo son, por una parte, que desplaza la atención también a otros aspectos del programa aparte del alumno y que permite tomar decisiones de mejora en futuras aplicaciones del programa. No obstante, también pueden identificarse limitaciones que en su

día catalizaron el proceso de eclosión de nuevos modelos evaluativos. Entre otros podemos señalar que:

1. Mediante este modelo no sabemos cómo se consiguen o no los objetivos.
2. Exige la operativización de todos los objetivos, aspecto harto difícil si no imposible en muchas situaciones.
3. La evaluación es considerada como un proceso terminal, lo que unido a la crítica anterior la convierte en una evaluación con función sumativa y no formativa.
4. No tiene en cuenta los efectos colaterales, secundarios o no esperados, y es bien sabido que éstos a veces son tan importantes o más que los preestablecidos.
5. El rendimiento del estudiante es tomado como criterio exclusivo del éxito o fracaso del programa.

2.3.3. El modelo CIPP (facilitador de toma de decisiones)

El modelo CIPP (*Context, Input, Process, Product*) fue propuesto inicialmente por Stufflebeam y Guba a principios de la década de 1970, pero posteriormente ha sido Stufflebeam quien ha seguido profundizando en el modelo y hoy en día muchos le otorgan su autoría en exclusividad.

Este modelo orientado a la gestión y a la toma de decisiones es según muchos autores el modelo más utilizado hasta la fecha en la evaluación de programas.

Este modelo comprensivo considera la evaluación como un proceso continuo y cíclico que debe llevarse a cabo de forma sistemática. Este proceso evaluativo necesita de tres fases o etapas:

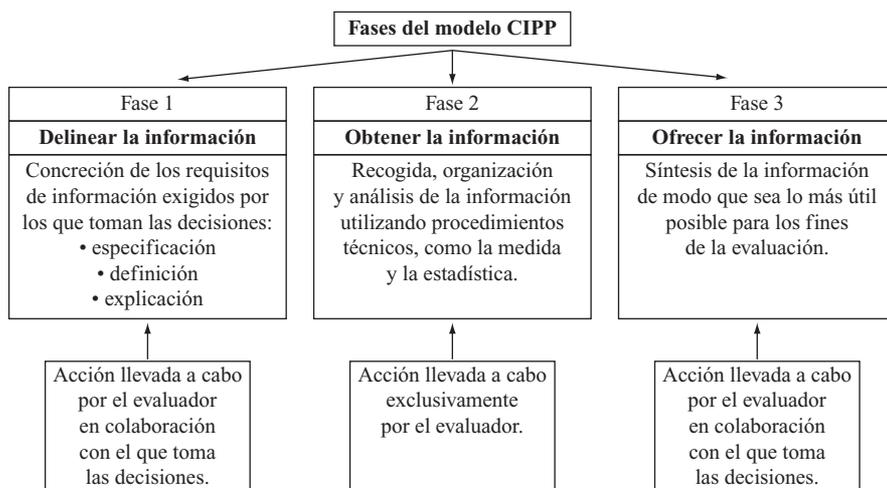
1. Delineación/delimitación, en la que los evaluadores se reúnen con las personas encargadas de tomar decisiones para identificar la información necesaria.
2. Obtener la información previamente seleccionada.
3. Devolución de la información a los responsables de las decisiones.

La figura 4.2, tomada de Monedero Moya (1988, p. 88), muestra claramente el proceso en sus tres fases.

Cuando el modelo CIPP es aplicado específicamente al ámbito educativo posee cuatro estadios característicos que constituyen en sí mismos el núcleo del programa. Éstos son:

1. *La evaluación del contexto (Context)*, que implica analizar todas las circunstancias alrededor del programa. Se trata, pues, de una fase diagnóstica cuyo objetivo es definir y analizar el escenario donde se

Figura 4.2. Fases del modelo de evaluación CIPP



encuadran los elementos relevantes del contexto. Fundamentalmente, se analiza la adecuación de los objetivos del programa con las necesidades apreciadas a partir del análisis realizado (Mateo, 2000a).

2. *La evaluación de las entradas (Input)*, en la que se recogen datos para identificar la forma de utilizar óptimamente los recursos existentes en aras de conseguir las metas establecidas.
3. *La evaluación del proceso (Process)*, cuyo objetivo consiste en establecer la congruencia entre lo programado y la realidad. Para ello, se deberá, por una parte, identificar o predecir los defectos del diseño de implementación del programa y, por otra, proveer de forma continuada la información para la toma de decisiones durante el desarrollo del programa.
4. *La evaluación del producto (Product)*, que exige la obtención de información sobre los efectos que ha producido el programa para tomar decisiones finales para continuar, modificar o desechar el programa.

El cuadro 4.2, adaptado de Monedero Moya (1999, p. 90), sintetiza lo que acabamos de presentar.

2.3.4. El modelo de evaluación respondiente de Stake

En la década de 1960, Stake (1967) presentó el modelo de congruencia-contingencia clasificado entre los modelos más tradicionales. Sin embargo, con el paso de los años, el mismo autor fue publicando nuevos escritos remodelando su primera propuesta. En 1975 publicó lo que él denominó como la evaluación respondiente en la que pasaba de una propuesta anterior positivista a una de corte humanístico (Mateo, 2000a).

Cuadro 4.2. Tipos de evaluación de acuerdo con las situaciones de decisión, según el modelo CIPP

	Evaluación de contexto	Evaluación de las entradas	Evaluación de proceso	Evaluación de producto
<i>Finalidad</i>	Planificar las decisiones para determinar los objetivos.	Estructurar las decisiones para diseñar los procedimientos de instrucción.	Instrumentalizar las decisiones para utilizar, controlar y mejorar estos procedimientos.	Reciclar las decisiones para juzgar y reaccionar ante los resultados.
<i>Métodos</i>	Descriptivos y comparativos.	Analizar las posibilidades disponibles y las estrategias necesarias para lograr los objetivos.	Describir acontecimientos y actividades del procedimiento para descubrir y anticipar deficiencias.	Similares a los del evaluador de consecución de metas, enfatizando los criterios extrínsecos.
<i>Conclusión</i>	Identificar objetivos específicos para diseñar un programa.	Proporcionar información respecto a cómo han de emplearse los recursos.	Controlar los procedimientos que se llevan a cabo.	Ayudar a decidir si hay que continuar, terminar, modificar o reenfocar un programa.

Este modelo respondiente acentúa la necesidad de responder a las necesidades del usuario. De tal manera que es necesario el intercambio constante de comunicación entre el evaluador y las audiencias con el propósito de descubrir, investigar y solucionar problemas. Cuando Stake propone su nuevo modelo, podemos caracterizar sus concepciones evaluativas como (Stufflebeam y Shinkfield, 1987; Tejada, 1999):

- Las evaluaciones deben ayudar a los usuarios a observar y mejorar su práctica respondiendo de forma holística a las demandas presentadas por las distintas audiencias.
- Los evaluadores deben describir programas teniendo en cuenta los antecedentes, las operaciones y transacciones y los resultados.
- Los efectos colaterales del programa deben ser tan buscados como los efectos esperados.
- Los evaluadores deben considerar las perspectivas y los puntos de vista de las personas implicadas en el programa.
- Los evaluadores deben evitar la presentación de informes resumidos, presentando descripciones completas recogiendo las perspectivas de todos los interesados.

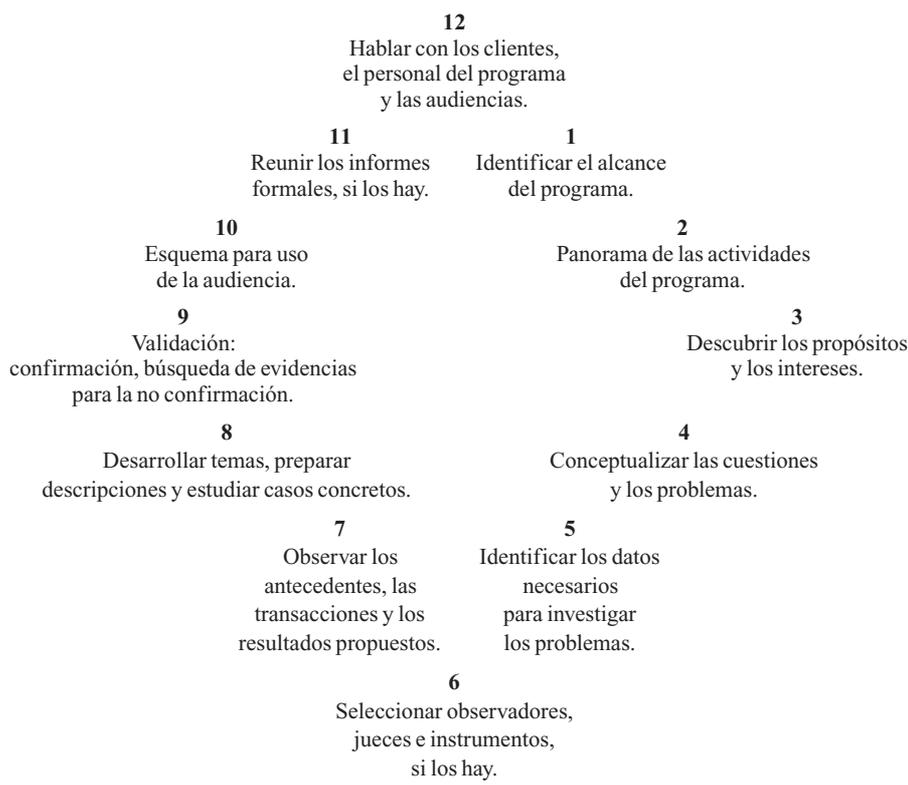
- Los evaluadores deben utilizar una variedad de métodos para la recogida de información (los tests estandarizados no son suficientes).

Este modelo se fundamenta en cuatro grandes elementos (Stake, 1975):

- a) *Los problemas.* Para detectar, comprender y definir las áreas problemáticas la comunicación con los evaluados es fundamental.
- b) *El diseño para la recogida de la información* es el factor organizador del modelo.
- c) *Los observadores* conforman la base para la recogida de información significativa.
- d) *La validación del proceso* de evaluación supone una recogida amplia de la información (utilizando múltiples fuentes) realizada por varias personas y debidamente cruzada.

Por último, para explicar de forma gráfica las fases de la estructura funcional de la evaluación respondente, Stake utiliza la figura de la «evaluación del reloj» (véase cuadro 4.3).

Cuadro 4.3. La evaluación del reloj (estructura funcional de la evaluación respondente)



El funcionamiento de este reloj es peculiar dado que las agujas pueden colocarse en cualquier hora y puede moverse tanto en un sentido como en el otro.

2.3.5. El modelo de evaluación iluminativa de Parlett y Hamilton

El modelo de evaluación iluminativa de Parlett y Hamilton (1972, 1983) surge como reacción a los modelos de corte positivista situándose en la corriente naturalista-interpretativa.

El objetivo del modelo es obtener una visión lo más completa y global posible del programa; para ello, se realiza un estudio intensivo teniendo en cuenta todos sus componentes con el propósito de «iluminar» todos sus problemas, cuestiones y características. Se insiste en el carácter complejo y dinámico de interacciones que es la educación y, a través de la evaluación iluminativa, estas interacciones deben ser observadas de manera holística y subjetiva, ya que no pueden ser divididas en variables para su medición objetiva (Mateo, 2000a).

La evaluación iluminativa no es un método regularizado sino una estrategia global que puede adoptar diversas formas. Pretende ser adaptable para descubrir las dimensiones, metas y técnicas de evaluación, y ecléctica en el sentido de proporcionar al evaluador una serie de tácticas investigativas, que dependerán del problema que deba investigarse. La preocupación del evaluador será familiarizarse con la realidad cotidiana en la cual está investigando, sin intentar manipular, controlar o eliminar las variables situacionales, sino abarcar toda la complejidad del contexto.

(Tejada, 1999, p. 51)

Señala este mismo autor que las finalidades de la evaluación iluminativa son las siguientes (p. 50):

- Estudiar el programa, teniendo en cuenta cómo funciona, cómo influyen en él las situaciones educativas en las que se aplica, ventajas y desventajas, cómo se ven afectadas las tareas intelectuales y las experiencias académicas de los participantes.
- Descubrir y documentar participando en el programa como formador o como participante.
- Discernir y comentar las características más significativas del programa, las concomitancias recurrentes y los procesos críticos.

Las características más relevantes del modelo evaluativo las podemos resumir de la manera siguiente:

- Tiene carácter holístico y tiene en cuenta el contexto en el que se desarrolla.

- Se preocupa más por la descripción y la interpretación que por la medición y la predicción.
- Se orienta al estudio de los procesos más que a los productos.
- Se desarrolla bajo condiciones naturales y no bajo condiciones experimentales.
- Se utilizan la entrevista y la observación como métodos principales para la recogida de información.
- Se emplean diseños flexibles y emergentes que van acomodándose con el desarrollo de la acción.
- Se rechazan los modos y las formas de actuar con los que el evaluador no aceptaría ser investigado.
- Se deben comprender las perspectivas de los implicados sin manipular el proceso.
- Se reconocen los valores y los derechos de los informantes.
- Se asegura que los evaluados sientan que su participación en la evaluación ha hecho movilizar su pensamiento y que su comprensión se ha desarrollado.
- Se tienen muy en cuenta el contexto de aprendizaje (el contexto material, psicológico y social en el que trabajan conjuntamente formadores y participantes) y el sistema de instrucción (el conjunto de elementos que componen un plan coherente que al ser aplicado sufre cambios debido al medio, a los participantes y a los formadores).

Parlett y Hamilton diferencian tres fases para llevar a cabo una evaluación iluminativa de un programa:

FASE 1: *Observación*, en la que el evaluador debe familiarizarse con el programa para tener una visión global de éste y de su contexto.

FASE 2: *Ampliación de la indagación*. Si en la fase de observación el propósito estaba en saber si el programa funcionaba, en la fase de ampliación de la indagación interesa saber por qué funciona. Para ello, se realiza una focalización progresiva e intensiva del programa.

FASE 3: *Explicación*, en la que se exponen los principios subyacentes y se explican los modelos causa-efecto a los implicados.

Con la presentación de estos cinco modelos de evaluación de programas representativos de los distintos enfoques evaluativos terminamos con este apartado. Sin embargo, antes nos parece pertinente presentar la siguiente cita de Martínez Mediano (1996, pp. 135-136), ya que resume perfectamente lo que debería ser un modelo de evaluación adecuado:

La evaluación es la recogida de información sistemática acerca de un programa elaborado para atender a determinadas necesidades, centrada esta recogida de información en la cali-

dad del propio programa, en su administración, en su implementación, en su desarrollo y en sus resultados, con el objetivo de comprobar si ha sido bien elaborado, bien implantado, bien desarrollado y si consigue los resultados previstos comparados con las metas pretendidas, a partir de lo cual el investigador evaluativo o evaluadores emitirá un juicio evaluativo claro basado en la información recogida de modo riguroso en todas las etapas y aspectos mencionados, que posibilite la toma de decisiones pertinentes para el cambio y mejora del programa. En esto coinciden la mayoría de los enfoques evaluativos.

3. El diseño de evaluación

3.1. Concepto de diseño

Una vez elegido el modelo o enfoque a utilizar en la evaluación del programa, debemos optar por el diseño más adecuado para llevar a cabo la evaluación.

El diseño de una evaluación alude a la planificación del proceso mediante el cual se va a llevar a cabo el estudio de un fenómeno que ha sido modelizado de antemano. Debemos, pues, elaborar un plan en el que se especificarán la estructura, los procedimientos y los sistemas de control que se seguirán para dar respuesta fiable y válida a los interrogantes planteados. El diseño, por lo tanto, hace referencia al proceso metodológico que se debe seguir en la práctica, ofreciendo al evaluador una lógica científica que aporte validez tanto a los procesos de evaluación como a los resultados (De Miguel, 2000b).

En todos los casos, independientemente del modelo elegido, un diseño deberá atender a los aspectos siguientes (Tejedor, 2000):

- El establecimiento de las tareas a evaluar.
- El establecimiento de los criterios de evaluación.
- La elección de las estrategias, métodos y técnicas para la recogida de la información.
- El análisis de la información recogida.
- La toma de decisiones según la información analizada.

3.2. Criterios para la clasificación de los diseños

Como señala De Miguel (2000a y b), los diseños evaluativos pueden ser clasificados teniendo en cuenta distintos criterios. Así, por ejemplo, Tejedor (2000) citando a Carballo (1990b) distingue entre diseños emergentes *versus* prefijados, diseños de evaluaciones formativas *versus* sumativas y diseños teniendo en cuenta distintos tipos de evaluaciones (evaluación de necesidades, evaluación del diseño o evaluación de entrada, evaluación de la implementación, evaluación de resultados y metaevaluación). De Mi-

guel (2000a y b) presenta los tres criterios más divulgados para clasificar los diseños:

1. *En función del grado de planificación o del enfoque teórico adoptado* se distinguen los diseños preordenados (o prefijados) y los diseños respondentes (o emergentes) basándose en los paradigmas cuantitativo y cualitativo. En un diseño preordenado se siguen los procedimientos y la lógica científicos aplicados a la investigación social en la que todos los elementos del proceso (desde los objetivos hasta los criterios de emisión de juicios) están especificados previamente. Los diseños respondentes, por el contrario, se basan en un marco teórico flexible en el que los criterios evaluativos provienen de las valoraciones de las personas implicadas en función de sus expectativas y necesidades. Es por ello por lo que se necesita un diseño flexible y emergente que vaya acomodándose al desarrollo de la propia evaluación.
2. *En función de la metodología o del sistema de control* que permitan las variables del programa, De Miguel (2000a y b) siguiendo a Anguera (1995) distingue los diseños observacionales, selectivos, cuasiexperimentales y experimentales. En este caso se está aludiendo al concepto de diseño² con terminología positivista y su clasificación se realiza a partir de las condiciones que debe cumplir toda investigación: aleatorización de los sujetos, manipulación de la variable independiente y grupo de control. En función del grado o nivel de garantía de cumplimiento de las condiciones citadas distinguen los cuatro tipos de diseños mencionados:
 - a) *Los diseños experimentales* pueden aplicarse en los programas en los que es posible el empleo de metodología experimental debido a que:
 - Las estrategias de intervención están muy estructuradas y son manipulables.
 - Existe la posibilidad de utilizar el grupo de control.
 - Los sujetos pueden ser elegidos aleatoriamente.
 - b) *Los diseños cuasiexperimentales* pueden aplicarse en programas donde existe un gran control sobre la intervención, pero se efectúan sobre sujetos o grupos no aleatorios o en contextos de difícil control.
 - c) *Los diseños selectivos* pueden aplicarse en programas con un cierto grado de formalización o estructuración de la intervención, aunque con un control débil sobre ella y con sujetos elegidos de forma no aleatoria la mayoría de las veces. En estos casos se emplean diseños prefijados en el que se especifican los procedimientos a utili-

zar para que aprovechando la variabilidad existente se puedan establecer las diferencias entre grupos.

- d) *Los diseños observacionales* pueden aplicarse en programas individualizados puestos en marcha en contextos naturales en donde el control sobre la intervención es muy bajo y los grupos, al ser naturales, no han sido elegidos aleatoriamente. En este caso se utilizan diseños muy flexibles y con instrumentos elaborados ad hoc.

Por su parte Tejedor (2000, p. 329) presenta otra clasificación de los diseños de investigación aplicables a la evaluación de programas más exhaustiva que la anterior que se presenta en el cuadro 4.4.

No obstante, en el capítulo 6 («Diseños de investigación aplicados a la evaluación») presentamos de manera más exhaustiva los diseños descriptivos, correlacionales, experimentales, etnográfico, estudio de casos e investigación-acción.

Cuadro 4.4. Diseños de investigación utilizables en la Evaluación de Programas

1. *Diseños para la investigación cuantitativa*
 - a) Diseños descriptivos-comparativos (estudios tipo *survey*)
 - b) Diseños longitudinales:
 - i. secuenciales (diacrónicos)
 - ii. transversales (sincrónicos)
 - c) Diseños correlacionales:
 - i. correlación y regresión
 - ii. técnicas de clasificación
 - d) Diseños causales
 - e) Diseños experimentales:
 - i. preexperimentales
 - ii. cuasiexperimentales
 - iii. experimentales de campo (intergrupo e intragrupo)

 2. *Diseños para la investigación cualitativa*
 - a) Observación naturalista
 - i. estudios de casos
 - ii. estudios del grupo
 - iii. estudios de comunidades (etnográficos)
 - iv. estudios de situaciones o contextos
 - b) Diseños participativo-críticos
 - i. investigación-acción
-

3. *En función del sistema de control sobre la información recabada.* Tal y como señala De Miguel (2000a y b), los diseños presentados en el apartado anterior hacen referencia únicamente a los resultados e impactos del programa, sin tomar en consideración que la evaluación debe también atender otros aspectos del programa como la detección de necesidades, el diseño del programa efectuado y la implementación de la práctica. Lógicamente, también en estas etapas se necesita un diseño y un sistema de control para garantizar la objetividad y validez de los juicios de valor. Se tendrá que determinar el procedimiento mediante el cual se va a contrastar la información recogida sobre el programa y para ello hay que recurrir a otros sistemas de control, aunque sean menos potentes, diferentes a los experimentales. De Miguel (2000b), citando a Rossi y Freeman (1993), clasifica los diseños utilizando dos criterios: la existencia o no de grupo de control y el procedimiento utilizado para ello, obteniéndose dos diseños:

- a) *Los diseños comparativos*, en los que la información recogida sobre el programa es confrontada con otros grupos que no están recibiendo el programa. La calidad de la selección de estos grupos (controles aleatorios, contruados o estadísticos) condiciona la calidad del sistema de control utilizado.
- b) *Los diseños no comparativos*, en los que la información recogida sobre el programa no puede ser contrastada con grupos control por su inexistencia. En estos casos, muy típicos, en las fases intermedias de la evaluación deben arbitrarse otro tipo de criterios con los que contrastar la información. Caben diferentes alternativas:
 - *Controles genéricos*, cuando se utilizan como criterios de comparación las normas de la población de referencia.
 - *Controles reflexivos*, cuando se utilizan como criterios distintas mediciones realizadas sobre el mismo grupo.
 - *Controles en la sombra*, cuando se utilizan como criterios las valoraciones que hacen sobre el programa los participantes y no participantes.

Las correspondencias entre los distintos tipos de diseños clasificados en función de los tres criterios mencionados son bastante evidentes: mientras que un diseño preordenado opta por diseños experimentales o cuasiexperimentales con controles comparativos, un diseño emergente necesita diseños observacionales y sistemas de control no comparativos.

4. Elección del modelo y del diseño de evaluación

Evidentemente, la relación entre el modelo de evaluación (dimensión teórica) y el diseño de evaluación (dimensión práctica) es necesaria. En un primer momento deberemos elegir el modelo que va a servir de guía del proceso evaluativo basándonos en tres criterios fundamentales (véase cuadro 4.1) (De Miguel, 2000b):

- El propósito de la evaluación.
- Dada la naturaleza del programa y el propósito de la evaluación, la aproximación paradigmática consecuente.
- El rol que debe/puede asumir el evaluador.

Como puede observarse en el cuadro 4.1 existe una correspondencia basándonos en los tres criterios mencionados, de tal manera que no tiene sentido, por ejemplo, recabar las valoraciones de los implicados desde un enfoque positivista en el que el evaluador asume el rol de experimentador. Debe haber, pues, una congruencia entre los tres criterios.

Tras la elección del modelo, habrá que optar por algún diseño evaluativo. Hemos visto que los diseños pueden clasificarse según tres criterios que el evaluador deberá tener en cuenta a la hora de seleccionar el diseño más adecuado a sus necesidades. Sin embargo, la elección del modelo de evaluación ya ha prefijado de antemano las características del diseño puesto que éste debe ser congruente con el modelo. El cuadro 4.1 muestra la congruencia que debe darse entre el modelo y el diseño. Por ejemplo, si hemos elegido una evaluación cuyo propósito es evaluar los resultados de eficacia del programa desde un enfoque positivista y en el que hemos asumido el rol de experimentador, lo congruente sería que eligiéramos un plan preordenado con metodología experimental o cuasiexperimental, utilizando sistemas de control comparativos.

Preguntas y ejercicios

1. Define lo que es un programa.
2. ¿Cuáles son los componentes básicos que caracterizan a la evaluación de programas?
3. En opinión de De Miguel, ¿qué criterios hay que tener en cuenta para realizar la clasificación de los modelos de evaluación?
4. ¿Cuáles son las aportaciones y las limitaciones del modelo de consecución de metas de Tyler?
5. ¿Qué significan cada una de las siglas del modelo CIPP de Stufflebeam?
6. ¿Cuáles son las características de la evaluación responderte de Stake?
7. ¿Cuáles son las características de la evaluación iluminativa?

8. ¿Están relacionados el modelo y el diseño a utilizar en la evaluación de un programa? ¿Cómo? ¿En qué hay que basarse para la elección de un modelo y un diseño de evaluación?

Notas

1. La parte derecha del cuadro relativa a los diseños de la evaluación se explicará más adelante, pero como es parte del cuadro, se ha creído conveniente presentarlo en este momento.
2. Desde este punto de vista, Tejedor (2000, p. 321) define el término diseño como «planificación puntual del análisis de los efectos de la intervención realizada (los diseños o diseño de investigación)».

5. Ámbitos de la evaluación educativa

Los ámbitos de la evaluación educativa hacen referencia a los distintos campos o subcampos de estudio. Si en el origen de la evaluación educativa su único campo se circunscribía a los alumnos, hemos visto que en su proceso evolutivo se ha ido ampliando su ámbito de acción hacia otros aspectos de la realidad educativa. Por ello, en este capítulo vamos a exponer las líneas principales de los distintos ámbitos de la evaluación educativa. Los ámbitos básicos de la evaluación educativa son la mencionada evaluación de alumnos, la evaluación del sistema educativo, la evaluación de programas ya tratada en el capítulo anterior, la evaluación de profesores y la evaluación de centros.

1. Evaluación del alumnado

Tal y como se ha señalado de manera reiterada en las páginas previas, durante muchos años el único objeto educativo susceptible de ser evaluado fue el alumnado. Hemos apuntado también que el foco de atención de la evaluación ha cambiado y hoy en día se tienen en cuenta otros ámbitos educativos. Sin embargo, ello no significa que la evaluación del alumnado haya dejado de ser importante. Es más, para muchos la evaluación del alumnado es y será siempre el ámbito educativo más importante. No obstante, la concepción de la evaluación del alumnado, así como los contenidos a evaluar y las formas de llevarlo a cabo, han ido variando a lo largo del tiempo. Si a principios del siglo XX se evaluaban (se medían) únicamente los conocimientos de los alum-

nos, con objeto de certificar esos saberes y el instrumento utilizado era el examen oral, ha habido una evolución en la que se ha pasado del examen oral al escrito, después a las pruebas objetivas. Eso en lo que se refiere a la técnica o instrumento utilizado para evaluar. Pero además, se ha pasado de una evaluación de carácter puramente sumativo a una evaluación con carácter marcadamente formativo. Hay que añadir que se han ido ampliando los contenidos a evaluar y como consecuencia de todo este desarrollo, en los últimos años se han criticado las formas de evaluar tradicionales y se ha abogado por otras maneras de evaluar.

No obstante, para realizar el análisis de la evaluación, hay que tener en cuenta la acción educativa y el concepto de aprendizaje que resulta de ella. Es decir, el modelo de enseñanza-aprendizaje utilizado estará en consonancia con el modelo de evaluación del alumnado aplicado. En este sentido, Bélair (2000) presenta en la figura 5.1 las relaciones que se han establecido entre determinadas corrientes de pensamiento actuales sobre la enseñanza, el aprendizaje y la evaluación.

Como puede observarse en el eje vertical izquierdo aparecen las corrientes de pensamiento más tradicionales. En ésta, el profesor ostenta el poder y el alumno escucha, repite, copia y memoriza, por lo que la evaluación es sobre todo sumativa, responde a criterios de uniformidad, pretende clasificar a los alumnos basándose en las puntuaciones obtenidas en los exámenes en los

Figura 5.1. Clasificación de las opciones pedagógicas y de evaluación. (Bélair, 2000, p. 14)

SUMATIVA	MODELO BASADO EN LA AUTORIDAD Estrategia militar	MODELO BASADO EN LAS DESTREZAS Estrategia abierta	FORMATIVA
	Orientado por el profesor Pedagogía de la huella Evaluación sanción	Orientado por el proceso Pedagogía constructiva Gestión del error	
	Estrategia militar Alumno pasivo	Estrategia centrífuga Alumno activo	
	DISMINUCIÓN DE LAS DIFERENCIAS INDIVIDUALES Estrategia popular	ÉNFASIS EN LAS DIFERENCIAS INDIVIDUALES Estrategia libre	
	Orientado por el producto Pedagogía del conocimiento Evaluación nota	Orientado por el alumno Pedagogía humanística Autoevaluación	
	MODELO BASADO EN EL PRESTIGIO	MODELO BASADO EN LA DISPONIBILIDAD	

que se le exige la repetición de los contenidos (sobre todo cognitivos) transmitidos durante el proceso de enseñanza.

En el eje vertical derecho aparecen las corrientes de pensamiento predominantes en la actualidad. En este caso, el profesor, más que ostentar el poder y transmitir conocimientos, adopta un papel de receptor y de disponibilidad. Se considera que cada alumno es diferente y que a través de su estilo, sus percepciones y su herencia cultural le hacen entender el aprendizaje desde distintos ángulos. La evaluación es sobre todo formativa.

Siguiendo esta misma línea, Mateo (2000a) señala que en los últimos diez años se han dado una serie de cambios en cuanto a la evaluación del aprendizaje del alumnado. Apunta tres cambios:

- a) *Cambios en el enfoque de enseñanza-aprendizaje*: del énfasis en la enseñanza al énfasis en el aprendizaje. Hoy en día, debido a la nueva cultura del usuario/alumno, a la petición de cuentas desde las administraciones y las demandas de la sociedad, la tendencia se centra en evaluar los resultados de los aprendizajes del alumnado, obviando los medios y los recursos utilizados.
- b) *Cambios en los contenidos sujetos a evaluación*. A los contenidos conceptuales clásicos hasta la fecha se han unido los contenidos procedimentales y actitudinales y éstos también deben ser evaluados. De la misma manera deben evaluarse otro tipo de habilidades, capacidades y valores de tipo transversal. En este sentido, existe una preocupación mayor por cómo maneja el alumnado la información que por conocer cuánta información ha acumulado.
- c) *Cambios en la lógica de la evaluación de los aprendizajes*. De una evaluación de carácter sumativo se ha pasado a una evaluación de carácter formativo. Se ha pasado de una evaluación centrada en los productos a una evaluación centrada en los procesos.

Las implicaciones de las nuevas teorías sobre los aprendizajes en la evaluación de los aprendizajes del alumnado son ineludibles. En el cuadro 5.1 se presenta una síntesis de las nuevas concepciones pedagógicas y de sus implicaciones para la evaluación.

1.1. Modelos de evaluación del aprendizaje de los alumnos

En la actualidad, se distingue el modelo de evaluación tradicional del modelo de evaluación alternativa (véase figura 5.2).

Durante prácticamente todo el siglo xx, la evaluación tradicional ha sido la preponderante (y todavía continúa siéndolo). El profesor utilizaba para la recogida de información pruebas construidas por él mismo o tests estandarizados. Este tipo de evaluación proporciona información cuantitativa válida para realizar comparaciones entre diferentes sujetos. Mediante este tipo de

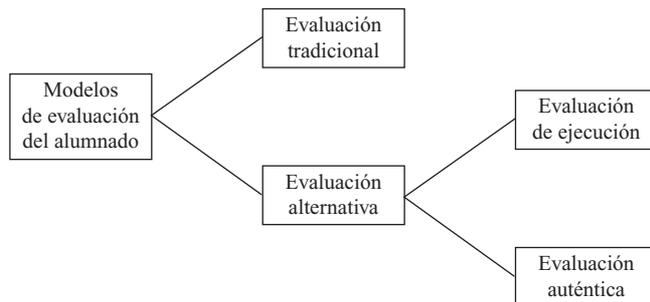
Cuadro 5.1. La evaluación del aprendizaje a la luz de las actuales concepciones sobre el aprendizaje. (Tomado de Cabrera 2000a, p. 123)

Teorías del aprendizaje	Implicaciones para la evaluación
<p>El conocimiento es algo que se construye. El aprendizaje es un proceso de creación de significados a partir de los conocimientos previos: es un proceso de transformación de las estructuras cognitivas del alumnado como consecuencia de nuevos conocimientos.</p>	<ol style="list-style-type: none"> 1. Promover acciones evaluativas que pongan en juego la significatividad (funcionalidad) de los nuevos aprendizajes. 2. Evitar los modelos memorísticos en los que sólo se pone de manifiesto la capacidad para reconocer o evocar. 3. Promover actividades y tareas de evaluación que tengan sentido para el alumnado.
<p>Hay variedad en los estilos de aprendizaje, la capacidad de atención, la memoria, el ritmo de desarrollo y las formas de inteligencia.</p>	<ol style="list-style-type: none"> 1. Promover distintas formas de evaluación. 2. Dar oportunidades para revisar y repensar. 3. Utilizar procedimientos que permitan al alumnado aprender a construir su forma personal de aprender, a manejar autónomamente procedimientos de evaluación y corregir los errores que pueda detectar.
<p>Las personas tienen una ejecución mejor cuando conocen la meta, observan modelos y saben los criterios y estándares que se tendrán en cuenta.</p>	<ol style="list-style-type: none"> 1. Promover que el alumnado haga suyos los objetivos del aprendizaje y los criterios que se van a utilizar para evaluarlos. 2. Proporcionar una amplia gama de modelos de ejemplo sobre trabajos de los alumnos para discutir sus características. 3. Discutir los criterios que se utilizarán para juzgar la ejecución.
<p>Se reconoce que el conocimiento y la regulación de los propios procesos cognitivos son la clave para favorecer la capacidad de aprender a aprender. Es importante saber manejar su propio proceso de aprendizaje.</p>	<ol style="list-style-type: none"> 1. Promover la autoevaluación. 2. Estimular los procesos de coevaluación entre el profesor y el alumno y entre éstos entre sí. 3. Ofrecer retroalimentación tanto de los errores como de los aciertos. 4. Hacer tomar conciencia al alumnado de lo que ha aprendido así como de los procesos que le han permitido construir nuevos aprendizajes.
<p>La motivación, el esfuerzo y la autoestima afectan al aprendizaje y la ejecución de la persona.</p>	<ol style="list-style-type: none"> 1. Atribuir los fracasos a razones temporales y externas y los éxitos a razones internas y perdurables. 2. Estimular relaciones entre el esfuerzo y los resultados. 3. Incorporar de manera natural tareas de evaluación durante el proceso de enseñanza aprendizaje que puedan servir al alumnado para tomar conciencia de lo que ha aprendido y de las dificultades o lagunas que todavía tiene.

evaluación no se sabe muy bien cómo aprenden los alumnos y cómo les afecta la aplicación del currículum. Además, muchas de las acciones que se pide realizar a los alumnos en el proceso evaluativo no tienen relación con el mundo real (Mateo, 2000a).

En contraposición a este tipo de evaluación, ha surgido una nueva forma de evaluación denominada alternativa en la que se da importancia a la observación directa del trabajo del alumnado y de sus habilidades (Mateo, 2000a).

Figura 5.2. Modelos de evaluación del alumnado



En principio, cualquier modelo de evaluación que utilice instrumentos o estrategias para la recogida de información distintos a los exámenes tradicionales (pruebas objetivas, tests estandarizados, etc.) podría ser considerado como evaluación alternativa. Prácticamente en todos estos casos, el alumno debe ejecutar algo. Pero no todas las ejecuciones pueden ser consideradas como la denominada evaluación auténtica. Para que se dé ésta, además de implicar al alumnado en la realización de algún tipo de ejecución, esta ejecución debe tratarse de alguna tarea perteneciente a la vida real o de una gran similitud con ésta. Para que se dé la evaluación auténtica, Bélair (2000, pp. 53-54) señala que deben tenerse en cuenta los aspectos siguientes:

- a) Debe implicar al alumnado en todas las etapas del proceso evaluativo.
- b) El alumno se sentirá implicado en el proceso de evaluación sólo si tiene confianza en el profesor, si tiene conciencia de lo que le aportará y a condición de que su valoración sea tenida en cuenta de la manera más equitativa posible.
- c) La autoevaluación es uno de los instrumentos más significativos para tener en cuenta estrategias de aprendizaje evidenciadas por el alumno.
- d) El proceso de evaluación compromete al docente a una serie de cambios, sobre todo en el poder y en la supremacía, para dejar sitio a la colaboración y a la negociación en el acuerdo y en la decisión.

En el cuadro 5.2, basado en las características señaladas por Wolf y Reardon (1996) y Mateo (2000a), se resumen los aspectos más significativos de cada uno de los dos enfoques.

Cuadro 5.2. Características de la evaluación alternativa y tradicional

	Evaluación alternativa	Evaluación tradicional
<i>Muestras</i>	Experimentos del alumnado, debates, portafolios, exposiciones, productos del alumnado.	Pruebas de elección múltiple, de emparejamiento, de verdadero-falso, de ordenamiento, de completar, etc.
<i>Juicio evaluativo</i>	Basado en la observación, en la subjetividad y en el juicio profesional.	Basado en la recogida de información objetiva e interpretación de las puntuaciones.
<i>Focalización</i>	De manera individualizada sobre el alumno a la luz de sus propios aprendizajes.	Fundamentalmente sobre la puntuación del alumno en comparación con la puntuación de los otros alumnos.
<i>Informe</i>	El evaluador crea una historia evaluativa respecto del individuo o del grupo.	El evaluador presenta el conocimiento del alumno mediante una puntuación.
<i>Generalización</i>	No (idiosincrásica).	Sí (generalizable).
<i>Acción curricular</i>	Es facilitada por medio de la información recogida.	Es inhibida debido a la información recogida.
<i>Agente</i>	Los estudiantes pueden participar en su propia evaluación.	El profesor u otra fuerza externa conduce la evaluación.

1.2. El proceso de evaluación del alumnado

Independientemente del enfoque utilizado, Mateo (2000a), basándose en Rodríguez Espinar (2000), señala las siguientes fases en el proceso evaluativo:

1. Establecimiento de los objetivos de evaluación.
2. Asignación de las tareas a realizar por el alumnado.
3. Fijación de sus criterios de realización.
4. Explicitación de sus estándares o niveles de logro.
5. Tomar muestras de las ejecuciones de los alumnos.
6. Valoración de dichas ejecuciones.
7. Retroalimentar adecuadamente al alumno.
8. Tomar decisiones.

1.3. La evaluación de las competencias

En la actualidad es necesario abordar la evaluación de competencias al hablar de la evaluación del alumnado. El término competencia está de moda, aunque su significado haya resultado muchas veces controvertido.

Para el desarrollo de una competencia, es necesaria la interacción entre el conocimiento y la realidad. El uso de conocimientos de manera competente exige (Mateo y Martínez, 2008):

- a) La comprensión profunda desde su propia lógica de los conocimientos.
- b) Leer e interpretar la realidad donde se deben aplicar.
- c) Mostrar una actitud decidida por incidir significativamente en la realidad.

Como consecuencia, se supone que nuestra propia percepción y comprensión de la realidad aumentará, así como nuestra capacidad de actuar en otros contextos. Es decir, poseer una competencia es algo más que su simple dominio o su aplicación mecánica sobre un contexto o realidad.

El uso de las competencias en las diferentes etapas educativas ha hecho que deban replantearse los currículos dado que sus contenidos deben relacionarse e interactuar necesariamente con contextos de realidad para poder así transformar el conocimiento en competencia.

Obviamente, la evaluación de las competencias supone decantarnos por modelos de evaluación alternativos al tradicional. Es decir, deberíamos situarnos en la evaluación de ejecución.

2. Evaluación de centros

La evaluación de centros es otro ámbito de evaluación importante en la actualidad. Hasta hace unas décadas no obstante, el centro no suscitaba el interés que causa hoy en día y ello ha sido debido a diversas razones. En primer lugar, debemos citar el famoso informe Coleman (1966) que demostró que los antecedentes socioculturales y económicos de los alumnos condicionaban el rendimiento posterior de éstos. Sin embargo, no es menos cierto que en investigaciones posteriores se descubrió que alumnos con antecedentes similares que asistían a centros educativos distintos obtenían un rendimiento diferente. Es decir, se descubrió otro factor diferente de los antecedentes que podía explicar parte de las diferencias entre los alumnos. A este nuevo factor se le denominó efecto escuela. Sin embargo, existen otra serie de razones que han contribuido también a que los administradores, educadores e investigadores centren su atención en la escuela (Pérez Juste, 1999; Tiana Ferrer, 1999; Lukas *et al.*, 2000b):

- La necesidad de información que muestra la sociedad tanto con respecto al sistema educativo en general como respecto a cada centro educativo en particular.
- La idea de que el centro educativo contribuirá a la mejora cualitativa de la educación.
- El aumento de la autonomía de los centros que lleva como contrapartida la rendición de cuentas.

Evaluar un centro educativo supone responder a tres cuestiones fundamentales. A saber: qué queremos evaluar, cómo lo vamos a hacer y para qué vamos a evaluar (De Miguel, 1997).

Al evaluar un centro educativo se ha de tener en cuenta que hoy en día éste constituye una organización compleja que puede ser analizada desde distintos enfoques. De Miguel (1997) y Mateo (2000a) señalan los siguientes:

- a) Desde una perspectiva descriptiva, un centro es considerado como una realidad dinámica en la que se aglutinan un conjunto de personas (profesores, alumnos, padres, etc.) cuya meta es alcanzar unos objetivos de aprendizaje. Desde este punto de vista de la eficacia se analizan el contexto, las variables de entrada, de proceso y de producto.
- b) Desde una perspectiva de comunidad educativa es considerado como espacio de relación interpersonal, ámbito donde conviven un conjunto de personas dentro de ciertas normas y en relación a ciertos fines. La evaluación se orienta a la comprensión en profundidad de los hechos.
- c) Desde una perspectiva del centro como organización en continuo proceso de desarrollo y que mantiene estrecha relación con el medio. La evaluación se orienta a buscar un mayor ajuste entre el centro y el contexto social mediante un proceso continuo de aprendizaje y de desarrollo interno.

Tal y como señala Mateo (2000a, p. 163) «el diseño específico de un modelo evaluativo por un centro estará condicionado: por el concepto que tengamos respecto a él, por el carácter que asignemos al término calidad educativa y por la manera en que ambos extremos estén interrelacionados».

Por ello, un segundo aspecto que debemos analizar es el relativo al concepto de calidad de los centros partiendo de los criterios más utilizados (De Miguel, 1997; Mateo, 2000a).

1. *La calidad como prestigio/excelencia*: Se entiende por calidad el prestigio o la reputación académica y social que tienen determinadas instituciones acreditadas, no mediante procedimientos objetivos, sino mediante el juicio de expertos en la materia.
2. *La calidad en función de los recursos*: Se utiliza como criterio de calidad la dotación del centro en cuanto a recursos humanos y materiales se refiere.
3. *La calidad en función de los resultados*: Los resultados de los alumnos egresados son el juicio utilizado como criterio de calidad del centro.
4. *La calidad como valor añadido*: El centro que contribuye de manera más decisiva en el cambio de conducta (impacto en los conocimientos de los alumnos, la evaluación de su personalidad y el desarrollo de su carrera) de los alumnos es el considerado de mayor calidad desde esta perspectiva. Se define la calidad en términos de valor añadido (*value added*).

5. *La calidad como adecuación a propósitos*: Desde este punto de vista se asume la calidad como algo práctico y no abstracto; es decir, se utilizan como criterios los fines educativos señalados en el marco legal, los objetivos del centro o la satisfacción de los alumnos.
6. *La calidad como perfeccionamiento*: Se apuesta por una cultura de calidad basada en una mejora continua que consiste en llevar a cabo tareas con la mejor perfección posible y no sólo como respuesta a requisitos legales o controles técnicos.
7. *La calidad total como meta*: Desde esta perspectiva se sustituye el control de calidad sobre el producto por la comprobación de que dicho producto ha sido elaborado a través de un proceso controlado.

2.1. Enfoques en la evaluación de centros educativos

En las últimas décadas se han desarrollado dos movimientos relacionados con el conocimiento y el cambio de los centros educativos: los que han investigado la eficacia de las escuelas y los proyectos de mejora y desarrollo de las escuelas (De Miguel, 1997). Los dos movimientos han ejercido, sin ningún género de dudas, una gran influencia en la conceptualización del funcionamiento de los centros y en el análisis de los factores relacionados con la calidad y la mejora de las escuelas (Lukas *et al.*, 2000b). Sin embargo, estos dos movimientos se han ido desarrollando de manera independiente sin tener apenas en cuenta los avances que se daban en el otro movimiento.

Los estudios sobre las escuelas eficaces se sitúan dentro del campo de la investigación científica y, por lo tanto, su meta principal es ampliar el conocimiento existente acerca del funcionamiento de las escuelas, analizando los factores que puedan resultar influyentes. Además, los resultados obtenidos en estos estudios no pueden llevarse directamente a la práctica. Son necesarios modelos que facilitan el cambio educativo para pasar de la teoría a la práctica. Diametralmente opuesta es la situación que se da en los programas de mejora escolar, puesto que su objetivo es más aplicado y su meta principal se sitúa en la verificación de la utilidad de ciertas estrategias de cambio. Normalmente, el conocimiento es obtenido de forma menos rigurosa; por lo tanto, sus conclusiones están limitadas y las posibilidades de generalización son escasas (Lukas *et al.*, 2000b). Parafraseando a Muñoz-Repiso y Murillo (2000) diríamos que el primero señala qué se debe cambiar, y el segundo, cómo realizar ese cambio.

Una segunda diferencia entre estos movimientos debemos buscarla en la elección de los indicadores de calidad. Los estudios de las escuelas eficaces se basan en los resultados de los alumnos. El propósito de estos estudios es analizar cuáles son los procesos de centro o de aula que están más relacionados con los resultados de los alumnos. En el otro extremo, los proyectos de mejora de la escuela analizan los procesos y las condiciones del cambio y tienen una perspectiva más amplia y más sistemática de la realidad educativa.

Por último, también existen diferencias en la metodología utilizada. Mientras que los investigadores de las escuelas eficaces prefieren utilizar metodología cuantitativa y correlacional, los investigadores orientados a la mejora de la escuela se decantan por la metodología cualitativa.

En los últimos años se han llevado a cabo intentos para acercar esos dos movimientos que se conocen con el nombre de «Mejora de la eficacia escolar» (*Effective School Improvement*) (Muñoz-Repiso y Murillo, 2000). En el cuadro 5.3 se resumen las aportaciones de cada uno de los dos movimientos al nuevo movimiento de «Mejora de la eficacia escolar».

Cuadro 5.3. Aportaciones del movimiento de las escuelas eficaces y del movimiento de mejora de las escuelas. (Adaptado de Stoll y Wikeley, 1998)

Aportaciones del movimiento de las escuelas eficaces	Aportaciones del movimiento de mejora de las escuelas
<ul style="list-style-type: none"> • La importancia dada a los resultados. • La importancia dada a la equidad. • La utilización de datos para tomar decisiones. • La comprensión de que la escuela es el centro del cambio. • La utilización de la metodología cuantitativa. 	<ul style="list-style-type: none"> • La importancia dada a los procesos. • La importancia dada al desarrollo y a la acción. • La mejora en los aspectos elegidos por el centro. • La comprensión de la importancia de la cultura escolar. • La importancia de centrarse en la enseñanza. • La visión de la escuela como agente de cambio. • La utilización de la metodología cualitativa.

De la misma manera, en el mundo anglosajón han surgido movimientos que plantean la evaluación de centros sobre la base de cuatro procedimientos agrupados en torno a la evaluación interna y externa de los centros (De Miguel, 1997; Mateo, 2000a):

- *Evaluación interna*: en la que la responsabilidad de la evaluación recae sobre el propio centro. Las estrategias son:
 - a) Procedimientos para estimar la calidad de los resultados (eficacia/eficiencia).
 - b) Sistemas establecidos para asegurar la calidad en el funcionamiento (mejora).

- *Evaluación externa*: en la que la responsabilidad de la evaluación recae sobre órganos externos al centro. Las estrategias son:
 - a) Control externo o auditoría sobre la adecuación del centro a la normativa vigente.
 - b) Valoración externa de la eficacia del centro en función de los resultados.

A partir de estos cuatro procedimientos se tienen en cuenta las necesidades tanto de las audiencias externas como internas, de la misma manera que se satisfacen las exigencias de los enfoques centrados en la eficacia y los centrados en la mejora. El procedimiento a utilizar en cada caso dependerá de qué se quiere evaluar.

2.2. Criterios que determinan los modelos evaluativos

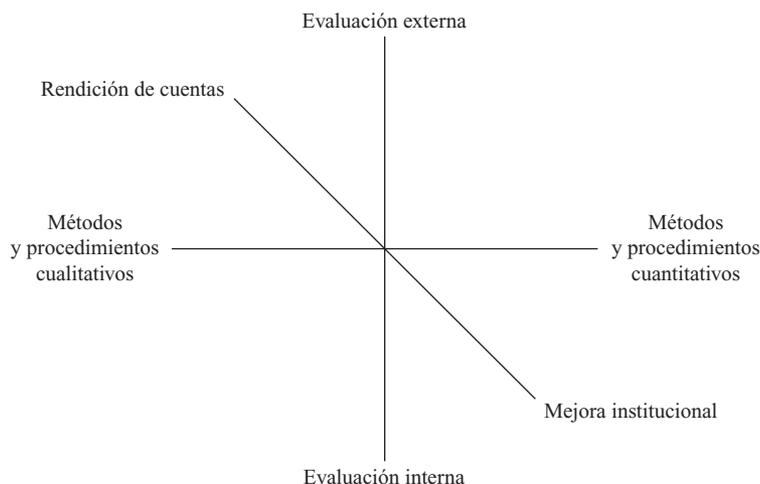
Tras definir lo que queremos evaluar deberemos responder a la pregunta de cómo evaluar. Para ello, tendremos que tener en cuenta una serie de criterios que condicionarán el modelo evaluativo a utilizar. Tiana Ferrer (1999) señala que para analizar los diversos modelos deben tenerse en cuenta tres elementos que los definen: los propósitos de la evaluación, los métodos y procedimientos utilizados y las perspectivas adoptadas.

Cada uno de estos elementos viene definido mediante un eje de dos polos que constituye un continuo en el que pueden adoptarse diferentes posiciones intermedias. Como el autor señala, estos tres elementos podrían representarse en un espacio tridimensional continuo, cuyos ejes perpendiculares entre sí sirven de referencia para definir cualquier posición espacial (Tiana Ferrer, 1999). En la figura 5.3 aparecen los tres ejes mencionados con sus polos de referencia (Lukas *et al.*, 2000b, p. 43).

El primero de los ejes se refiere a los fines de la evaluación. Como puede apreciarse en la figura, en uno de los polos se sitúa la rendición de cuentas (*accountability*) y en el polo opuesto la mejora institucional. Tal y como señala Tiana Ferrer (1999) y como se ha mencionado anteriormente, un modelo de evaluación no va a situarse necesariamente en uno de los dos polos, sino que caben alternativas intermedias que nos señalarán una tendencia más o menos acusada hacia cualquiera de los polos. Lógicamente, la rendición de cuentas tendría una función de tipo sumativo (certificación, acreditación) y la mejora institucional tendría una función de tipo formativo.

El segundo de los ejes representa la dimensión de los métodos y los procedimientos utilizados en la evaluación. Tal y como puede observarse, en uno de los polos tenemos los procedimientos cuantitativos y en el polo opuesto los procedimientos cualitativos. Cada uno de ellos tiene sus ventajas y limitaciones, así como sus defensores y detractores. Sin embargo, hace algunos años que diversos autores plantean la utilización de ambos tipos de procedi-

Figura 5.3. Criterios que determinan los modelos educativos



mientos en un mismo modelo de evaluación, intentando hacer resaltar el carácter complementario y no excluyente de ambos.

El tercero de los ejes es el relativo al agente evaluador. En uno de los polos se encuentra la evaluación interna; es decir, aquella que es llevada a cabo por los propios actores del centro. En el otro polo se sitúa la evaluación externa que se da cuando los evaluadores son ajenos al centro. Cada una de estas modalidades tiene sus ventajas y desventajas que podrían resumirse en el cuadro 5.4.

Hoy en día diversos autores plantean modelos de evaluación mixtos en los que se insiste en la participación externa en la evaluación interna. Nevo habla de *diálogo para la mejora*. Otros, como Santos Guerra (1993), proponen la *coevaluación*, la combinación de la evaluación externa e interna.

De Miguel (1997), por su parte, además de tomar como criterios a) la evaluación interna y externa y b) la orientación sumativa y formativa, añade otros tres:

- a) Evaluación regulada o autorregulada; es decir, quién es el promotor de la evaluación. Si está regulada desde el exterior, el objetivo que se persigue es el control, mientras que si la evaluación es autorregulada (promovida desde el interior), la finalidad será conocer la propia situación para tomar medidas que la optimicen.
- b) La evaluación global y parcial; es decir, puede pensarse en una evaluación holística del centro, o limitar la evaluación a algún aspecto del mismo (equipo directivo, clima, profesorado...).
- c) Indicadores de rendimiento y revisión de pares; es decir, la elección de la estrategia que se va a emplear en la evaluación. La utilización de indicadores de rendimiento supone traducir la información recogida en

Cuadro 5.4. Ventajas y limitaciones de la evaluación interna y externa.
(Tomado de Lukas *et al.*, 2000b, p. 44)

	Ventajas	Desventajas
<i>Evaluación interna</i>	<ul style="list-style-type: none"> • Legitimidad conferida al proceso. • Mejor comunicación. • Mayor capacidad de dinamizar la vida del centro. • Estrecha conexión con las decisiones de mejora que puedan adoptarse. • Mejor conocimiento del contexto. • Menos amenazador para los que van a ser evaluados. 	<ul style="list-style-type: none"> • Falta de tiempo. • La falta de distancia emocional y el compromiso de los agentes puede convertir el proceso evaluador en un factor de autojustificación. • Menor objetividad y credibilidad. • Ausencia de contraste externo.
<i>Evaluación externa</i>	<ul style="list-style-type: none"> • Mayor independencia. • Más tiempo dedicado a la tarea. • Mayor rigor debido a la especialización técnica. • Mayor credibilidad de sus conclusiones. • Imparcialidad del juicio. • Inclusión de una referencia externa en la valoración. 	<ul style="list-style-type: none"> • Dificultad de adentrarse en la vida del centro. • Mayor rechazo de la comunidad escolar. • Sesgo informativo.

términos de indicadores —cuantitativos o cualitativos— objetivables en relación con un objetivo que permita conocer el funcionamiento del centro y proponer las decisiones oportunas. En el caso de la revisión de pares por el contrario, se supone que son los colegas ajenos al propio centro los que están más capacitados para emitir juicios de valor ya que utilizan criterios más contextualizados.

2.3. Estrategias utilizadas para la evaluación de centros

Tal y como señala De Miguel (1997), los criterios que hemos apuntado definen el marco a partir del cual se constituyen las distintas estrategias para la evaluación de centros. Según él, cabe por lo tanto identificar en cada uno de los procedimientos cada uno de los criterios mencionados (regulado o autorregulado, orientación formativa o sumativa...). No obstante, en muchos casos pueden aparecer dos dimensiones y además tal y como hemos mencionado anteriormente estas dicotomías podrían entenderse como un eje en el que ca-

brían diferentes posiciones. Vamos a señalar los procedimientos de evaluación más utilizados indicando las características más importantes de cada uno en el cuadro 5.5 (De Miguel, 1997):

1. *La revisión y permiso de los estados.* Se centra en la verificación de las condiciones establecidas legalmente para el funcionamiento de un centro educativo. El objetivo, por lo tanto, es determinar si cumple la normativa; por ello, normalmente se realiza a priori, es decir, antes de poner en funcionamiento un centro educativo.
2. *La acreditación.* Este proceso supone que un centro por voluntad propia ofrece información sobre su funcionamiento y sus resultados a una agencia externa independiente que, a la vista de esa información y de su análisis, emitirá un juicio de valor a partir de unos estándares conocidos sobre el valor y calidad de la institución. Si el juicio es positivo, la agencia externa otorgará la acreditación al centro en cuestión.
3. *Las auditorías.* Éstas tienen como finalidad efectuar un chequeo de los centros para comprobar si los mecanismos establecidos internamente funcionan para asegurar la calidad de los procesos y los resultados y no sólo su rentabilidad o eficacia. Es decir, sería como una estrategia de acreditación, pero en este caso el promotor es un organismo externo (administración, consejo escolar...).
4. *Los autoestudios.* La finalidad de éstos es mejorar la calidad del funcionamiento del centro mediante un proceso de autorreflexión colectivo. Es, pues, un proceso autorregulado e interno ya que es el propio colectivo el que fija los objetivos, elige la metodología y toma las decisiones. Estaría caracterizado por la participación, la responsabilidad y el compromiso de los miembros del centro (Mateo, 2000a). Metodológicamente se concibe como un proceso de revisión interna con el propósito de detectar los puntos fuertes y débiles y efectuar propuestas de mejora.

Cuadro 5.5. Características de las estrategias utilizadas para la evaluación de centros

	Revisión y permiso de los estados	Acreditación	Auditoría	Autoestudio	Evaluación institucional
<i>Regulación</i>	Regulada	Autorregulada	Regulada	Autorregulada	Autorregulada
<i>Alcance</i>	Global	Parcial	Parcial	Global	Global
<i>Agente</i>	Externo	Externo	Externo	Interno	Interno y externo
<i>Orientación</i>	Sumativa	Sumativa	Sumativa	Formativa	Formativa

5. *La evaluación institucional.* Ésta intenta conjugar la doble perspectiva interna y externa de la evaluación de centros. La evaluación interna y autorregulada es necesaria para «*la generación de compromiso entre el centro y su propia búsqueda de calidad y es un elemento clave de desarrollo institucional*» (Mateo, 2000a, p. 176). Por otra parte, se considera que la evaluación interna adquiere credibilidad en la medida en que es contrastada externamente. Es decir, se integran la evaluación interna (autoestudio) y la externa con propósitos claramente formativos. Además del análisis de la eficacia y de la eficiencia del centro, la evaluación institucional se centra en la revisión de los procesos.

2.4. El proceso de la evaluación de centros

En términos generales, Mateo (2000a) señala estas siete etapas en el proceso de evaluación de centros:

1. *Análisis de necesidades*
Consiste en realizar un diagnóstico acerca de las necesidades y problemas del centro para poder elegir los contenidos objeto de evaluación.
2. *Elección y concreción de los objetivos de evaluación*
Debe tenerse en cuenta que el tema elegido sea relevante y significativo, por una parte, y esté delimitado y sea evaluable, por otra.
3. *Procedimientos, recursos y normas*
Tras delimitar el objeto evaluativo se deben elegir los procedimientos para llevarlo a cabo. Habrá que seleccionar los instrumentos para recoger la información y diseñar el plan de recogida de información previendo los recursos que se precisarán.
4. *El análisis y la interpretación de la información*
Tras el vaciado de la información recogida y su posterior análisis vendrá la interpretación contrastándola con los criterios de valor previamente formulados.
5. *El informe evaluativo y la toma de decisiones*
6. *Uso intensivo y extensivo de la información*
El centro debe reflexionar, debatir y definir un plan estratégico de desarrollo y mejora para hacer discurrir la información evaluativa generada y de esta manera diseñar las propuestas de mejora en un marco de desarrollo para el centro.
7. *La metaevaluación*

3. Evaluación del profesorado

3.1. Concepto y problemática

La evaluación del profesorado es otro de los ámbitos de la evaluación educativa importante en la actualidad dado que es innegable el papel que juega en la mejora de la escuela. No es posible la mejora de la escuela sin la participación directa y activa de los profesores. Tal y como señalan Abalde, Muñoz y Ríos (2002) si bien son muchas las correlaciones existentes entre los elementos que inciden en el sistema educativo, señalan al profesorado como especialmente sensible con respecto a su calidad.

Marcelo y Villar (1994, p. 314) definen la evaluación del profesor como «el proceso de concebir, obtener y comunicar información para adoptar decisiones educativas sobre el profesor. En la evaluación del profesor se formulan preguntas (derivadas de los objetivos planteados para la función o rol docente), se recopila información a través de un enfoque modélico seleccionado (por medio de instrumentos válidos y fiables) para establecer conclusiones (basadas en criterios de valor o de mérito) que permiten proponer recomendaciones (para tomar decisiones sobre la mejora o la selección), durante el proceso evaluador (evaluación formativa) o una vez realizada la valoración (evaluación sumativa)».

El Joint Committee on Standards for Educational Evaluation, por su parte, lo define como «la evaluación sistemática de la actuación profesional y/o cualificaciones de una persona en relación a un cometido profesional y algún fin institucional y defendible» (2001, p. 26).

Casi nadie pone en duda la necesidad de la evaluación del profesorado, dada su relación con la mejora de la calidad de la enseñanza, pero, sin embargo, es un tema espinoso por los propósitos de la evaluación y por los procedimientos utilizados. Como señalan Villa y Morales (1993), la evaluación del profesorado es algo **complejo** porque puede centrarse en aspectos muy diversos (su competencia, formación, resultados que obtiene, etc.), es algo **controvertido** porque muchos profesores no quieren ser evaluados y es algo que supone implícitamente un modelo de enseñanza.

Siguiendo esta línea, Jiménez Jiménez (1999c, pp.183-184) señala los problemas que habrá que superar en cualquier proceso de evaluación del profesorado:

- El tiempo, dado que los momentos de aplicación no siempre satisfacen a todos.
- La idoneidad de los instrumentos a aplicar (observación, diarios, etc.).
- Las personas y las instituciones implicadas, puesto que en un principio pueden ser reacios a la evaluación.
- Las exigencias que desde la administración, las personas, el centro y el propio equipo y objetivos de la propia evaluación nos requieren, que no siempre se cumplen.

- La coordinación entre las actividades de la evaluación y el desarrollo cotidiano de la práctica docente.
- Los problemas que pueden darse entre los profesores debido a que la propia evaluación (por no haber explicitado claramente su finalidad o no haber sido bien entendida) puede enturbiar las relaciones entre los profesores.

Para superar estos problemas, es necesario que previamente a la evaluación haya un proceso de negociación con los directamente implicados independientemente de si la cuestión ha sido solicitada de forma externa o interna. En este sentido, Santos Guerra (1990, p. 63) señala los acuerdos a los que habría que llegar:

- La naturaleza del trabajo que se va a realizar.
- Los fines que se persiguen.
- Los métodos que se van a utilizar.
- El tipo de colaboración que se requiere.
- La confidencialidad de los datos.
- El calendario de trabajo.
- El momento y la forma de entregar los informes.
- El contenido de los informes.
- La utilización de informes por otras personas ajenas al centro.
- El equipo que va a realizar el trabajo.

Continuando por los mismos derroteros, Stronge (1997) señala que para obtener una evaluación de alta calidad en la que se conjuguen la mejora de la escuela y la mejora del profesorado deben tenerse en cuenta las siguientes acciones:

1. Establecer objetivos recíprocamente beneficiosos. Tanto el centro como los profesores deben valorar y percibir los objetivos establecidos como importantes.
2. Establecer un proceso sistemático de comunicación.
3. Cuidar la aplicación técnica del sistema de evaluación.
4. Uso de fuentes de datos múltiples. El uso de fuentes de información diversas es uno de los aspectos en los que coinciden la mayoría de los evaluadores por las múltiples ventajas que ofrece frente a la utilización de una sola fuente:
 - El marco de actuación evaluado es más rico y contextualizado.
 - La información es recogida en escenarios y situaciones más naturales.
 - Se integran fuentes de datos primarias y secundarias.
 - Mayor fiabilidad.

- La documentación de la actuación docente es más cercana a la realidad.
- Bases legalmente más defendibles respecto a las decisiones evaluativas.

Además de las fuentes de datos múltiples, deberían utilizarse también evaluadores múltiples que aporten perspectivas diferentes en la emisión de juicios de valor a partir de la información disponible.

V. Creación de un clima adecuado para la evaluación.

3.2. Propósitos de la evaluación del profesorado

Tradicionalmente se han distinguido dos finalidades en la evaluación del profesorado: la petición de responsabilidades y la mejora de la actividad docente. En esta línea, Villa y Morales (1993) señalan que los propósitos de la evaluación del profesor se pueden clasificar en cuatro metas básicas en función del carácter formativo o sumativo y en función de si evaluamos a un solo profesor o a un grupo. En el cuadro 5.6 se resumen estas cuatro finalidades.

Cuadro 5.6. Propósitos básicos de la evaluación del profesor.
(Tomado de Villa y Morales 1993, p. 32)

		NIVEL	
		Individual	Organizativo
Propósito	<i>Mejora docente (evaluación formativa)</i>	Desarrollo individual (perfeccionamiento)	Mejora institucional
	<i>Rendición de cuentas (evaluación sumativa)</i>	Decisiones sobre un individuo (estatus laboral)	Decisiones sobre el estatus escolar o institucional (certificación)

Sin embargo, reconocemos junto con Mateo (2000a) que esta dicotomía no es suficiente, deben aunarse dinámicamente los intereses y las necesidades del profesor con los de la institución para que se desarrolle simultáneamente la mejora de ambos. Este modelo estaría enraizado en dos amplios propósitos:

- a) *Orientado a los resultados* que contribuyan al logro de los objetivos personales del profesor y del centro. También debería ofrecer información para la petición de responsabilidades por la actuación docente en función de los objetivos planteados.
- b) *Orientado a la mejora* que contribuya al desarrollo personal y profesional del profesor, así como la mejora del centro.

3.3. El proceso de evaluación del profesorado

Aunque autores como Jiménez Jiménez (1999c) han propuesto otras fases en la evaluación del profesorado, consideramos que la propuesta presentada por Mateo (2000a) es la más completa y adecuada, y es por ello por lo que vamos a presentarla aquí. Este autor distingue cuatro grandes fases (ideación, desarrollo, implementación y metaevaluación) en cada una de las cuales se diferencian otras subfases (véase cuadro 5.7).

Cuadro 5.7. Fases en el diseño de la evaluación del profesorado

FASE 1: IDEACIÓN

- 1.1. Identificación del sistema de necesidades.
 - 1.2. Derivación de los objetivos institucionales.
 - 1.3. Definición del modelo de calidad para la docencia.
-

FASE 2: DESARROLLO

- 2.1. Determinación de las funciones y responsabilidades del profesorado.
 - 2.2. Desarrollo de los criterios de evaluación y de sus indicadores.
 - 2.3. Fijación de estándares para la valoración de la actividad docente del profesor.
-

FASE 3: IMPLEMENTACIÓN

- 3.1. Documentación de la actividad profesional del profesor.
 - 3.2. Emisión de los juicios de valor respecto a la docencia.
 - 3.3. Uso intensivo y extensivo de la información evaluativa: la derivación de las propuestas de mejora.
-

FASE 4: META-EVALUACIÓN

FASE 1: IDEACIÓN

La fase de ideación se compone de:

1. *La identificación del sistema de necesidades*; es decir, cada centro debe realizar un examen minucioso de sus necesidades que ayude a clarificar los objetivos institucionales y a definir el proyecto educativo.
2. *La derivación de los objetivos institucionales*; es decir, se deberán explicitar las metas que se pretenden alcanzar.
3. *La definición de los modelos de calidad para la docencia*; es decir, será la que sustantive y oriente el proceso de evaluación del profesorado, siendo el vínculo entre ésta y la mejora de la calidad del centro.

FASE 2: DESARROLLO

En la fase de desarrollo se distinguen tres subfases:

1. *Determinación de las funciones y responsabilidades del profesorado*

Sólo partiendo de modelos de docencia bien establecidos y definidos se podrán establecer descripciones apropiadas y ajustadas de las funciones y responsabilidades del profesorado. Por lo tanto, deberán establecerse las áreas de responsabilidad del profesor en el contexto institucional y sus funciones más importantes. En el cuadro 5.8 tomado de Scriven (1994), citado por Mateo (2000a), el autor recoge el listado de deberes del profesor, aunque no agotan la concepción y la práctica de la evaluación del profesorado.

2. *Desarrollo de los criterios de evaluación y de sus indicadores*

Los criterios proveen una dimensión general de lo que será posible valorar en el nivel de realización alcanzado por el profesor, mientras que los indicadores permitirán cuantificar las variables criterio facilitando de esa manera la objetivación de la evaluación.

3. *Fijación de estándares para la valoración de la actividad docente del profesor*

Establecer los estándares o los referentes exige determinar los niveles de actuación del profesor considerados como aceptables. Obviamente, es necesario que de antemano tengamos bien fijados los objetivos institucionales, su naturaleza evaluativa y el marco general de la docencia.

FASE 3: IMPLEMENTACIÓN

En la fase de implementación se distinguen tres etapas:

1. *Documentación de la actividad profesional del profesor*

Al hablar de la documentación, Mateo entiende el proceso de recogida de información acerca de la actividad del profesor, teniendo siempre en cuenta que ésta debe hacerse a partir de múltiples fuentes de información. El propio autor, citando a Stronge (1997), señala siete procedimientos para obtener información en el proceso de evaluación del profesorado:

1. La observación en el aula.
2. La opinión de los padres.
3. La opinión de los alumnos.
4. La evaluación a partir del juicio emitido por los padres.
5. El rendimiento en el aprendizaje de los alumnos.
6. El autoinforme.
7. El portafolios.

Cuadro 5.8. Contenidos de la función docente (Scriven, 1994; tomado de Mateo, 2000a, p. 102)

Conocimiento de la materia	<ul style="list-style-type: none"> — En las áreas de su competencia. — En las materias transversales del currículum.
Competencias instruccionales	<ul style="list-style-type: none"> — Destrezas comunicativas. — Destrezas de gestión: <ul style="list-style-type: none"> • Gestión del proceso. • Gestión del progreso. • Gestión de las emergencias. — Destrezas de programación y desarrollo: <ul style="list-style-type: none"> • Planificación del curso. • Selección y creación de material. • Utilización de los recursos disponibles. • Evaluación del curso, de la enseñanza, de los materiales y del currículum.
Competencias evaluativas	<ul style="list-style-type: none"> — Conocimientos sobre la evaluación de los estudiantes. — Destrezas de construcción y administración de pruebas. — Clasificación, puntuación y calificación: <ul style="list-style-type: none"> • Proceso. • Producto. — Registro e informe del rendimiento de los estudiantes. — Conocimiento sobre informes de rendimientos.
Profesionalidad	<ul style="list-style-type: none"> — Ética. — De actitud. — De mejora. — De servicio: <ul style="list-style-type: none"> • Conocimiento acerca de la profesión. • Ayuda a los docentes noveles y colegas. • Trabajo para las organizaciones profesionales. • Investigación sobre la enseñanza. • Conocimiento de los deberes. • Conocimiento de la escuela y sus deberes.

2. *Emisión de los juicios de valor respecto de la docencia*

3. *Uso intensivo y extensivo de la información evaluativa: la derivación de las propuestas de mejora*

Tal y como señala Mateo (2000a, p. 114) «la derivación de las propuestas de mejora deberá asentarse sobre un análisis intensivo de la información y en su uso extensivo». Para ello, el autor propone un análisis en profundidad, con relevancia e intensidad, orientado por el modelo de calidad de la docencia, por un lado, y para incidir extensivamente, por otro; debe existir un plan estratégico de centro en donde integrar la información procedente de la evaluación para proyectar cualquier propuesta de mejora.

FASE 4: LA META-EVALUACIÓN

Como ya se ha mencionado en el apartado 3.3 relativo a las fases del proceso evaluativo en general y de la metaevaluación en particular, con la creación a partir de 1975 del Joint Committee on Standards for Educational Evaluation se han ido proporcionando diferentes recomendaciones para realizar metaevaluaciones. Las recomendaciones más conocidas son las relativas a la evaluación de programas, proyectos y materiales, cuya última revisión data de 1994 (de 1998 la traducción española). De la misma manera, el mismo comité publicó en 1994 (la traducción española data de 2001) los estándares¹ de evaluación del personal. Éstos reflejan las características que debiera tener una evaluación del profesorado considerada como aceptable. Los estándares están clasificados en cuatro categorías:

1. *Estándares de propiedad*

Estos estándares requieren que las evaluaciones del profesorado se lleven a cabo legal, éticamente y con el debido respeto hacia el bienestar de los evaluados y los clientes de las evaluaciones (véase cuadro 5.9).

Cuadro 5.9. Estándares de propiedad

Orientación hacia el servicio	Las evaluaciones de los profesores deberían promover principios pedagógicos bien fundamentados, el cumplimiento de los fines institucionales y el desempeño eficaz de las responsabilidades profesionales con el fin de que se cubran las necesidades educativas de los alumnos, la comunidad y la sociedad.
Orientaciones formales de evaluación	Para que las evaluaciones sean consistentes, equitativas y conformes con la legislación correspondiente y con los códigos éticos, las orientaciones para la evaluación de los profesores deberían registrarse en declaraciones políticas o en acuerdos negociados.
Conflicto de intereses	Los conflictos de intereses deberán detectarse y abordarse abierta y honestamente para que no pongan en peligro el proceso y los resultados de la evaluación.
Acceso a los informes de evaluación del personal	El acceso a los informes de evaluación de los profesores debería limitarse a aquellas personas con un interés legítimo de revisar y actualizar los informes, garantizando de esta manera un uso apropiado de la información.
Interacciones con los evaluados	La evaluación debería tratar a los evaluados de una forma profesional de manera que su autoestima, motivación, actitud profesional y actitud hacia la evaluación mejoren o por lo menos no se vean perjudicadas.

2. Estándares de utilidad

Estos estándares están pensados para orientar las evaluaciones para que sean informativas, oportunas en el tiempo e influyentes (véase cuadro 5.10).

Cuadro 5.10. Estándares de utilidad

Orientación constructiva	Las evaluaciones deberían ser constructivas para que ayuden a los centros a desarrollar los recursos humanos y animar y ayudar a los evaluados a prestar un servicio excelente.
Usos definidos	Para que la evaluación pueda abordar las cuestiones adecuadas, los usuarios y todos los usos pretendidos de una evaluación de los profesores deberían ser identificados.
Credibilidad del evaluador	La evaluación debería ser dirigida y ejecutada por personas con las cualificaciones, destrezas y autoridad necesarias, y los evaluadores, comportarse profesionalmente para que los informes de evaluación sean respetados y utilizados.
Informes funcionales	Los informes deberían ser claros, oportunos, precisos y pertinentes para que sean de valor práctico para el evaluado y otros destinatarios pertinentes.
Seguimiento e impacto	Debería hacerse un seguimiento de la evaluación para que se ayude a los usuarios y evaluados a comprender los resultados y a adoptar las acciones adecuadas.

3. Estándares de viabilidad

Estos estándares requieren unos sistemas de evaluación tan fáciles de aplicar como sea posible, eficientes en su utilización del tiempo y recursos, financiados adecuadamente y viables desde diversos puntos de vista (véase cuadro 5.11).

Cuadro 5.11. Estándares de viabilidad

Procedimientos prácticos	Los procedimientos deberían ser planificados y llevados a cabo de forma que produzcan la información necesaria al tiempo que reducen el costo y las molestias.
Viabilidad política	La evaluación debería desarrollarse y supervisarse en colaboración para que todas las partes implicadas se involucren de forma constructiva en hacer que funcione el sistema.
Viabilidad fiscal	Se deberían asignar tiempo y recursos adecuados para las actividades de evaluación de manera que ésta pueda ser aplicada eficaz y eficientemente.

4. Estándares de precisión

Estos estándares exigen que la información recogida sea técnicamente precisa y que las conclusiones se enlacen lógicamente con esa información (véase cuadro 5.12).

Cuadro 5.12. Estándares de precisión

Definición de funciones	La función, responsabilidades, objetivos de actuación y las cualificaciones necesarias del evaluado deberían ser definidas claramente para que el evaluador pueda determinar criterios de evaluación válidos.
Entorno de trabajo	Para que las influencias y las limitaciones del entorno sobre la actuación docente sean recogidas en la evaluación, debería identificarse, describirse y registrarse el contexto en el que trabaja el evaluado.
Documentación de procedimientos	Para que los evaluados y otros usuarios puedan evaluar los procedimientos reales en relación con los pretendidos, los procedimientos de evaluación seguidos en realidad deberían ser documentados.
Medición válida	Los procedimientos de medición deberían ser elegidos o desarrollados y aplicados sobre la base de las funciones descritas y el uso pretendido para que las inferencias que tienen relación con el evaluado sean válidas y precisas.
Medición fiable	Los procedimientos de medición deberían elegirse o desarrollarse para asegurar la fiabilidad con el propósito de que la información recogida proporcione indicaciones consistentes y precisas de la actuación del evaluado.
Control sistemático de los datos	La información utilizada en la evaluación debería guardarse en lugar seguro.
Control del sesgo	Para que las cualificaciones del evaluado sean justamente valoradas, el proceso de evaluación debería garantizar la inexistencia de sesgos.
Seguimiento de los sistemas de evaluación	Los sistemas de evaluación de los profesores deberían revisarse periódica y sistemáticamente de forma que pudieran hacerse las modificaciones oportunas.

4. Evaluación del sistema educativo

4.1. Concepto y antecedentes

El último ámbito de la evaluación educativa que vamos a considerar se refiere a la evaluación del sistema educativo. Cuando hacemos referencia al sistema educativo, estamos planteando el «conjunto de instituciones, programas y acciones que una sociedad política destina intencional y sistemáticamente a la educación y/o a la instrucción de sus miembros; especialmente, a los

más jóvenes» (García Garrido, 1996, p. 554). Como señala García García (2000), la evaluación del sistema educativo es un caso particular de la evaluación de programas cuyas características serían las siguientes:

- Es un instrumento de retroalimentación y mejora.
- Está vinculado a la política educativa.
- El objeto de evaluación es muy amplio (el sistema en su conjunto).
- La audiencia es muy numerosa (toda la sociedad).

García García (2000) recoge diversas denominaciones y acepciones que se le han dado a la evaluación del sistema educativo. De esta manera, D'Hainaut (1988) habla de *regulación pedagógica* para referirse al proceso para la mejora del funcionamiento y resultados del sistema educativo. Por su parte, Landsheere (1996) distingue entre *control del sistema educativo* (observación sistemática y regular de la evolución de un sistema educativo a fin de mantenerlo en la línea de los objetivos fijados y de descubrir las eventuales disfunciones) y *pilotaje* (toma de decisiones a nivel macroscópico sobre la base de hechos, tan objetivos como sea posible, relativos al estado, al funcionamiento o a los productos del sistema). La evaluación para él sería, por lo tanto, una fase intermedia en el pilotaje. Sin embargo, la mayoría de los autores aceptan la definición de evaluación del sistema educativo señalada por García García (2000, p. 234) en la cual se tiene en cuenta tanto «el proceso sistemático de recogida y análisis de información periódica como la emisión de juicios de valor sobre la bondad del sistema educativo evaluado y los logros conseguidos».

En el ya famoso informe «La educación encierra un tesoro» realizado por la Comisión Internacional sobre la Educación para el Siglo XXI y dirigido por Delors (1996) se señala que la educación a lo largo de la vida se basa en cuatro pilares: aprender a conocer, aprender a hacer, aprender a convivir y aprender a ser. Estos pilares implican repensar y modificar el conjunto de conocimientos habituales y clásicos y sustituirlos o complementarlos con aquellas destrezas y actitudes que ayuden al desarrollo de la capacidad autónoma de aprendizaje. En este sentido, Mateo (2000a) señala que se deben establecer mecanismos de análisis de necesidades y determinar cuáles deben ser las habilidades y actitudes más importantes a educar o desarrollar en los alumnos. De la misma manera, en el citado informe, se señala la necesidad de establecer dispositivos objetivos y públicos para que la sociedad en general esté en condiciones de comprender la situación del sistema educativo y su impacto sobre la sociedad (Mateo, 2000a). Es decir, se está apuntando a la necesidad de evaluar los sistemas educativos. Es lógico pensar que todo sistema político contemple al lado de la planificación y gestión del sistema educativo, los sistemas de control sobre el funcionamiento o el producto del sistema educativo (García García, 2000).

Por lo tanto, hoy en día, el diseño de la evaluación del sistema educativo debe contemplar una perspectiva global ya que los objetivos a evaluar no pue-

den reducirse al rendimiento de los estudiantes, sino que han de tener en cuenta otros ámbitos, como los programas, el currículum, el profesorado, la administración, los procesos de cambio, etc., y una perspectiva integral, puesto que cada uno de los ámbitos mencionados presenta elementos complementarios para la comprensión y evaluación de los otros (Mateo, 2000a).

No obstante, para llegar a la situación actual, en las décadas anteriores, se han dado una serie de pasos en la evaluación de los sistemas educativos. Koretz (1992) señala tres momentos diferentes en el desarrollo de la evaluación de los sistemas educativos (García García, 2000):

1. *Primer periodo* (década de 1970), relacionado con el movimiento de evaluación de competencias mínimas. Se medían fundamentalmente los rendimientos de los estudiantes para acreditar su paso de un nivel a otro, aunque también tenía un componente de mejora. A nivel de España podemos citar el «Libro Blanco» de 1969 cuyo título era *La educación en España: Bases de una política educativa* que precedió a la reforma educativa de 1970 (Ley General de Educación).
2. *Segundo periodo* (años ochenta), relacionado con el movimiento de las reformas educativas. Durante esta época, la detección de anomalías en los sistemas educativos condujo a muchos países a reformar sus sistemas educativos, centrándose la evaluación en la rendición de cuentas. Se desarrollan tests estandarizados ampliándose el número de destrezas y conocimientos evaluados. De esta época podemos citar el informe *Examen de la política educativa española* realizado por la OCDE en 1986 y el *Libro Blanco para la Reforma del Sistema Educativo* publicado en 1989 por el Ministerio de Educación y Ciencia.
3. *Tercer periodo* (años noventa), relacionado con el movimiento de la calidad educativa y la descentralización de la educación pública. En esta época la mayoría de los Estados ha creado algún instituto de evaluación para evaluar los sistemas educativos. En el caso de España, el INCE en 1994 y más recientemente el ISEI/IVEI en 2001 en la Comunidad Autónoma Vasca. Estos institutos tienen entre sus funciones la de comprobar la calidad en cuanto a gestión, funcionamiento, procesos y resultados educativos de los sistemas educativos. De la misma manera, se han creado organismos interestatales que se están ocupando de la evaluación de los sistemas educativos para lograr estándares comunes de calidad que faciliten la comparación de los sistemas de diferentes países. Entre estos podemos citar el proyecto INES (Proyecto Internacional de Indicadores de la Educación) coordinado por el CERI (Centro para la Investigación e Innovación Educativa) y el proyecto PISA (Proyecto Internacional para la Producción de Indicadores de Rendimiento de los Alumnos) llevado a cabo por la OCDE. Además, se ha ampliado el objeto evaluativo, dado que también se han empezado a tomar en cuenta los profesores, los centros, la administración, etc.

La llegada a este tercer periodo en la evaluación de los sistemas educativos y en su generalización no se ha dado de forma casual, sino que hay que atribuirlo a varios factores. Podemos destacar entre otros (García García 2000, p. 229):

- La democratización de los sistemas políticos que lleva aparejada la necesidad de transparencia y de información ante y para la sociedad acerca de las acciones emprendidas y de los logros alcanzados.
- La competitividad económica entre países desarrollados que ha llevado a la comparación sistemática y periódica de la posición relativa que ocupa cada país en relación con el resto de los países. Estas comparaciones se fundamentan en el hecho de que sistemas educativos mejores lograrán ciudadanos mejor formados que a la larga contribuirán a posicionar al país entre las potencias mundiales.
- La descentralización de la educación o, en palabras de Hutmacher (1999), el principio de subsidiariedad que lleva asociada la necesidad de tomar decisiones allí donde se desarrolla la acción educativa. Esto ha conducido, por ejemplo, en España a la transferencia de competencias educativas a las comunidades autónomas con la necesidad posterior de asegurar la consecución de unos mínimos comunes consensuados por todos.
- La masificación del sistema educativo ha llevado a la creación de organismos intermedios poniéndose en evidencia la heterogeneidad de objetivos, calificaciones otorgadas, etc., lo que ha supuesto buscar nuevos instrumentos que permitan la comparación de los logros del sistema.

Ante esta situación, mediante la Ley Orgánica de Ordenación General del Sistema Educativo (LOGSE) promulgada en 1990 se incluye en España la evaluación del sistema educativo. En su preámbulo se señala que la evaluación es fundamental para analizar en qué medida los distintos elementos del sistema están contribuyendo al logro de los objetivos preestablecidos y, como consecuencia, se señala que la evaluación es uno de los factores que va a contribuir a la mejora de la calidad de la enseñanza. La LOPEG, promulgada en 1995, completó algunos aspectos de la acción evaluativa, pero en ambos casos se alejan de la concepción evaluativa que se ciñe exclusivamente a la medición del rendimiento de los alumnos, puesto que en sus artículos 62 (LOGSE) y 27 (LOPEG) se dice explícitamente que la evaluación del sistema educativo se aplicará sobre los alumnos, el profesorado, las instituciones escolares, los procesos educativos y sobre la propia administración educativa. En estas leyes se otorgan tres finalidades a la evaluación del sistema educativo:

1. La mejora de la calidad de la educación.
2. La utilidad para la toma de decisiones.
3. La contribución al debate social sobre educación.

También en la LOCE (2002) se señala explícitamente en el artículo 95 del Título VI («De la evaluación del sistema educativo») que la evaluación se extenderá a todos los ámbitos educativos, entre los que se señalan los procesos de aprendizaje de los alumnos, los procesos educativos, la actividad del profesorado, los centros docentes, la inspección de educación y la propia administración educativa.

4.2. Funciones de la evaluación del sistema educativo

Suelen distinguirse diversas funciones de la evaluación del sistema educativo, aunque todas ellas están relacionadas entre sí y normalmente se tiende a utilizarlas conjuntamente (García García, 2000). No obstante, el diseño de una evaluación estará íntimamente relacionado con la importancia relativa dada a las distintas funciones de la evaluación del sistema educativo. De todas maneras, hoy en día nadie niega la función prevalente que cumple la evaluación del sistema educativo como instrumento en la toma de decisiones políticas de un país. García García (2000, pp. 234-235) distingue cinco funciones, pero nosotros hemos descompuesto una de sus funciones en dos diferentes, por lo que señalamos seis funciones de la evaluación del sistema educativo:

- a) *Función diagnóstica*: La evaluación trata de obtener una fotografía parcial del sistema educativo para conocerlo y observar su evolución. Es decir, se trata de lograr información relevante en un momento determinado.
- b) *Función de toma de decisiones*: La evaluación ofrece información a los políticos educativos para que éstos tomen las decisiones pertinentes sobre las acciones a emprender.
- c) *Función de rendición de cuentas*: La evaluación ofrece información sobre el sistema educativo como convicción de que éste debe responder a las demandas de la sociedad. Esta rendición de cuentas ha de contemplar tres ámbitos: los usuarios (la sociedad en general y la comunidad educativa en particular), el colectivo profesional docente y la propia administración (Tiana Ferrer, 1997).
- d) *Función de comparación*: La evaluación, utilizando sistemas de indicadores comunes, ofrece información comparativa acerca de subsistemas educativos de un Estado (en España, por ejemplo, entre las distintas comunidades autónomas) o entre diferentes Estados.
- e) *Función de control o acreditación*: La evaluación ofrece información sobre los estándares de calidad de los principios, gestión y componentes del sistema educativo.
- f) *Función prospectiva*: La evaluación ofrece información para predecir posibles cambios y anticipar necesidades futuras basándose en el comportamiento y resultados del sistema educativo actual.

4.3. Evaluación del sistema educativo y calidad educativa

Durante el siglo xx en gran parte de Europa, Estados Unidos y el resto de los países desarrollados se consiguió que la educación se extendiera a toda la población; es decir, se logró el principio de universalización de la educación. En el desarrollo de esa consecución primaba la cantidad; esto es, cuántos ciudadanos tienen acceso a la educación. Sin embargo, tras el logro mencionado, el acento ha pasado de la cantidad a la calidad. En este sentido, Pérez Juste (1998) señala tres perspectivas al hablar de calidad educativa:

1. *La perspectiva absoluta que relaciona la calidad con excelencia*, de tal manera que un sistema de calidad será aquél capaz de lograr la formación integral del alumno con grandes conocimientos, mente crítica, capacidad para comprender la realidad social, con conciencia de sus derechos y deberes, con actitud de compromiso, etc.
2. *La perspectiva relativa que relaciona la calidad con superioridad*, de tal manera que un sistema será de más calidad tras compararlo con el sistema anterior o con otros sistemas diferentes.
3. *La perspectiva integral que relaciona la calidad con fines y logros y con procesos educativos*, de tal manera que hablaremos de «diseñar un sistema en el que la calidad de las metas se articule con unos procesos eficientes que den lugar a resultados elevados, de forma que pueda hablarse de niveles adecuados de eficacia» (García García, 2000, p. 240).

Como señala esta misma autora (pp. 240-241), la calidad del sistema educativo puede definirse y consecuentemente controlarse desde cuatro dimensiones interdependientes:

- *Calidad como relevancia* que asegura que lo que aprenden los alumnos responde a las necesidades sociales e individuales, desarrollándose a la vez como ciudadanos y como personas integrales física, intelectual, afectiva y socialmente.
- *Calidad como equidad* que asegura que todos los alumnos, independientemente de su origen y clase social tienen igualdad de oportunidades, procesos y resultados.

Se asume que las dos dimensiones mencionadas, relevancia y equidad, deben estar presentes independientemente del planteamiento a seguir. Sin embargo, el modelo de evaluación variará en función de la importancia relativa dada a las otras dos dimensiones, eficacia y eficiencia, que vamos a señalar a continuación:

- *Calidad como eficacia* que asegura que los alumnos aprenden aquello establecido y consensuado en los programas educativos en el tiempo

señalado. Trata de evaluar los resultados alcanzados por el sistema en relación con los objetivos propuestos previamente. En este caso, la estrategia metodológica se ciñe a la medición periódica de los resultados del rendimiento de los alumnos en distintos momentos del sistema en diferentes áreas curriculares. El término todavía novedoso de valor añadido utilizado cada vez con mayor asiduidad es el componente que relaciona las dimensiones de eficacia y equidad.

- *Calidad como eficiencia* que asegura que los medios, estrategias, procedimientos y recursos utilizados favorecen las experiencias educativas de los estudiantes y que permiten en último término aumentar el nivel tecnológico y económico del país. Trata de evaluar la racionalización de los costes y la relación entre los recursos utilizados, los procesos llevados a cabo y los resultados obtenidos. Cuando en el modelo de evaluación prima la eficiencia, entonces se plantea para mejorar los procesos evaluando las estrategias de gestión y didácticas realizadas en los centros educativos.

4.4. Componentes de la evaluación del sistema educativo

Aunque, al contrario de lo que sucede con los otros ámbitos mencionados, no es fácil señalar un modelo general de evaluación del sistema educativo, vamos a señalar sus componentes principales.

4.4.1. El objeto de la evaluación

El objeto de evaluación lo constituye el sistema educativo en general, pero como señala Mateo (2000a, p. 243) «para poder articular el sistema y su evaluación, para situar la acción en definitiva, precisamos percibirlo más bien como el conjunto articulado de organizaciones, que tienen como misión el desarrollar la oferta educativa sistemática de un país».

4.4.2. Los objetivos de la evaluación

Hutmacher (1999) dice que en un Estado democrático, el sistema educativo debe presentar las cuentas de su actividad y de su rendimiento tanto al parlamento como a la sociedad en general. Por ello, los sistemas educativos tienen entre sus objetivos evaluativos más corrientes los siguientes (Mateo, 2000a, p. 243):

- Evaluar hasta qué punto el sistema educativo es coherente con las necesidades de la sociedad.

- Evaluar la situación del sistema y su dinámica de desarrollo.
- Evaluar los resultados obtenidos y su nivel de logro.

En este sentido, en el estudio de evaluación de la Educación Primaria que realizó el Departamento de Educación del Gobierno Vasco en 1999 en colaboración con el INCE y con las administraciones educativas del resto de las comunidades autónomas, se plantearon los siguientes objetivos (ISEI/IVEI, 2002):

- Conocer el nivel de conocimientos de los alumnos de la etapa de Primaria en las áreas de Lengua Castellana, Matemáticas y Lengua Inglesa.
- Contrastar los resultados con los obtenidos en la evaluación anterior realizada cuatro años antes, en 1995.
- Relacionar el rendimiento del alumnado con los factores contextuales y los procesos educativos.

4.4.3. Los agentes de la evaluación

La mayoría de los Estados democráticos han creado organismos especializados en la evaluación del sistema educativo. En este sentido podemos señalar el Instituto Nacional de Calidad y Evaluación (INCE) creado en 1994 en España en cumplimiento del artículo 62 de la LOGSE (posteriormente denominado INECSE y, en la actualidad, Instituto de Evaluación tras la aprobación de la LOE), la DEP (Dirección de la Evaluación y Prospectiva en Francia), la *Audit Unit* en Escocia, la QCA (Agencia para la Calidad Curricular), en el Reino Unido, la *Skolverket* en Suecia, el Instituto para la Innovación y la Evaluación en Portugal, etc. (Mateo, 2000a). De la misma manera, en las diversas comunidades autónomas del Estado español se han ido creando paralelamente agencias encargadas de la evaluación del sistema educativo de la comunidad autónoma en cuestión. A modo de ejemplo, podemos señalar la creación del Instituto Vasco de Evaluación e Investigación Educativa (ISEI/IVEI) en la Comunidad Autónoma Vasca en 2001.

En prácticamente todos los casos, estos organismos cumplen el primero de los requisitos en la evaluación del sistema educativo señalado por Ferrer (1999), el de tener un elevado nivel técnico-pedagógico que garantice la calidad de los procesos de validez, fiabilidad y credibilidad de sus resultados. No obstante, el segundo de los principios señalado por Ferrer (1999) hace referencia a la independencia política a la hora de diseñar un modelo de evaluación del sistema educativo. En este sentido, la dependencia de estos institutos de los Ministerios de Educación correspondientes no genera toda la confianza exigible (Mateo, 2000a). Por ello, cada vez es más frecuente la petición de que estos organismos dependan directamente de los parlamentos y no de los Ministerios de Educación.

4.4.4. Las audiencias de la evaluación

En la evaluación del sistema educativo en general podemos distinguir tres audiencias interesadas en los resultados obtenidos (Mateo, 2000a):

- La sociedad y los ciudadanos en general deben tener acceso a la información generada tras la evaluación del sistema educativo como garante de su buen funcionamiento.
- La administración educativa como depositaria de la responsabilidad del óptimo funcionamiento del sistema educativo.
- Los colectivos profesionales de la educación, teniendo en cuenta desde los que ocupan puestos altos en la administración como los profesores que trabajan diariamente en el aula.

4.4.5. El proceso de la evaluación

La evaluación del sistema educativo es concebida como un proceso vinculando en todo momento la teoría con la acción y asociado a la política educativa y a los ámbitos a los que se aplica (Mateo, 2000a).

Como señala Mateo (2000a), el modelo CIPP (*Context, Input, Process y Product*) de Stufflebeam ha sido el más utilizado en la evaluación de los sistemas educativos.

Sin embargo, vamos a señalar en el proceso de evaluación algunos de los aspectos clave a tener en cuenta (García García, 2000; Mateo, 2000a):

1. *La obtención de información* suele hacerse siguiéndose alguno de estos tres procedimientos: explotar los datos disponibles en la administración educativa, los procedentes de los centros educativos u organizar procedimientos ad hoc para la consecución de la información necesaria.
2. *La determinación del contenido y del momento de la evaluación* supone responder a las preguntas de qué evaluar, quién evaluará, cómo evaluar y cuándo evaluar.
3. *La elección de los sujetos* susceptibles de ser evaluados supone decidir entre la utilización del censo completo (población total de sujetos) o de una muestra de éste, con las ventajas e inconvenientes de cada una de las opciones. Normalmente, debido a los costes elevados, se suelen elegir procedimientos muestrales utilizando algún tipo de estratificación (comunidad autónoma, zona rural o urbana, titularidad del centro, nivel socioeconómico y cultural de las familias, género de los alumnos, etc.).
4. En cuanto a *los procedimientos de medida*, deben tomarse una serie de decisiones en cuanto a:
 - Utilizar pruebas de referencia normativa o criterial.
 - Tipos de ítems incluidos en las pruebas (objetivos y/o abiertos).

- La utilización del muestreo matricial para comprobar el rendimiento de un gran número de alumnos, pero sin necesidad de administrar el mismo instrumento, ni los mismos ítems a todos los sujetos.
- La utilización de la teoría de respuesta a los ítems (TRI) que posibilita que los resultados no dependan de las características técnicas del test y de la población a la que se administra. De esta manera se posibilita aplicar distintos subconjuntos de ítems a distintos alumnos, pero sin perder la posibilidad de compararlos debido a la utilización de los denominados ítems de anclaje. De esta manera, los resultados pueden describirse en escalas de rendimiento, permitiendo referencias criterioles en distintos puntos de la escala en función de lo que los alumnos conocen o saben hacer. Así, cuando un alumno está situado en un determinado nivel podemos decir que es competente en las habilidades de dicho nivel y en las de los niveles anteriores. A modo de ejemplo presentamos la evaluación de Lengua Castellana de 6.º de Primaria realizada por el ISEI/IVEI (2002) en 1995 (véase cuadro 5.13).

5. Por último, se deberán determinar los *procedimientos de difusión de la información*.

Cuadro 5.13. Nivel y características de Lengua Castellana de 6.º de Primaria

Nivel	Características
150	<ul style="list-style-type: none"> • Comprenden literalmente textos literarios (descripciones) y verbo-icónicos (anuncios). • Realizan inferencias sobre personajes que aparecen en textos verbo-icónicos (viñetas). • Interpretan la intención comunicativa de una frase de un hablante. • Conocen el significado de palabras de uso habitual, aunque no frecuente.
200	<ul style="list-style-type: none"> • Comprenden literalmente el contenido de una carta. • Realizan inferencias sobre la importancia que tienen los personajes que aparecen en un texto informativo (noticia) y sobre la información proporcionada en un texto verbo-icónico (anuncio). • Reorganizan un texto informativo. • Hacen apreciaciones sobre un texto informativo.
250	<ul style="list-style-type: none"> • Comprenden literalmente el contenido de un texto literario (monólogo). • Realizan inferencias sobre la forma de ser de los personajes que aparecen en un texto teatral.
300	<ul style="list-style-type: none"> • Comprenden literalmente un diálogo de un texto literario (novela). • Reorganizan un texto divulgativo. • Realizan inferencias sobre el diálogo mantenido por dos personajes en un texto literario (novela).

4.5. La utilización de indicadores en la evaluación del sistema educativo

Si bien la utilización de indicadores surgió en el campo económico y social, posteriormente fue trasladado a otros campos entre los que se incluye el educativo.

Un indicador es un elemento externo de una situación, significativo e importante, que ayuda a definir y a explicar un fenómeno y permite establecer previsiones sobre su evolución futura. Por su parte, Oakes (1986) señala que un indicador es un estadístico que se refiere al sistema educativo y que revela algo sobre su funcionamiento o salud. En este mismo sentido se pronuncia Tiana Ferrer (1997) cuando define el término indicador como una medición o estimación de una característica relativa a una determinada población, con la intención de aportar una información significativa y relevante sobre su estado o situación (Mateo, 2000a, p. 254). Desde un punto de vista funcional estamos hablando de un índice estadístico obtenido a raíz de una serie de operaciones matemáticas a las que han sido sometidas un conjunto de datos sobre el sistema educativo (Mateo, 2000a). Shavelson, McDonnell y Oakes (1991) van más allá cuando señalan que un indicador que puede ser una estadística simple o compuesta relacionada con un constructo educativo básico debe ser útil en el contexto político. Es decir, aunque el indicador es una estadística, no toda estadística es un indicador. Para ello, deberá informar de algún aspecto importante y relevante del sistema y poner en evidencia cómo está funcionando, cuál es su salud (García García, 2000).

En un primer momento de desarrollo de un país, los indicadores educativos realizan la labor de señalar la cantidad de personas que están accediendo a la educación (alumnos que acceden, número de horas que pasan en la escuela, etc.). Sin embargo, tras la universalización de la educación a todos los ciudadanos, los indicadores mencionados pierden poder informativo ya que es la calidad del sistema la que interesa dado que la cantidad está ya asegurada y por ello son necesarios otro tipo de indicadores más complejos.

García García (2000, pp. 250-251) señala que en la actualidad los indicadores educativos se usan de tres maneras:

- Como estadísticos descriptivos de recursos (qué se pone) y resultados (qué produce) del sistema.
- Como sistemas de indicadores globales sobre contexto, resultados y procesos poniendo en relación factores de entrada y proceso del modelo con los resultados a nivel macro.
- Sistemas de indicadores sobre la escuela y procesos educativos que pretenden identificar escuelas eficaces.

Para terminar, vamos a señalar las posibilidades que ofrecen los indicadores como elementos de la evaluación del sistema educativo (García García, 2000; Mateo, 2000a):

- Permiten describir y comparar sistemas educativos de distintos países o incluso de distintas autonomías.
- Proporcionan información útil respecto al funcionamiento del sistema educativo que describen.
- Posibilitan analizar los cambios o la evolución de un sistema educativo proporcionando pistas para su prospectiva (permitiendo plantear hipótesis explicativas del cambio, prediciendo cambios futuros y estudiando el impacto de reformas o decisiones políticas).
- Permiten conocer los problemas, puntos críticos, necesidades o potencialidades del sistema educativo marcando posibles direcciones de trabajo.
- Ayudan a informar a la sociedad de los resultados obtenidos, recursos empleados, etc., en la educación.
- Permiten valorar las decisiones políticas sobre las acciones emprendidas.
- Ayudan a tomar decisiones sobre la política educativa del país, informando sobre los objetivos.

Preguntas y ejercicios

1. ¿Qué cambios se han dado en los últimos años en cuanto a la evaluación del aprendizaje del alumnado?
2. ¿Cuáles son las diferencias entre la evaluación tradicional y alternativa del alumnado?
3. ¿El desarrollo de las competencias en el proceso educativo va a suponer cambios en los procedimientos evaluativos?
4. ¿Qué razones han influido en el interés que suscita en la actualidad la evaluación de centros docentes?
5. ¿De qué formas diferentes puede entenderse la calidad de un centro educativo?
6. Existen diferentes criterios para caracterizar los distintos modelos de evaluación de centros. Realiza un esquema recogiendo dichos criterios y señalando el tipo de evaluación que surgiría en cada caso.
7. ¿Qué funciones puede cumplir la evaluación del sistema educativo?
8. ¿Los objetivos de las evaluaciones como PISA son los mismos que los de la evaluación diagnóstica propuesta en la LOE?

Nota

1. El propio comité define estándar como «un principio comúnmente acordado por personas implicadas en la práctica profesional de la evaluación para la medición de la valía o la calidad de una evaluación» (2001, p. 26).

6. Diseños de investigación aplicados a la evaluación

El objetivo de este capítulo es presentar algunos de los diseños utilizados en la investigación educativa, que también pueden denominarse «métodos de la investigación educativa» y que son aplicables a la evaluación de programas.

A pesar de que el concepto de «diseño» en la literatura científica ha estado asociado a la metodología experimental, como ya hemos dejado claro en el capítulo 4, partimos de una concepción de diseño, en su sentido más amplio, como aquel conjunto de decisiones a tomar en una investigación y por extensión en un proceso de evaluación de programas.

A pesar de que en el capítulo 1 han quedado claras las diferencias y las similitudes entre evaluación e investigación, vamos a incidir nuevamente en el carácter investigativo que tiene la evaluación de programas.

Siguiendo a Tejedor (2000), la evaluación de programas desde el punto de vista metodológico puede entenderse como un proceso de investigación evaluativa que implica un proceso riguroso, controlado y sistemático de recogida y análisis de información a fin de tomar decisiones sobre un programa educativo. Coincidimos con Alvira (1991) en que la investigación evaluativa lo es en la medida en que es un proceso cuya finalidad es la aplicación de procedimientos científicos para acumular evidencias válidas y fiables en la línea de determinar el valor de la acción educativa y buscar soluciones a problemas concretos.

En consecuencia, el proceso de evaluación de programas hay que concebirlo como una auténtica estrategia de investigación sobre los procesos educativos que como tal investigación puede utilizar estrategias diversas.

El número de clasificaciones de diseños es abundante, por lo que las posibilidades de plantear una clasificación de los posibles métodos son muchas. Una forma sencilla, y a la vez clásica, de clasificar los distintos métodos podría ser diferenciar entre diseños cuantitativos y cualitativos. De hecho en cantidad de manuales de metodología de la investigación se utiliza. Sin embargo, este criterio de clasificación no está exento de problemas. Así por poner algún ejemplo podría citarse el del estudio de casos que para algunos autores puede ser un método cuantitativo (diseños $N = 1$) y para otros es claramente cualitativo (el estudio etnográfico de casos). Un criterio más novedoso y más relacionado con la Educación podría ser el de dividir los métodos entre empíricos y no empíricos. Este criterio tiene el problema de la concepción que se tenga de lo empírico, que es un término controvertido acerca de cuyo significado no hay una completa unanimidad. En el apartado 3.2 del capítulo 4 («Criterios para la clasificación de los diseños») hemos presentado más modelos de diseños aplicables a la evaluación de programas. Ante esto hemos decidido presentar una lista con los diseños más utilizados en la investigación en Educación aplicables a la evaluación, destacando las características más importantes de cada uno de ellos.

La elección de un diseño u otro vendrá determinada por distintas circunstancias, como pueden ser: los objetivos de la investigación, los recursos, el tiempo disponible, etc.

En concreto, los diseños que vamos a describir son los siguientes: descriptivos, correlacionales, experimentales, etnográfico, estudio de casos e investigación-acción.

1. Diseños descriptivos

La investigación basada en métodos o diseños¹ descriptivos constituye el primer nivel del conocimiento científico. Aunque no existe acuerdo entre los distintos autores, se suele identificar este tipo de investigación con lo que Fox (1981) denominaba «aproximación descriptiva». El objetivo básico de la investigación descriptiva es la descripción de los fenómenos, hechos y situaciones educativos.

Desde el punto de vista de la Metodología Experimental, una de las características de los métodos descriptivos es la falta de manipulación en el sentido de que no se pretenden cambios en el objeto de estudio, sino únicamente recoger información relevante acerca de ese objeto, analizarla y llegar a unas conclusiones.

Los métodos descriptivos suelen utilizarse muchas veces como paso previo a investigaciones de tipo correlacional o experimental, ya que nos pueden ofrecer un mayor y mejor conocimiento del objeto de estudio. Generalmente, la investigación descriptiva parte de unos supuestos teóricamente muy pobres acerca del problema a investigar. Por medio de la recogida de datos se puede

llegar a un conocimiento más profundo de dicho problema que más adelante puede permitir la comprobación empírica de ciertas hipótesis que en un primer momento no podrían formularse.

Antes de continuar adelante, quisiéramos mencionar que en este caso, al hablar de métodos descriptivos, nos estamos limitando a un conjunto de métodos basados en la metodología cuantitativa. No incluimos entre ellos a otros como el etnográfico, que aunque también tiene como objetivo la descripción de fenómenos utiliza la metodología cualitativa, por lo que le daremos un tratamiento diferenciado.

Como algunas características de los métodos descriptivos se pueden mencionar las siguientes:

- Se suelen realizar en ambientes naturales y no en los laboratorios. El laboratorio en investigación educativa debemos entenderlo en su sentido más amplio y no en el de ubicación física de la investigación. Se puede entender por laboratorio una situación en la que, mucho o poco, se han alterado las condiciones normales de funcionamiento de una determinada realidad. En este sentido bastaría con alterar la composición de un grupo de alumnos y alumnas de un determinado centro, para considerar que esa nueva situación es diferente de la que es habitual. De alguna manera se ha artificializado la situación.
- A pesar de que los métodos descriptivos permiten investigar prácticamente cualquier tema, tienen serias limitaciones para establecer relaciones causales entre variables.
- Utilizan el método inductivo.
- Resultan apropiados para identificar problemas relevantes y/o hipótesis.

En cuanto a las fases del proceso de investigación se presenta a continuación un esquema general de todo el proceso. Como se verá más adelante existen diferentes tipos de métodos o estudios descriptivos. Cada uno de ellos adecuará este esquema general a sus peculiaridades. Por lo tanto un posible esquema podría ser el siguiente:

- Formulación del problema de investigación.
- Formulación de los objetivos de la investigación.
- Planteamiento metodológico:
 - Definición de variables.
 - Selección o elaboración del o de los instrumentos de recogida de datos.
 - Selección de la muestra.
- Recogida, análisis e interpretación de los datos.

- Redacción del informe de investigación.
- A la hora de clasificar los diversos métodos descriptivos nos encontramos con el problema de la heterogeneidad de procedimientos que pueden encuadrarse dentro de esta categoría.

No obstante, siguiendo a Hernández Pina (1995) y Colás y Buendía (1998) hemos distinguido cuatro tipos fundamentales: los estudios tipo encuesta, los estudios analíticos, los estudios observacionales y los estudios de desarrollo. Ésta es la clasificación que desarrollaremos en las siguientes líneas, ya que, además de no incluir los métodos correlacionales que pensamos que necesitan un tratamiento aparte, nos parece que es la que mejor distingue, dentro de los métodos descriptivos, entre métodos o formas de proceder independientes. Veamos, pues, las características de cada uno de estos métodos.

1.1. Estudios tipo encuesta

Estos estudios también conocidos como *survey* consisten en la recopilación de datos en un momento concreto con la intención de:

- a) Describir la naturaleza de las condiciones existentes.
- b) Identificar normas o patrones contra los que se puedan comparar las condiciones existentes.
- c) Determinar las relaciones que existen entre acontecimientos específicos.

(Cohen y Manion, 1990, p. 131)

Latorre, Del Rincón y Arnal definen estos estudios de la siguiente forma:

[...] método de investigación basado en una serie de preguntas dirigidas a sujetos que pueden constituir una muestra representativa de una población con el fin de describir o relacionar características personales y ciertos ámbitos de información necesarios para responder al problema de investigación, como pueden ser: intereses, motivaciones, creencias, actitudes, intenciones, deseos, percepciones, vivencias y conductas personales o de grupo.

(1996, p. 182)

Como puede verse, básicamente lo que se pretende es obtener información descriptiva de una población a partir de una muestra representativa de ella. Las fases de un estudio de encuesta son las siguientes:

- Definición de los objetivos de la investigación.
- Decisión de la información necesaria.
- Selección de la muestra.

- Diseño del instrumento.
- Estudio piloto.
- Recogida de la información.
- Análisis e interpretación.
- Redacción del informe.

Las principales técnicas de recogida de datos son los cuestionarios, las entrevistas estructuradas o semiestructuradas, los tests estandarizados y las escalas de actitudes. En cuando a la selección de los sujetos y dado que lo que se persigue es llegar a amplias poblaciones se utilizan sistemas de muestreo fundamentalmente aleatorios en sus distintas variantes: simples, estratificados, por conglomerados, etc. Este problema de la selección de la muestra constituye uno de los aspectos más importantes por el hecho de recoger opiniones representativas y no sesgadas del grupo de personas sobre el que se investiga (Cea, 1996). Por último y en cuanto al análisis de los datos recogidos, las técnicas básicas son las propias de la estadística descriptiva: medidas centrales y de desviación, frecuencias, porcentajes, etc., y en algún caso técnicas como las correlaciones o la chi cuadrado.

Se puede decir que, además de la elección adecuada de la muestra, la mayor dificultad técnica de este tipo de estudios está en la construcción del instrumento de recogida de datos, lo cual es costoso ya que debe cumplir una serie de requisitos, tales como minimizar los errores potenciales de los informantes, atraer el interés y la cooperación de las personas que van a responder y que las respuestas sean lo más cercano a la realidad; en definitiva que garantice su fiabilidad y validez.

1.2. Estudios observacionales

En nuestra vida cotidiana utilizamos la observación de una manera continua a fin de adquirir nuevos conocimientos. Sin embargo, estas observaciones en la mayoría de los casos no cumplen los requisitos necesarios para ser consideradas como métodos de investigación. Entre estos requisitos podemos destacar los siguientes:

1. Sirve a un objetivo de investigación ya formulado en ella.
2. Está planificada ya que previamente ha delimitado lo que se va a observar y en qué condiciones ocurrirá la observación, así como las conductas que deben registrarse y cómo hacerlo.
3. Está relacionada con proposiciones o conceptos más generales como hipótesis o teorías.
4. Es controlada en el sentido de que es objetiva y está sometida a comprobaciones para asegurar la validez y fiabilidad de la información recogida.

El proceso de los estudios observacionales puede resumirse en las siguientes fases o etapas:

- Formulación del problema y de las hipótesis.
- Identificación de la muestra.
- Selección de los observadores y las técnicas de observación.
- Formación de los observadores.
- Estudio piloto.
- Diseño del plan de recogida de datos.
- Recogida de datos.
- Análisis e interpretación de los datos.
- Redacción del informe.

Los aspectos que ofrecen mayor especificidad a estos métodos y que deben ser desarrollados con mayor atención son el tipo de observación, las técnicas de registro, el muestreo y la selección y la formación de observadores.

Existen diferentes tipos de observación según los criterios de clasificación que se utilicen. Así se pueden destacar los siguientes grupos de tipos de observación (Anguera, 1989):

- no sistemática u ocasional / sistematizada o controlada / muy sistematizada;
- natural o espontánea / de laboratorio;
- participante / no participante / autoobservación.

En cuanto a las técnicas de registro de datos, los más habituales son el registro anecdótico, el sistema de signos, el sistema de categorías, el vídeo y las escalas estimativas.

No obstante, en el primer apartado del capítulo 7 se presenta de manera más pormenorizada el tema relativo a la observación como estrategia de recogida de información.

En cuanto al denominado muestreo, en el caso de los estudios observacionales se suelen utilizar tres procedimientos específicos:

- a) El muestreo de tiempo supone tomar decisiones sobre cuánto se va a observar, con qué frecuencia y en qué momentos.
- b) El muestreo de situaciones por el que se eligen las situaciones que van a permitir generalizaciones de las conductas o fenómenos observados.
- c) El muestreo de sujetos que como el anterior tiene también por objeto garantizar la generalización de resultados.

La selección y la formación del personal que va a observar constituye uno de los apartados cruciales de este método. Aspectos tales como las características,

proceso de selección y entrenamiento, problemática de la interacción observador-observado constituyen cuestiones que tienen gran importancia con vistas a la fiabilidad de los datos recogidos.

El método observacional ha tenido y tiene una gran importancia en el campo de la Educación; entre otras razones, debido a su capacidad de amoldarse a cantidad de hechos y circunstancias en sus contextos naturales. Otra de las grandes ventajas de este método es la posibilidad que ofrece de realizar investigaciones y evaluaciones en aquellos ámbitos en los que no es posible informarse de una forma verbal (niños pequeños, retrasados...).

1.3. Estudios analíticos

Estos estudios se caracterizan porque tratan de extraer información a través de análisis minuciosos y profundos de determinados contextos, sujetos y materiales. Sus aplicaciones cubren distintos ámbitos educativos como estudios sobre la composición factorial de un conjunto de datos, construcción de instrumentos de recogida de datos y análisis de contenido de discursos orales, visuales e impresos, utilizándose para ello técnicas como el análisis factorial, correlaciones y análisis de contenido.

El proceso de investigación se puede sintetizar en las fases siguientes:

- Definición de los objetivos de la investigación.
- Determinar la unidad de contenido analizable.
- Elaboración de hipótesis de trabajo para su contraste.
- Elaboración de indicadores.
- Recogida de datos.
- Análisis e interpretación de los datos.
- Redacción del informe.

Algunas aplicaciones de este método a temas educativos son las siguientes: análisis de contenidos de textos escolares, análisis de documentos curriculares de un centro, nivel de lectura de materiales impresos, etc.

1.4. Estudios de desarrollo

Estos estudios se caracterizan porque pretenden determinar el desarrollo o el cambio que puede sufrir un determinado fenómeno a lo largo del tiempo. Los rasgos principales que definen este método son, por una parte, el interés por los cambios de comportamiento y, por otro, el interés por la descripción y explicación de ese cambio. En general, son estudios centrados en diferencias unidas a la edad; por ejemplo, las investigaciones de la corriente cognitivo-evolutiva sobre el pensamiento infantil de Piaget.

El proceso de investigación que se sigue en estos estudios no ofrece variaciones sustanciales con el modelo general, por lo que puede quedar de la siguiente forma:

- Definición del problema.
- Revisión bibliográfica.
- Diseño.
- Recogida de datos.
- Análisis e interpretación de los datos y resultados.
- Redacción del informe.

Entre los estudios de desarrollo se distinguen dos diseños básicos, los estudios transversales y los longitudinales.

En los estudios **transversales** se describe el estado de un determinado fenómeno en dos o más grupos diferentes que se encuentran en distintos momentos de maduración. La característica de este método estriba en que se recogen datos de todos los sujetos simultáneamente; es decir, sin que transcurra el tiempo. La ventaja de este tipo de estudios radica en su economía ya que en un breve espacio de tiempo se pueden analizar los cambios que presumiblemente se suelen dar como consecuencia del cambio a lo largo del tiempo, aunque hay que reconocer que no controla el efecto que puede producir el propio tiempo en las distintas generaciones. Otras ventajas de los métodos transversales son la mayor facilidad para obtener la cooperación de los sujetos y la posibilidad de trabajar con un número mayor de ellos. Este método es aplicable en temáticas propias de la pedagogía diferencial así como a fin de obtener información relevante para la planificación de programas educativos, como, por ejemplo, conocer los cambios generacionales en determinados comportamientos y actitudes. Sin embargo, resultan poco efectivos para identificar las variaciones individuales o para establecer relaciones causales entre variables.

En los estudios **longitudinales** se estudia a un mismo grupo de sujetos en diversos momentos, a lo largo del tiempo, para conocer la evolución que van manifestando. Estos estudios permiten describir las evoluciones intra-individuales aunque tienen algunos inconvenientes. El primero de ellos es que pueden llegar a ser demasiado costosos para la información que se va a lograr. Otro inconveniente puede ser que los momentos elegidos para la recogida de los datos no coincidan con los cambios más significativos con lo que la información recogida no sería relevante. Por último, mencionar los problemas derivados de las denominadas «mortalidad experimental» y «efecto de la administración de la prueba» o el efecto de los cambios de motivación a lo largo del tiempo. En definitiva, los mayores problemas vienen de la dificultad de mantener la planificación inicial del estudio debido a los cambios que se producen a lo largo del tiempo. A pesar de estos inconvenientes, estos métodos

ofrecen grandes posibilidades en Educación. Entre las ventajas se pueden mencionar las siguientes:

- Resultan adecuados cuando se intentan establecer relaciones causales en una primera fase, ya que se puede demostrar que ciertas variables (las supuestamente causa) preceden en el tiempo a otras (supuestamente efecto). Esto puede ser punto de partida para una investigación experimental.
- Permiten obtener información acerca de los cambios producidos por una determinada intervención educativa.

Como caso particular de los estudios longitudinales cabe mencionar los denominados **estudios descriptivos de caso único**, cuya principal característica es que, en lugar de centrarse en un grupo de sujetos, se centra en un único caso. Ello evidentemente reduce considerablemente las posibilidades de generalización de los resultados.

Por último, tenemos los denominados **estudios de cohortes** o de panel. Mientras que para algunos no son más que sustituciones del término longitudinal, para otros son una combinación de los transversales y longitudinales. Según estos últimos se trata de una sucesión de estudios transversales que se repiten en diversos periodos de tiempo pero considerando cada vez a sujetos distintos. Esta última característica, la de la observación en cada una de las ocasiones a sujetos distintos es lo que les diferencia de los estudios longitudinales pero al mismo tiempo y dado que se prolonga en el tiempo, les priva de un carácter totalmente transversal.

2. Diseños correlacionales

Algunos autores consideran que los diseños correlacionales son una variante o están incluidos dentro de los diseños descriptivos. Sin embargo, hay otros que consideran que estos métodos tienen sus propias especificidades y deben ser analizados independientemente de los descriptivos. En este libro se adopta la segunda alternativa, es decir, la de considerar los diseños correlacionales como independientes, con entidad propia independiente de los diseños descriptivos y los experimentales.

Estos diseños incluyen todas aquellas investigaciones y evaluaciones que intentan descubrir o clarificar la relación entre variables a través del uso de procedimientos estadísticos correlacionales. Básicamente lo que se hace es recoger datos acerca de algunas variables y ver la intensidad de la relación entre ellas, en condiciones naturales sin que el investigador o el evaluador provoque ningún cambio. Esta falta de manipulación de variables es lo que posiblemente ha hecho que algunos autores hayan incluido estos métodos dentro de los descriptivos en oposición a los experimentales.

La investigación correlacional se caracteriza por plantear sus hipótesis u objetivos de investigación, sobre hechos y fenómenos ya acontecidos, o que no son intrínsecamente manipulables, con el fin de descifrar y analizar las relaciones subyacentes entre ellos, tratando de aproximarse a la causalidad de los fenómenos.

En realidad, estos diseños se podrían situar a caballo entre los métodos descriptivos y los experimentales en cuanto que superan el objetivo de la mera descripción tratándose de acercarse a las explicaciones causales.

Pérez Juste (1985) plantea los métodos correlacionales como previos a la utilización de los experimentales, ya que el conocimiento de la existencia o no de correlación entre variables no tiene valor en sí mismo si ello no tiene continuidad en la búsqueda de las razones de esa correlación, lo cual puede darse a través de métodos más rigurosos. En esta línea, los estudios correlacionales pueden ser de gran utilidad al permitir identificar posibles variables extrañas a controlar o conocer el peso relativo en la predicción de las variables independientes según el criterio elegido por el investigador.

La investigación correlacional es una alternativa válida y de gran utilidad en aquellas situaciones en las que no es posible realizar una investigación con las exigencias de los diseños experimentales o cuasiexperimentales. Para Mateo y Rodríguez (1984), desde la investigación correlacional se puede abordar la problemática de la causalidad, teniendo en cuenta algunas circunstancias como las siguientes:

- a) Covariación temporal entre la presunta causa y el presunto efecto.
- b) Orden temporal adecuado, precediendo la causa al efecto.
- c) Eliminación de explicaciones alternativas verosímiles para las relaciones observadas.

Las técnicas estadísticas avanzadas (regresión múltiple, análisis de sendas, análisis causal, etc.) han permitido trabajar en las investigaciones con grandes cantidades de datos y variables y contrastar modelos teóricos que proporcionen ciertas explicaciones de complejos fenómenos educativos (Álvaro, 1990; Fuentes, 1986).

Entre las aplicaciones de los diseños correlacionales podemos encontrar las siguientes: conocer la magnitud de la relación entre variables, descubrir el grado de predicción de determinadas variables, elaborar y refinar instrumentos de medición a través de los análisis de fiabilidad y validez de dichos instrumentos, ver diferencias entre grupos naturales en cuanto a distintas variables, recopilación de datos necesarios para establecer hipótesis, etc.

En cuanto al proceso o fases de la investigación basada en el diseño correlacional, se puede decir que no difiere significativamente de la descriptiva. Por lo que el esquema podría quedar de la forma siguiente:

- Formulación del problema de investigación.
- Formulación de hipótesis.
- Planteamiento metodológico:
 - Identificación y definición operativa de las variables.
 - Selección o elaboración de los instrumentos de recogida de datos.
 - Selección de la muestra.
- Recogida de los datos.
- Análisis estadístico de los datos e interpretación.
- Redacción del informe de investigación.

Las clasificaciones de los diseños correlacionales generalmente suelen estar ligadas a las técnicas estadísticas utilizadas para el análisis de los datos recogidos. Colás y Buendía (1998) presentan la clasificación siguiente:

- *Estudios comparativo-causales*. Estos estudios trabajan con grupos que presentan características diferentes por sí mismos, y a partir de dichos grupos tratan de obtener una correlación entre las diferentes variables estudiadas. Al no existir manipulación y control de la variable independiente no se podrá nunca decir que una variable es causa de la otra. Como mucho puede servir para detectar factores que parecen hallarse asociados a ciertos hechos.
- *Estudios de relación o correlacionales*. Se pretende averiguar la magnitud de la relación entre dos o más variables. Los análisis se basan en coeficientes de correlación diferentes: de Pearson, biserial, biserial-puntual, phi, parcial, chi cuadrado, etc. Normalmente, se han aplicado en la elaboración de técnicas de recogida de datos para el análisis de ítems, cálculo de fiabilidad y validez, etc.
- *Estudios predictivos*. Intentan predecir ciertas conductas conociendo el valor de otras predictoras. Se basan en la correlación y la regresión. Entre las técnicas más utilizadas se encuentran la regresión múltiple, la regresión polinómica, las técnicas de segmentación y la regresión múltiple no lineal. A través de éstas se define un grupo de variables como predictores y otro como variable o variables criterio con el fin de ver los niveles de contribución de cada variable predictora al pronóstico de la variable criterio. Además de para la validación de instrumentos de medida, sus mayores aplicaciones se han dado en el campo de la orientación escolar y profesional y el diagnóstico pedagógico.
- *Estudios basados en el análisis factorial*. Estos estudios pretenden conocer la estructura interna de determinados factores, a través de la detección de aquellos factores que explican un número más amplio de variables. Cuando esta búsqueda es más empírica que teórica se utiliza el análisis de componentes principales. Su mayor aplicación ha sido en el análisis y el refinamiento de instrumentos de medida.

- *El análisis discriminante*. Permite diferenciar grupos de individuos según sus perfiles característicos en un conjunto de variables. Este método presenta claras aplicaciones en pedagogía diferencial en el establecimiento de grupos diferenciados de individuos como base para diseñar programas o modelos educativos.
- *Análisis de conglomerados o clusters*. Consiste en clasificar a los sujetos en grupos o conglomerados lo más similares posible entre sí y lo más diferentes posible en relación con otros grupos. Además de su utilización en la elaboración y refinamiento de instrumentos de medida, se ha aplicado también para detectar constructos personales y percepciones de grupos sobre temáticas educativas. A éstos y a los anteriores también se les suele llamar *estudios de clasificación*.
- *Análisis de sendas*. Sirven para contrastar la plausibilidad o coincidencia empírica de modelos causales teóricos. Su principal objetivo es estudiar los efectos directos e indirectos de variables, consideradas causas, sobre variables efecto, sin manipular tales variables. Se pueden considerar como una alternativa los experimentos puros que son tan difíciles de llevar a la práctica en educación, ya que aunque no están diseñados para descubrir relaciones de tipo causal sí lo están para comprobar el ajuste o coincidencia a modelos causales previamente planeados.

Por último, una mención a las ventajas y a los inconvenientes de los métodos correlacionales. Entre las ventajas se pueden mencionar las siguientes: su utilidad en aproximaciones exploratorias cuando el conocimiento del fenómeno que se va a estudiar es escaso, no se necesitan muestras excesivamente amplias y permiten trabajar con un número elevado de variables estudiando las relaciones entre ellas de forma simultánea. Entre los inconvenientes se podría mencionar la imprecisión de los coeficientes de correlación que están muy limitados por la fiabilidad en la recogida de los datos.

3. Diseños experimentales

Los métodos experimentales son los más directamente relacionados con la perspectiva positivista acerca de la investigación y de la evaluación. De hecho hay que reconocer que los inicios de la investigación y de la evaluación en Educación se dan basándose en las investigaciones experimentales. Son los diseños experimentales los que reflejan de una forma más fidedigna u ortodoxa el proceso de investigación del modelo hipotético-deductivo de la ciencia. La investigación experimental se asienta sobre los presupuestos del método científico, tal como se aplica en las Ciencias Naturales.

La esencia del diseño experimental es que el investigador o evaluador deliberadamente manipula o controla una serie de condiciones (variable independiente) que influirán en los fenómenos que se pretenden observar (varia-

ble dependiente) a la vez que se mantienen controladas otras variables (variables extrañas o intervinientes).

La variable dependiente es el fenómeno que se estudia, el efecto que pretendemos encontrar como consecuencia de las modificaciones realizadas por el investigador en la variable independiente. Una buena medición de esta variable dependiente es imprescindible para que se pueda confirmar o rechazar que los cambios o efectos son debidos a la variable independiente. En la investigación o evaluación experimental en Educación, la variable dependiente suele recoger aquellos comportamientos o conductas de los sujetos que se dan como consecuencia de programas de intervención que, a la vez de buscar conocimiento objetivo, pretenden una mejora de la realidad educativa.

La variable independiente es aquella sobre la que el investigador o evaluador interviene. En esta intervención se modifican intencionalmente los valores de dicha variable para estudiar los efectos de esa modificación en la dependiente. Es aquí donde se da la llamada manipulación experimental que no es otra cosa que una modificación de las condiciones en que normalmente se dan unas situaciones (en nuestro caso educativas) para estudiar lo que ocurre tras esa modificación. En un lenguaje más psicológico a la variable independiente se le suele denominar «estímulo» (a la dependiente «respuesta»). En Educación, estos estímulos suelen ser muchas veces intervenciones basadas en programas educativos de distinta índole, como, por ejemplo, programas basados en el desarrollo de habilidades diferentes: cognitivas, de valores, de personalidad, sociopersonales, afectivas...

Por último, las variables extrañas o intervinientes son aquellas sobre las que el investigador no actúa directamente en un principio, pero que pueden influir en la variable dependiente que se pretende estudiar, de forma tal que interactúa con la variable independiente. En este caso se puede decir que la relación causa (variable independiente)-efecto (variable dependiente) queda contaminada (estas variables también son denominadas variables contaminadoras) por fuentes de variación ajenas a la variable independiente. El investigador o evaluador debe intentar controlar o neutralizar el efecto de estas variables, para lo que se pueden plantear diferentes técnicas.

Es la característica del «control» la que marca la diferencia fundamental entre el diseño experimental y los descriptivos y correlacionales. Este control debe ser entendido en una doble vertiente, como manipulación y como control de variables.

Por una parte, podemos hablar de control como manipulación de la variable independiente en el sentido de que las variaciones que se den en esa variable van a ser previamente determinadas por el investigador que espera que en función de los cambios en esa variable independiente se den variaciones en la variable dependiente. Puesto que la finalidad última de las investigaciones experimentales es establecer relaciones de causa-efecto, la manipulación de la presunta causa tiene como objetivo observar los cambios que ella provoca en el supuesto efecto.

Por otra parte, este control se extiende a todo el proceso, evitando la aparición de otras variables relevantes que pueden influir en los resultados, y si aparecieran, intentando que su influencia sea mínima. Lo que se pretende es poder llegar a establecer con cierto grado de seguridad que las variaciones en la variable dependiente son como consecuencia de los cambios en la variable independiente, y sólo por ellos, para lo que el investigador o evaluador debe ejercer un buen control sobre la situación del estudio. Una forma de realizar este control es recurrir a la aleatorización o asignación aleatoria de sujetos a los distintos grupos de la investigación o evaluación. Aunque existen otros criterios o técnicas de control, la ventaja de la aleatorización es que no es necesario conocer cuántas variables extrañas existen, ni cuáles son, ni cómo influyen en la variable dependiente.

Uno de los objetivos fundamentales del diseño experimental es la comprobación de hipótesis. Las hipótesis experimentales reflejan la relación entre la posible causa o causas (variable independiente o independientes) y el posible efecto o efectos (variable dependiente). En algunos casos se trata de una hipótesis propuesta para aumentar el conocimiento científico en campos todavía poco conocidos. En otros, lo que se pretende es decidir entre dos hipótesis contrarias para llegar a la explicación correcta de un fenómeno. En tercer lugar, la hipótesis formulada puede servir para confirmar la validez de una teoría previa para lo cual se suelen replicar experiencias realizadas anteriormente. Estas tres posibilidades corresponderían a los tres tipos de experimentos que diferencia Arnau (1982): experimentos exploratorios, cruciales y confirmatorios.

El proceso de investigación de los diseños experimentales es el que se adecua de una manera más rigurosa a las fases del método científico de las Ciencias Naturales. Los pasos principales en el planteamiento de una investigación experimental son los siguientes:

- Identificación y formulación del problema de investigación.
- Formulación de la o las hipótesis de investigación.
- Planteamiento metodológico:
 - Definición de las condiciones experimentales o niveles de la variable independiente.
 - Definición de los procedimientos de recogida de datos acerca de la variable dependiente.
 - Definición de las técnicas de control de variables extrañas.
 - Selección de la muestra.
 - Diseño y planificación del experimento y la recogida de datos.
- Realización del experimento y recogida de datos.
- Análisis estadístico de los datos para contrastar las hipótesis.
- Formulación de conclusiones, sugerencias, etc.
- Redacción del informe.

Entre los experimentos, podemos distinguir entre experimento de laboratorio y experimento de campo. Mientras este último se realiza en situaciones reales, el experimento de laboratorio supone que el investigador decide las condiciones en que se va a realizar aquél con vistas al mayor control posible de todas las variables.

Entre las ventajas de los experimentos de laboratorio o controlados se puede destacar la posibilidad de obtención y registro de datos con gran precisión, lo cual posibilita una mayor probabilidad de validez de las conclusiones. Otra de las ventajas es la posibilidad de repetir el experimento en condiciones idénticas. Sin embargo, tienen el grave problema de su artificialidad en cuanto a las condiciones en las que se desarrolla, lo cual en muchas ocasiones invalida su aplicación en el campo de la educación por su falta de sintonía con lo que ocurre en la realidad. Además, cuestiones éticas impiden la manipulación experimental de muchas variables de interés para la investigación educativa.

Los experimentos de campo tienen la ventaja de una mayor generalización o transferibilidad a situaciones educativas debido a su realización en contextos reales o muy similares a los reales. Esto hace que sean aplicables a gran variedad de problemas, existiendo además la posibilidad de estudiar fenómenos complejos en situaciones similares a como se dan de una forma natural. Sin embargo, adolecen del problema de que al disminuir el grado de control también disminuye su validez interna. Kerlinger (2002) considera estos experimentos como apropiados para la demostración de teorías y la resolución de problemas prácticos.

Aunque hay diferentes criterios para mencionar los tipos de validez, podríamos decir que desde la perspectiva experimental fundamentalmente se habla de tres tipos de validez: la validez interna, la validez externa y la validez teórica o conceptual. Definiremos brevemente cada una de estas variantes o aspectos de la validez.

- *Validez teórica o conceptual.* Se dice que un experimento o investigación tiene validez teórica cuando está ligada a un cuerpo o marco teórico de forma coherente y asimismo aporta algo a ese marco.
- *Validez interna.* Se refiere a que desde el planteamiento del problema hasta la formulación de las conclusiones, los pasos recorridos metodológicamente y técnicamente son correctos.
- *Validez externa.* Es la capacidad del diseño de aplicar las conclusiones obtenidas a otros sujetos, poblaciones, casos, condiciones, situaciones, etc.

En relación con el diseño en cuanto a garante de la calidad de la estrategia utilizada para aclarar las dudas de investigación, es a la validez interna a la que se debe dar prioridad ya que nos garantizará que la investigación esté bien hecha y además es requisito para que se den las otras dos formas de validez.

A partir de estas ideas se han desarrollado diferentes estrategias para asegurar la validez de los diseños.

En cuanto a los tipos de diseños, se pueden hacer diferentes clasificaciones según el criterio que se utilice. He aquí algunas de ellas:

1. *Según el número de grupos:*
 - Diseño entregrupos o entresujetos: se dan comparaciones entre grupos compuestos de sujetos distintos.
 - Diseño intragrupos o intrasujetos: se compara un grupo consigo mismo en momentos diferentes. Una variante de estos diseños son los denominados diseños intrasujeto ($N = 1$).
 - Diseño mixto: se dan comparaciones entre grupos y dentro de un mismo grupo.

2. *Según el número de variables independientes:*
 - Diseño unifactorial: solamente se manipula una variable independiente.
 - Diseño factorial: se manipulan por los menos dos variables independientes.

3. *Según las técnicas de control de variables extrañas utilizadas:*
 - Diseño aleatorio: se ha utilizado la aleatorización.
 - Diseño de bloques: se ha utilizado el bloqueo.
 - Diseño de series temporales: se han utilizado diferentes mediciones a lo largo del tiempo.

4. *Según el grado de experimentalidad o nivel de control llevado a cabo:*
 - Diseño preexperimental: se puede hablar de algún nivel de manipulación de la variable independiente, pero el nivel de control de las variables secundarias es ínfimo.
 - Diseño cuasiexperimental: a pesar de que se da manipulación de la variable independiente y de que el nivel de control es bastante alto, al no darse la selección al azar quedan sin controlar algunas variables extrañas.
 - Diseño experimental: se da el máximo nivel de control y además se utiliza la aleatorización para la selección y asignación de sujetos.

Centrándonos en esta última clasificación, habría que hacer algunas puntualizaciones. Los diseños **preexperimentales** o seudoexperimentales, en cuanto que no ejercen un control suficiente sobre las variables secundarias y en consecuencia no garantizan la validez de las conclusiones, son diseños no

recomendables en la investigación educativa, ya que sus aportaciones son escasas.

En cuanto a los diseños **experimentales** o verdaderos experimentos, hay que decir que son los más potentes en cuanto a validez interna y, por lo tanto, los más deseables si queremos establecer relaciones de causa-efecto, como suele ocurrir cuando queremos conocer los efectos de una intervención educativa en un determinado ámbito. Sin embargo, su aplicación no es fácil en estudios educativos, ya que las condiciones exigidas suelen ser difíciles de cumplir, principalmente en los experimentos de campo, en los que en la mayoría de las ocasiones los grupos ya creados no pueden ser modificados por el investigador o evaluador.

Dentro de los diseños de corte experimental, son los diseños **cuasiexperimentales** los más utilizados en la investigación educativa. Estos diseños carecen del alto nivel de control de los diseños experimentales, ya que no se cumple el principio de la asignación aleatoria de los sujetos a los grupos y tratamientos. Al no darse esta asignación aleatoria no se garantiza la semejanza entre los grupos pero sí garantizan un control aceptable a través de diversas técnicas como el grupo control o la inclusión del pretest, entre otras, sin olvidar el desarrollo de procedimientos como el análisis de covarianza o regresión múltiple que, aunque a posteriori, también sirven para controlar. Sin embargo, hay autores que afirman que los diseños cuasiexperimentales suponen el grado máximo de control para el estudio de la Educación.

El recurso de los diseños cuasiexperimentales nos proporciona cierta flexibilidad, necesaria en el campo de la Educación, dado que siempre se interviene, se toman datos, se compara y se extraen conclusiones sobre la optimidad o no de determinados programas educativos, y si éstos se realizan en contextos formales de educación el acceso a los grupos, aunque naturales o ya formados, no suele ser demasiado costoso. Esto, unido a la utilización de escenarios naturales, los hace ser de gran aplicación en el campo educativo por la posibilidad que ofrecen de poder llegar a establecer resultados generalizables a ámbitos reales.

Entre los diseños cuasiexperimentales podemos destacar los siguientes:

1. *Diseños pretest-postest con grupo único.*
2. *Diseños pretest-postest con grupo de control no equivalente.*
3. *Diseños de series temporales simple.* Un solo sujeto o un grupo en el que se incrementan las medidas previas y posteriores al tratamiento.
4. *Diseños con grupo de control no equivalente.* Dos grupos no aleatorios: uno como en el caso anterior y otro de control.
5. *Diseños con variables dependientes no equivalentes.* Un solo grupo al que se aplican medidas diferentes en función del tiempo. Se espera que una de ellas cambie y la otra no.
6. *Diseños con réplicas múltiples.* Un solo grupo al que se aplican tratamientos múltiples para observar el efecto.

7. *Diseños con réplicas intercambiadas*. Dos grupos en los que ambos reciben tratamiento de forma alternativa: cuando uno hace de control otro recibe tratamiento, y viceversa.

En cuanto a los campos de interés en el ámbito educativo y a los que pueden vincularse los métodos experimentales, se podrían mencionar los siguientes (Tejedor, 1994):

- Problemas referentes al estudio y aprendizaje.
- Problemas de didáctica.
- Problemas de eficacia de métodos y procedimientos para mejorar el rendimiento.
- Problemas derivados del estudio de las diferencias individuales y su tratamiento adecuado.

Por último, hay que reconocer que la aplicación estricta de la metodología experimental puede llevar a tales niveles de artificialidad que no parece que deba ser el objetivo preferente para la investigación educativa desde el punto de vista metodológico. En este sentido son los experimentos de campo y/o los de corte cuasiexperimental los que pueden ofrecer aportaciones valiosas a la Educación, ya que al estar realizados en contextos naturales gozan de una alta validez externa.

4. Método etnográfico

En un sentido amplio del término, la expresión «investigación etnográfica» ha venido utilizándose como expresión equivalente a investigación cualitativa, investigación naturalista, trabajo de campo, observación participante, etcétera. En la medida en que para algunos la investigación etnográfica es sinónimo de investigación cualitativa, más que un método específico de investigación es considerado como una estrategia general que engloba diferentes tipos o métodos.

En sentido literal, significa la descripción de un modo de vida de una raza, o grupo de individuos (Woods, 1987) o de un grupo o cultura (Fetterman, 1989). Sus orígenes se sitúan a finales del siglo XIX y comienzos del XX, en algunos trabajos de antropólogos acerca del modo de vida de las tribus. Es entonces cuando se comenzaron a hacer los estudios de campo en los que los antropólogos se desplazaban a las tribus para estudiarlas «desde dentro».

Wolcott define la etnografía como «proceso» de investigación por el que se aprende el «modo de vida» de algún grupo identificable de personas como al «producto» de ese esfuerzo: un escrito etnográfico o retrato de un modo de vida (García Jiménez, 1994b, p. 344).

La investigación etnográfica consiste en la descripción de los sucesos que ocurren dentro de la vida de un grupo, con especial atención a las estructuras sociales y a la conducta de los individuos respecto a su pertenencia al grupo. Además se da gran importancia a la interpretación del significado de los mismos en relación a la cultura del grupo.

En definitiva, podemos decir que la preocupación fundamental del etnógrafo es el estudio de la cultura de una unidad social en sí misma, estudio a través del cual se pretende recoger y entender de la forma más fiel posible las percepciones, acciones y normas de juicio de esa unidad social.

Cuando esta forma de investigar se aplica a los fenómenos educativos es cuando comienza a denominarse etnografía educativa, cuyo objeto es definido por Goetz y Lecompte como «el objeto de la etnografía educativa es aportar valiosos datos descriptivos de los contextos, actividades y creencias de los participantes en los escenarios educativos» (1988, p. 41). Para Tójar (2001), la etnografía es la metodología antropológica que más ha influido en la investigación educativa.

Según Woods (1987), algunos campos de interés para la etnografía en el ámbito de la educación son los siguientes:

- Los efectos que tienen sobre individuos y grupos las estructuras organizativas y los cambios que en ellas se producen.
- Las relaciones sociales y la formación de los educandos y educadores.
- Las culturas de grupos particulares, tales como el profesorado de un centro, grupos determinados de alumnado, etc.
- Lo que las personas hacen, las estrategias que emplean y los significados que ocultan detrás de ellas.
- Las actitudes, opiniones y creencias de los docentes acerca de la enseñanza y alumnado, y del alumnado acerca del profesorado, la escuela, la enseñanza, el futuro de los estudios.
- La gestión de los centros, la toma de decisiones, la comunicación, la disposición de los medios y de los recursos, etc.

Algunas características del método etnográfico son las siguientes:

- *Holista, naturalista, descriptiva, inductiva y emergente.* Los fenómenos son descritos de manera global en sus contextos naturales (observando, escuchando y hablando, anotando anécdotas, haciendo entrevistas...) y a partir de lo observado. Muchos de los instrumentos se generan *in situ*, como resultado de la propia observación.
- *Carácter fenomenológico.* Se estudian los significados de los hechos desde el punto de vista de los agentes implicados en ellos.
- *Intensa implicación.* El investigador se implica intensamente en el entorno social que estudia como observador y participante.

- Tienen importancia los resultados, pero también los procesos.
- *Libre de juicios de valor.* Se debe evitar emitir juicios de valor acerca de las observaciones realizadas, para lo que el etnógrafo debe hacer una ruptura epistemológica con sus visiones, prejuicios y proyecciones.
- *Datos contextualizados.* Los datos se sitúan dentro de una perspectiva más amplia y pueden recogerse a través de diversos medios o recursos como la observación participante, entrevistas, documentos personales, historias de vida, técnicas audiovisuales...
- *Énfasis en la naturaleza de un fenómeno social.* Antes que la comprobación de hipótesis acerca del fenómeno, lo que se pretende es explorar su naturaleza.

Vamos a exponer un esquema basado en el planteado por García Jiménez (1994b), haciendo hincapié en aquellos aspectos o actividades que son más específicos del método etnográfico. Hay que volver a repetir que el hecho de plantear diferentes fases no quiere decir que lo debamos entender como una sucesión lineal de etapas, sino como un proceso de carácter cíclico, flexible y abierto en el que se puede dar solapamiento e influencia mutua entre las distintas fases. Por lo tanto, las fases del método etnográfico quedarían de la siguiente forma:

- *Planteamiento de la cuestión o problema de investigación.* Estos problemas suelen tener su origen en experiencias, compromisos personales e ideológicos del propio investigador o evaluador o en los fenómenos cotidianos que provocan su curiosidad.
- *Planificación de la investigación o evaluación.* Es lo que algunos llaman diseño de la investigación o evaluación y que es rechazado por otros. En definitiva, es la definición de un plan de acción encaminado a lograr las metas fijadas y que también estará determinado por el conocimiento y experiencia previa en este tipo de investigaciones.
- *Acceso al escenario y definición de roles.* El acceso al escenario o campo debe entenderse como un proceso por el que el etnógrafo va accediendo a la información fundamental para su estudio. Como dicen Taylor y Bogdan, «el escenario ideal para la investigación es aquel en el cual el observador obtiene fácil acceso, establece una buena relación inmediata con los informantes y recoge datos directamente relacionados con los intereses investigativos» (1987, p. 36). Para ello, deberá elegir el lugar y negociar los permisos necesarios. Ligado con el escenario se encuentra la cuestión de los roles de las personas participantes, tanto informantes como investigador o evaluador. La definición de los roles tiene relación con el grado de implicación en el escenario y proceso de la investigación o evaluación y tiene una naturaleza interactiva ya que a lo largo de la misma se pueden redefinir.

- *Elección de los informantes.* Es el propio investigador o evaluador quien fija los criterios de elección. Nos encontramos, pues, ante una elección no probabilística sino intencional, que, además es algo dinámico, pues puede cambiar a lo largo del proceso de investigación o evaluación. Entre las distintas estrategias de elección de sujetos tenemos las siguientes: selección exhaustiva, por cuotas, casos únicos, casos extremos...
- *Elección de las técnicas de recogida y análisis de datos.* Los procedimientos utilizados en la investigación o evaluación etnográfica para la recogida de datos son diferentes sistemas de observación (observación participante, diarios, grabaciones en vídeo...), de entrevista cualitativa y documentos de distinto tipo. En cuanto al análisis de datos, hay diferentes técnicas como análisis de dominio, de componentes, de temas, etc. Hay quien lo engloba dentro del denominado análisis de contenido. Lo que habría que destacar es la prácticamente total ausencia de análisis estadísticos.
- *Recogida de datos, análisis e interpretación.* Es lo que podríamos denominar el trabajo de campo. Hay que destacar que estas tres fases de recogida, análisis e interpretación se van simultaneando a lo largo de la investigación o evaluación en un proceso dialéctico en el que el acceso a una de las fases no implica la finalización de la anterior. El final de este proceso se da cuando el informe está concluido.
- *Redacción del informe.* El informe final, que también se suele conocer con el nombre de etnografía, es el documento por medio del cual se comunican los descubrimientos, razonamientos y forma de llegar a ellos. Desde este punto de vista se puede pensar que el informe etnográfico es una traducción de los significados de los participantes, que deberán coincidir con las interpretaciones del investigador o evaluador.

Spindler y Spindler (1992) plantean una larga lista de requisitos que debe cumplir una investigación etnográfica para considerar que está bien hecha. Vamos a destacar tres de ellos. Uno de estos requisitos es el de que la observación sea directa, que se dé donde la acción tiene lugar y tratando de modificar lo menos posible dicha acción. Otro requisito es el de pasar tiempo suficiente en el escenario para que la observación sea validada. Esto supone permanecer en el escenario hasta el momento en que no se recoja nada nuevo. Y en tercer lugar, mencionar la necesidad de registrar un gran volumen de datos; en principio todos aquellos que puedan estar relacionados con el objeto de estudio y que sean válidos para un mejor conocimiento de dicho objeto.

A los requisitos anteriores habría que añadir la necesidad de validación de los datos, para lo cual se suele recurrir a la triangulación, que supone la utilización de más de una fuente o estrategia de recogida de datos al estudiar las mismas unidades. Por ejemplo, además de la observación pueden utilizarse entrevistas y un cuestionario.

Las diversas modalidades de investigación etnográfica según Colás (1997) son las siguientes:

- *Etnografía antropológica*. Se centra en los aspectos relacionados con la forma en que las personas actúan y viven en sus ambientes, así como las creencias que tienen hacia el mundo.
- *Etnografía clásica*. El objetivo es la realización de descripciones amplias y significativas de las elaboraciones culturales que realizan las personas desde sus posiciones.
- *Etnografía sistemática*. Se interesa por definir la estructura de la propia cultura, quedando en segundo plano tanto las personas como su interacción social.
- *Etnografía interpretativa*. Esta modalidad, más que descripciones detalladas, pretende sacar al exterior los significados implícitos de las conductas humanas.
- *Etnografía crítica*. Estaríamos en el caso de la etnografía desde la perspectiva crítica. Dentro de esta corriente, la autora sitúa el posmodernismo, que se centra fundamentalmente en la retórica y en la forma de descripción, y el feminismo, que se orienta a poner de manifiesto las fuerzas de opresión que se ejercen sobre los grupos menos privilegiados socialmente.
- *Etnografía de la comunicación*. Se desarrolla desde la sociolingüística la antropología y la sociología. Su propósito es identificar los principios fundamentales de la organización social e identificar los patrones culturales de las organizaciones de acuerdo a cómo la gente interacciona.

5. Estudio de casos

El estudio de casos al que nos referimos en este apartado es el que tiene lugar dentro de la perspectiva cualitativa, ya que hay autores que también lo presentan desde una perspectiva cuantitativa, como una modalidad de investigación descriptiva e incluso de la experimental. En otras ocasiones, el estudio de casos es tratado como una variante dentro del método etnográfico considerándolo como una aplicación de éste a una unidad de análisis denominada «caso» para conocerla con mayor profundidad. Incluso en algunos ejemplos también es considerado como un enfoque dentro de la investigación-acción.

Para Latorre, Del Rincón y Arnal (1996), el estudio de casos constituye un método de investigación para el análisis de la realidad social de gran importancia en el desarrollo de las Ciencias Sociales y Humanas, y representa la forma más pertinente y natural de las investigaciones orientadas desde una perspectiva cualitativa.

Vamos a considerar el estudio de casos como diferenciado del método etnográfico, porque, aunque en ocasiones pueden coincidir metodológicamente, el objeto de estudio suele ser diferente. Según Yin (1994), el estudio de casos consiste en la descripción y el análisis detallados de unidades sociales (y no tanto de comunidades como el aula) o entidades educativas únicas. Para Tójar (2001), se trata de una investigación descriptiva y en profundidad de un caso en la que se trata de descubrir e identificar problemas y los orígenes de los mismos.

En definitiva, el objetivo del estudio de casos, que también es conocido como «estudio de caso único», es lograr un conocimiento profundo de realidades concretas tales como un individuo, una familia, un grupo, una comunidad, una situación o un suceso. Normalmente, es una investigación a la que se dedica poco tiempo, pero con una alta intensidad (en general, la etnografía suele llevar más tiempo). En este sentido, la diferencia con la investigación etnográfica reside en que ésta pretende fundamentalmente las relaciones entre los comportamientos y los contextos en que se dan. De todas formas debe quedar claro que en algunos casos pueden coincidir.

En cuanto a las características del estudio de casos cualitativo, habría que decir que sustancialmente no son muy diferentes de las del método etnográfico. Como ejemplo podemos mencionar las que Merriam (1990) considera como esenciales y que de alguna manera pueden reflejar las particularidades de este método. Son las siguientes:

- *Particularista*: los estudios de casos se centran en una situación, evento, programa o fenómeno particular que son afrontados de una manera holística.
- *Descriptivo*: el producto final de un estudio de casos es una descripción rica y completa del fenómeno estudiado.
- *Heurístico*: el estudio de casos ilumina al lector en cuanto a la comprensión del caso, puede dar pie a descubrir nuevos significados o confirmar conocimientos ya sabidos.
- *Inductivo*: la mayoría de los estudios de casos se basan en el razonamiento inductivo para llegar a generalizaciones, conceptos o hipótesis.

Los objetivos de los estudios de casos pueden ser muy variados. Ary *et al.* (1996) señalan como posibles objetivos los siguientes:

- Describir, diagnosticar y analizar situaciones específicas.
- Generar hipótesis.
- Adquirir nuevos conocimientos.
- Completar información ofrecida por investigaciones cuantitativas.

El proceso de investigación del estudio de casos sigue la lógica de las etapas de la investigación cualitativa. En este sentido se podrían distinguir cuatro grandes etapas o fases:

- Fase de exploración y de reconocimiento de las situaciones, lugares y sujetos como posibles fuentes de datos.
- Fase de selección de escenarios, aspectos, técnicas de recogida de datos, sujetos...
- Recogida, análisis e interpretación de los datos. Son la observación y la entrevista las utilizadas con mayor frecuencia, aunque en principio puede utilizarse cualquier técnica que el propio estudio precise.
- Elaboración del informe.

La característica más notable en este proceso de investigación es que las primeras etapas son muy abiertas y en la medida en que el trabajo va avanzando, el investigador va focalizando su atención en los aspectos o datos que más interesantes le resultan.

Los tipos de estudios de casos son múltiples así como sus clasificaciones. Una de ellas es la presentada por Arnal, Del Rincón y Latorre (1992, pp. 208-209) que los clasifican en:

- *Estudio de casos institucional*. Se centra en el estudio de una organización particular (por ejemplo, una escuela) durante un tiempo, describiendo y analizando su desarrollo. A esta modalidad de estudio de caso, Taylor y Bogdan (1987) la denominan como estudio de casos organizativo-histórico, con lo que podría entenderse también como una modalidad de investigación histórica.
- *Estudio de casos observacional*. El foco de estudio es una organización o aspecto concreto o actividad de ella (por ejemplo, el diseño curricular de un centro), o incluso una combinación de estos aspectos. Lo que les caracteriza es que la técnica básica de recogida de datos es la observación participante.
- *Historias de vida*. En este caso, la técnica básica de recogida de información es la entrevista. Son entrevistas intensivas dirigidas a personas que pueden servir como vehículos para entender aspectos básicos de la conducta humana o de instituciones actuales. Pujadas (1992) considera las historias de vida como una variante del método biográfico, y en este sentido, según él, se refieren al estudio de caso de una persona mediante el relato que hace esa persona de su propia existencia, relato obtenido a través de entrevistas sucesivas.
- *Análisis situacional o microetnografía*. Estos estudios se centran en contextos muy particulares como un evento particular (por ejemplo, expulsión de un alumno del centro escolar) que es estudiado desde distintos puntos de vista.

- *Estudios de casos múltiples*. Cuando se estudian dos o más sujetos, situaciones o conjuntos de datos.

Pero si atendemos a la naturaleza del informe final nos encontramos con la clasificación siguiente (Merriam, 1990, p. 27):

- Estudio de casos descriptivo: se presenta una descripción del fenómeno estudiado sin marco teórico previo.
- Estudio de casos interpretativo: se interpreta y teoriza acerca del fenómeno estudiado.
- Estudio de casos evaluativo: implican descripción, explicación y juicio.

Por último cabría destacar algunas de las posibles ventajas e inconvenientes del estudio de casos que también pueden aplicarse al método etnográfico.

La ventaja principal del estudio de casos radica en que al sumergirse en la dinámica de una única entidad, situación o evento, se pueden descubrir hechos o procesos que pasarían por alto si se utilizaran otros medios más superficiales.

Otra ventaja es la de que el estudio de casos acentúa el proceso, el cambio experimentado con el paso del tiempo ya que se revelan datos y situaciones sobre el momento en que comienza el estudio, sobre la manera de realizarse y sobre su conclusión. Por último decir que tanto el estudio de casos como la etnografía son valiosos para informar sobre la realidad educativa, para mostrar sus realizaciones, dilemas y contradicciones.

En cuanto a los inconvenientes, además de los inherentes a la investigación cualitativa habría que mencionar que uno de los mayores inconvenientes procede de la presencia prolongada del investigador o evaluador en el escenario, lo que puede producir una distorsión de la realidad objeto de estudio y un quedar atrapado por las situaciones o cuestiones que se están estudiando como consecuencia de su implicación en el escenario. A esto se podría añadir la posibilidad de que surjan conflictos interpersonales entre el investigador o evaluador y los sujetos sobre quienes se investiga.

6. Investigación-acción

La investigación-acción es un fenómeno complejo cuya naturaleza ha variado desde sus orígenes hasta nuestros días. Siguiendo a López Górriz (1998), se pueden distinguir cuatro etapas en la trayectoria histórica de la investigación-acción:

- *Raíces de la investigación-acción*. Se sitúan en el siglo XIX y principios del XX. En este periodo, se dieron una serie de hechos que fueron preparando el terreno para el surgimiento del método de la inves-

tigación-acción. Entre estos hechos podemos destacar los siguientes: cuestionario pasado por Karl Marx a los obreros de manufacturas para provocarles la reflexión sobre sus condiciones de vida; la búsqueda de encuadres diferentes del de las Ciencias Naturales para comprender e interpretar hermenéuticamente los hechos sociales por parte de autores como Dilthey, Simmel o Weber, o las investigaciones realizadas por la Escuela de Chicago acerca de los problemas sociales urbanos de la primera mitad de siglo.

- *Nacimiento de la investigación-acción.* La mayor parte de los autores reconocen a Kurt Lewin como el creador de esta línea de investigación. La idea básica de Lewin es la de poner la investigación al servicio de la acción, para lo que busca los medios de introducir la investigación en los procesos de la acción e intentar avanzar en el sentido de una ciencia práctica. Este periodo se sitúa entre la década de 1920 y finales de la de 1940.
- *Evolución y eclipsamiento de la investigación-acción.* Este periodo llega hasta la década de 1970. Dentro de él se distingue una fase de importante desarrollo de la investigación-acción hasta finales de la década de 1950 para continuar con otra de decaimiento y eclipsamiento. Según Barbier (1996) son cuatro las modalidades de investigación-acción que se originan en la fase de desarrollo: la investigación-acción diagnóstica, la investigación-acción participante, la investigación-acción empírica y la investigación-acción experimental. Es en este periodo cuando se sitúan las primeras aplicaciones de la investigación-acción a la Educación.
- *Resurgimiento e impulso de la investigación-acción.* Comprende el periodo desde la década de 1970 hasta el momento actual. Sus características fundamentales son: la reformulación del término «investigación-acción» y el desarrollo de una amplia y variada gama de enfoques que van desde los más experimentalistas a los más radicales y revolucionarios. Los autores más destacados de esta época son Moser en Alemania; Stenhouse y Elliott en Inglaterra; Carr y Kemmis en Australia; Barbier, Ardoino y Lapassade en Francia; Goyette y Lessard-Hébert en Canadá; etc.

Debido al carácter especial de la denominada investigación-acción, antes de describir el proceso de investigación expondremos algunas de sus características.

6.1. Naturaleza de la investigación-acción

Existen muchas definiciones o formas de conceptualizar la investigación-acción. López Górriz (1998) ha hecho un análisis de distintas definiciones y en-

foques de investigación-acción. A partir de dicho análisis, esta autora presenta una larga lista de las características que están detrás del concepto de investigación-acción. Algunas de éstas son las siguientes:

- Es un fenómeno complejo que, en la medida que se va poniendo en funcionamiento, se va haciendo más complejo.
- Es un enfoque contradictorio en el que se persigue la producción de «conocimiento científico», a la vez que la transformación de las prácticas sociales.
- Exige participación y colaboración, así como el establecimiento de relaciones horizontales, dialógicas, de escucha, de solidaridad, de compromiso con sus propias prácticas y con sus situaciones sociales.
- Se desarrolla siguiendo el proceso de una espiral, que comprende ciclos constantes de planificación, acción, observación, reflexión.
- El grupo es el instrumento básico de este proceso.
- Exige un trabajo sistemático de recogida de información, reflexión y análisis constantes.
- Desde el punto de vista metodológico se concibe de un modo amplio y flexible, con un cierto rigor metodológico, pero que es distinto al concepto de control y rigor de la investigación básica.
- Exige, a priori, un contrato que se va renegociando en la medida en que avanza el proceso.
- El investigador actúa desde la implicación y debe dominar diferentes tipos de saberes científicos.
- Es una investigación que se hace sobre una praxis social concreta y cuyo proceso es complejo y arduo, ya que exige un compromiso muy fuerte en las personas implicadas.

Goyette y Lessard-Hérbert (1988) consideran que la investigación-acción desarrolla tres dimensiones: la acción, la investigación y la formación. Estos autores afirman que este tipo de investigación beneficia tanto el desarrollo de destrezas como la expansión de teoría y la resolución de problemas.

Autores como Elliott (1990) y Stenhouse (1987) en sus definiciones aluden a la acción. La dimensión de la acción como praxis es un punto fundamental alrededor del cual gira la realización de una investigación-acción y ello exige la intervención e implicación de los actores sociales. Como afirma Kemmis (1988, p. 42), «la investigación-acción es una forma de investigación llevada a cabo por parte de los prácticos sobre sus propias prácticas». Esto llevado al ámbito de la Educación se transforma en la investigación por parte de los agentes educativos acerca de las acciones educativas.

En cuanto a la segunda dimensión, la investigación, habría que decir que la investigación-acción es entendida como un método de investigación riguroso, sistemático y fiable pero que necesita de la utilización de un método amplio y flexible que pueda recoger su complejidad.

Por último, en la dimensión de formación la investigación-acción supone, entre otros, un cambio de actitudes, capacidades y valores; un clima de comunicación, diálogo y reflexión; desarrollo del espíritu crítico y el trabajo en equipo; creación de espíritu investigador y desarrollo de la capacidad de escucha (López Górriz, 1998).

Estas tres dimensiones, *acción, investigación y formación* junto con la característica de su *perspectiva comunitaria*, en el sentido de que es necesaria la implicación grupal (Rodríguez, Gil y García, 1996), constituirían las cuatro características básicas de la investigación-acción independientemente de las diversas modalidades en que puede presentarse.

6.2. Enfoques de la investigación-acción

En lo referente a los enfoques o modelos de investigación-acción, el panorama también es muy complejo. Las clasificaciones son muy variadas en función de los criterios empleados.

Colás y Buendía (1998), basándose en el criterio de «formas de acción», distinguen entre investigación-acción técnica (sobre la acción), práctica (para la acción) y crítica (por la acción).

López Górriz presenta una amplia descripción de los enfoques fundamentales de investigación-acción y que presentamos de una forma resumida a través del cuadro 6.1, elaborado por la misma autora.

Centrándonos más en la investigación-acción relacionada con el ámbito de la educación, es de destacar la clasificación que realiza Reason (1994) entre investigación-acción del profesor, investigación cooperativa e investigación participativa. Para este autor, la diferencia fundamental entre estas tres modalidades de investigación-acción radica en que la participativa sirve a la comunidad; la cooperativa, al grupo, y la investigación-acción del profesor, al individuo.

Brevemente profundizaremos en las diferencias entre las tres modalidades exponiendo lo esencial de cada una de ellas. El propósito de la **investigación-acción del profesor** es que el docente profundice en la comprensión de su situación problemática a través de la interpretación de lo que ocurre desde el punto de vista de quienes actúan e interactúan en la situación problemática, como, por ejemplo, profesorado-alumnado o profesorado-dirección. Se le suele llamar también «investigación en el aula».

Bartolomé (1994, p. 379) define la **investigación cooperativa** como *metodología de investigación orientada al cambio social y educativo. [...] una modalidad de la investigación-acción*. La razón que justifica esta última afirmación es el hecho de compartir los siguientes rasgos comunes entre investigación cooperativa e investigación-acción:

- a) El objeto de la investigación acción lo constituye la práctica educativa y/o social, al tiempo que se procura una mayor comprensión de ella.

Cuadro 6.1. Enfoques fundamentales de investigación-acción.
(Tomado de López Górriz, 1998, p. 241)

Enfoques	Campos	Autores
<p>Enfoques de investigación en el aula y desarrollo profesional:</p> <ul style="list-style-type: none"> • Enfoque anglófono. • Enfoque australiano. 	<p>Enseñanza: aula, desarrollo profesional del centro y de comunidades educativas.</p>	<ul style="list-style-type: none"> • Stenhouse, Elliott • Carr y Kemmis
<p>Enfoque de formación y desarrollo comunitario:</p> <ul style="list-style-type: none"> • Investigación-acción participativa. 	<p>Formación formal y no formal. Adultos, animación sociocultural, desarrollo organizacional...</p>	<ul style="list-style-type: none"> • Fals Borda, Demo, Hall, Bartolomé, Ander-Egg
<p>Enfoques de intervención institucional:</p> <ul style="list-style-type: none"> • Análisis institucional externo. • Análisis institucional interno. • Investigación-acción institucional. 	<p>Formación y cambio institucional y profesional. Educación, empresa, organizaciones políticas...</p>	<ul style="list-style-type: none"> • Lapassade, Lourau, Hess • Boumard, Hess • Barbier
<p>Enfoque de investigación inter-institucional:</p> <ul style="list-style-type: none"> • Investigación cooperativa. 	<p>Formación y desarrollo profesional e institucional. Innovación educativa.</p>	<ul style="list-style-type: none"> • Bartolomé y Anguera, Santiago, Oja y Pine
<p>Enfoques de investigación y formación clínicos:</p> <ul style="list-style-type: none"> • El método de centración en el deseo. • Investigación-acción existencial. 	<p>Formación y desarrollo personal social y profesional.</p>	<ul style="list-style-type: none"> • Lobrot • Barbier

- b) Existe una articulación permanente entre la investigación, la acción y la formación a lo largo del proceso.
- c) Una manera particular de acercarse a la realidad, que persigue vincular conocimiento y transformación.
- d) Protagonismo de los educadores-investigadores.
- e) Interpelación del grupo.

Lo esencial de la investigación cooperativa es que se da cuando se agrupan personas de dos o más instituciones (por ejemplo, de una universidad y de una escuela) para resolver juntos problemas que atañen a la práctica profe-

sional de los miembros de una de ellas (generalmente, la escuela o centro escolar) y que suelen estar vinculados a los procesos de innovación o al desarrollo y formación profesional (Bartolomé, 1994). En este caso, el profesorado desarrolla competencias, habilidades y conocimientos de investigación, en tanto que los investigadores se reeducan a sí mismos con la utilización de metodologías naturalísticas y estudios de campo. Algunos autores la conocen también como investigación colaborativa.

La **investigación participativa** busca que a través de un conjunto de principios, normas y procedimientos metodológicos se logre una adquisición colectiva del conocimiento, su sistematización y su utilidad social (Rodríguez, Gil y García, 1996). El objetivo último es la mejora de los sujetos implicados. Es una forma de unir la teoría y la práctica a través de la combinación entre la participación y la investigación acentuándose el compromiso social desde una posición crítica.

6.3. El proceso de la investigación-acción

Indudablemente es difícil hacer un planteamiento general del proceso de investigación en la investigación-acción que sea válido para todos y cada uno de los enfoques presentados en el cuadro 6.1. Aunque cualquiera de los enfoques anteriores guarde unos pasos fundamentales como pueden ser: planteamiento del problema, planificación de la acción, desarrollo y seguimiento de la acción, evaluación y reformulación de la problemática a investigar, los matices pueden ser múltiples por lo que es muy difícil plantear un enfoque general con total garantía de que recoja todas las posibilidades.

Por otra parte presentar todas las variantes en cuanto a los procesos de investigación en la investigación-acción excede los objetivos de este libro. Por ello, hemos tomado la decisión de presentar uno de los modelos, el que López Górriz identifica como enfoque australiano dentro de los «Enfoques de investigación en el aula y desarrollo profesional».

Las razones de presentar este modelo no son otras que las de ser un modelo relativamente sencillo muy conocido en nuestro entorno y que además ha sido utilizado en el terreno de la Educación. Además este modelo refleja las cuatro fases que identificó Lewin, y que luego han ido configurándose a partir de numerosas aportaciones desde diferentes contextos geográficos e ideológicos.

Los autores y obras más representativos de este enfoque y en los cuales se basa el proceso de investigación que vamos a presentar a continuación son Carr y Kemmis (1988) y Kemmis y McTaggart (1988). La propuesta de estos autores es la que se conoce como la espiral autorreflexiva y está formada por ciclos sucesivos de planificación, acción, observación y reflexión.

Según los anteriores autores, las cuatro fases del proceso de investigación-acción son las siguientes:

1. *Planificación* o desarrollo de un plan de acción. Una vez que el área problemática y las cuestiones específicas han sido identificadas y consecuentemente el problema de investigación ha sido formulado de manera precisa y operativa, se estudian y establecen prioridades en las necesidades para optar entre diversas alternativas. Este plan de acción irá encaminado a la mejora de la situación actual por lo que deberá considerar los riesgos y las limitaciones a partir de un diagnóstico previo. Es por esto por lo que en algunas ocasiones el «diagnóstico y reconocimiento de la situación inicial» es considerado como una fase del proceso de investigación, previa a la planificación.
2. *Acción*. Es una actuación para poner en práctica el plan diseñado. Esta acción es flexible, permeable a los imprevistos y consecuente con las decisiones colectivas, por lo que en muchas ocasiones requiere de la negociación y el compromiso.
3. *Recogida de información*. Las técnicas más utilizadas son la entrevista en profundidad, la observación cualitativa, el diario de campo, los autoinformes, los registros anecdóticos, etc. La información recogida será acerca de los efectos de la acción en el contexto en que tiene lugar. Es por ello por lo que además de recoger datos acerca del proceso y los efectos de la acción deberán recogerse también acerca de las circunstancias en que se desarrolla.
4. *Reflexión*. La reflexión será acerca de los efectos de las acciones emprendidas a fin de tener base para una nueva planificación de nuevas acciones. La base de la reflexión está en la discusión entre participantes a través del intercambio de puntos de vista y la reconstrucción de significados.

Una característica de esta modalidad metodológica es que sus evidencias están apoyadas en la observación abierta, el diálogo permanente entre participantes, la revisión sistemática, el análisis y control de resultados, así como la estrecha colaboración entre los expertos y participantes. Además de éstas, las decisiones de cambio y reajuste surgen en el propio proceso de investigación y son tomadas de común acuerdo entre los integrantes del grupo de investigación.

Por último, habría que mencionar las áreas o campos de aplicación de la investigación-acción. Colás y Buendía (1998) mencionan los siguientes campos de aplicación de la investigación-acción en Educación:

- Movimientos comunitarios.
- Formación ocupacional.
- Formación del profesorado.
- Desarrollo curricular e innovación educativa.
- Organización escolar.
- Evaluación de centros, aprendizaje, instituciones, etc.
- Planificación y toma de decisiones.
- Educación de adultos.

Algunas cuestiones que pueden ser tratadas y resueltas con esta metodología son las siguientes:

- Resolución de problemas diagnosticados en una situación específica: disciplina, coordinación de la enseñanza, etc.
- Introducción de enfoques innovadores en la enseñanza y el aprendizaje dentro de un sistema de constantes innovaciones y cambio.
- Desarrollo de la comunicación entre docentes y los investigadores y remediar el fracaso de la investigación tradicional para ofrecer prescripciones claras.
- Proveer una alternativa a la solución de problemas en clase, tratando aspectos que no pueden abordarse desde el rigor de la investigación científica.

Preguntas y ejercicios

1. Elabora un esquema con los diseños de investigación que se ofrecen en el libro.
2. Plantea un problema de investigación o evaluación adecuado para llevarse a cabo siguiendo cada uno de los 6 tipos de diseños mostrados en el capítulo. Señala en cada caso algunos objetivos posibles.
3. De los 6 tipos de diseños mostrados en el capítulo, ¿cuál te parece que es el más adecuado para estudiar la presumible falta de educación en valores que se da hoy en día en el alumnado? Razona la respuesta.
4. ¿Y si lo que quisiéramos fuera conocer los valores mayoritariamente asumidos por la población en edad escolar? Argumenta la respuesta.
5. Si quisiéramos intervenir para mejorar la formación en valores del alumnado, ¿cuál de los diseños presentados en capítulo te parecería el más adecuado? Presenta las razones de tu respuesta.

Nota

1. En este capítulo se utilizan indistintamente los términos «método» y «diseño» como sinónimos.

7. Procedimientos para la recogida de la información

En este capítulo vamos a tratar el tema relativo a las técnicas, las estrategias y los procedimientos para la recogida de evidencias o información sobre el objeto que está siendo evaluado. Al final del capítulo haremos un breve repaso a los recursos que nos ofrece la informática para la recogida de información en el proceso evaluativo.

En todas las definiciones de evaluación se expresa claramente la necesidad de recabar información sobre el objeto a evaluar. Aunque el mero hecho de recoger información no es suficiente para completar una evaluación, sí es una condición necesaria sin la cual ésta no se dará. Es por ello por lo que en este apartado vamos a presentar, aunque sea de forma breve y concisa, los procedimientos más usuales utilizados en la evaluación educativa.

Hoy en día se utilizan numerosas técnicas evaluativas en el proceso de evaluación. Unas de carácter más bien cuantitativo (los tests estandarizados, por ejemplo) y otras de carácter más bien cualitativo (la observación participante, por ejemplo). No obstante, no se puede argumentar que una técnica sea superior a otra, sino que se utilizará(n) una(s) u otra(s) en función de su adecuación a la finalidad de la evaluación (criterio de utilidad) y en función de los recursos humanos, materiales y económicos disponibles (criterio de contextualización) (Cabrera, 2000a).

Existen diversas formas de clasificar las distintas técnicas evaluativas atendiendo a diferentes criterios. Sin embargo, ninguna de estas clasificaciones goza de la aceptación mayoritaria de la comunidad educativa.

En este sentido, algunos autores distinguen entre técnicas cuantitativas (pruebas objetivas, tests...) y técnicas cualitativas (observación participante, entrevista en profundidad...). Sin embargo, en muchas ocasiones, una misma técnica (el cuestionario, por ejemplo) puede ser cuantitativa o cualitativa, dependiendo de su construcción, aplicación y análisis.

Otros, como Del Rincón (1996), distinguen entre estrategias (entrevista y observación participante, por ejemplo), instrumentos (tests y pruebas objetivas, por ejemplo) y medios audiovisuales y documentales (el vídeo y la fotografía, por ejemplo).

González Such (2000) diferencia entre instrumentos estandarizados, entre los que sitúa los tests, y los instrumentos no estandarizados, entre los que señala la entrevista, el cuestionario, la observación, etc.

Más recientemente, Padilla Carmona (2002) clasifica las técnicas evaluativas en:

- a) Procedimientos observacionales (observación).
- b) Instrumentos de encuesta (entrevista y cuestionario).
- c) Tests y pruebas objetivas.
- d) Otras técnicas y estrategias (portafolios, historias de vida...).

Independientemente de cómo sea clasificada una técnica (en función de los distintos criterios usualmente empleados) debe sin ningún lugar a dudas aplicarse teniendo en cuenta los criterios antes mencionados de utilidad y contextualización.

Por otra parte, no debemos olvidar que cualquier técnica, instrumento o estrategia utilizado para la obtención de información debe cumplir dos objetivos básicos: conseguir las máximas cotas de **fiabilidad** y **validez**. La fiabilidad hace referencia a la ausencia de errores de medida y a la consistencia de la información obtenida. La fiabilidad es una condición necesaria, pero no suficiente, para asegurar la validez puesto que ésta hace referencia al grado en que las mediciones obtenidas realmente describen lo que pretendían describir.

Las técnicas utilizadas en la evaluación educativa son innumerables y cada día se están experimentando nuevos procedimientos, aunque éstos no lleguen a sustituir a los anteriores. De todas maneras, hay una serie de técnicas que consideramos que son las más utilizadas y son las que vamos a presentar a continuación.

1. La observación

1.1. Concepto

El método científico se basa en la observación, entendida ésta como cualquier procedimiento de recogida de información de la realidad que se preten-

de describir y analizar. Sin embargo, la observación también puede ser entendida como una técnica específica con la que se pretende recoger información directamente de la realidad. De tal forma que posibilita obtener datos de un fenómeno tal y como se produce.

Cabrera (2000a) señala que los procedimientos de observación abarcan distintas técnicas cuyo común denominador es «describir u obtener información mediante el registro del comportamiento que manifiesta más o menos espontáneamente una persona» (p. 185). La conducta de una persona y sus interacciones con el medio son observados directa e intencionalmente.

Padilla Carmona (2002, p. 71) nos presenta distintas definiciones dadas por diversos autores del concepto de observación.

Anguera (1982, p. 20) define observar como «advertir los hechos como espontáneamente se presentan y consignarlos por escrito». Posteriormente, esta misma autora la definió como «un procedimiento encaminado a articular una percepción deliberada de la realidad manifiesta con su adecuada interpretación, sistemático y específico de la conducta generada de forma espontánea en un determinado contexto, y una vez se ha sometido a una adecuada codificación y análisis nos proporciona resultados válidos dentro de un marco específico de conocimiento» (1988, p. 7).

Por último, Sierra Bravo (1983, p. 213) la define como «la inspección y estudio realizado por el investigador, mediante el empleo de sus propios sentidos, especialmente la vista, con y sin ayuda de aparatos técnicos, de las cosas y hechos de interés social, tal como son o tienen lugar espontáneamente en el tiempo en que acaecen y con arreglo a las exigencias de la investigación científica».

Tras el análisis de estas definiciones, Padilla Carmona (2002) señala que las características más sobresalientes de la observación son:

- Es un procedimiento de recogida de información que se basa en la percepción sensorial del investigador o evaluador.
- Estudia los fenómenos que existen natural y espontáneamente.
- Es un examen de fenómenos actuales tal y como son o tienen lugar en la realidad del momento presente.

1.2. Modalidades de observación

Para hablar de las distintas modalidades de la observación, debe resaltarse una distinción previa entre su observación y su registro. La observación hace referencia a la percepción e interpretación de la realidad por parte de la persona que la realiza, mientras que plasmar de forma escrita esa percepción e interpretación haría referencia al registro. Por ello, suelen distinguirse modalidades de observación y tipos de registros.

Las modalidades de observación pueden clasificarse atendiendo a distintos criterios:

1. Teniendo en cuenta el nivel de sistematización del proceso, se distinguen tres modalidades:
 - a) *Observación no sistemática* u ocasional, en donde el registro de los datos es muy poco estructurado y se realiza de manera ocasional.
 - b) *Observación sistematizada* o controlada, en donde las conductas objeto de observación están bien definidas y especificadas.
 - c) *Observación muy sistematizada*, en donde existe un sistema de categorías de observación muy estructurado.
2. Teniendo en cuenta la fuente de datos, se distinguen:
 - a) *Observación directa*, en donde se observan y registran los comportamientos y los fenómenos en el momento en que suceden.
 - b) *Observación indirecta*, en donde la información sobre los fenómenos y las conductas se obtiene a través de lo que el sujeto dice de sí mismo o de lo que otros informan sobre él.
3. Teniendo en cuenta el rol que adopta el observador, se distinguen:
 - a) *Observación participante*, en donde el observador participa y es una persona más en la dinámica de la situación observada.
 - b) *Observación no participante*, en donde el observador queda al margen del campo de observación.
4. Teniendo en cuenta el agente evaluador, se distinguen:
 - a) *Heteroobservación*, en donde el observador contempla las conductas de otras personas.
 - b) *Autoobservación*, en donde el observador contempla su propia conducta.

De todas formas, estas dicotomías no deben entenderse como absolutas, sino que representan los dos extremos de un continuo en el que tienen cabida distintas posturas intermedias.

Los tipos de registros para recoger los datos observacionales suelen clasificarse en narrativos y categoriales.

En los **registros de tipo narrativo**, el observador trabaja como un reportero reproduciendo los sucesos o comportamientos y dando cuenta de lo ocurrido. El objetivo de éstos sobrepasa la mera comprensión intentando identificar e interpretar la aparición de semejantes sucesos o comportamientos. El observa-

dor es el instrumento principal, y su observación depende de su capacidad perceptiva de los hechos y de su capacidad de transmitirlos. Los registros narrativos más usuales son los siguientes (Cabrera, 2000a):

- *Las notas de campo* utilizadas en la observación participante en donde se recogen los hechos y su interpretación.
- *Los diarios* son una modalidad de registro muy abierto de tipo longitudinal que contienen observaciones repetidas de una misma persona o grupo.
- *Los registros anecdóticos* son descripciones de algún suceso o conducta puntual que es considerada como importante.
- *La descripción de muestras de conducta* utilizada en la observación no participante supone la transcripción detallada, precisa y completa de un suceso o escena.

En los **sistemas de registros categoriales**, el observador dispone de antemano de un listado o catálogo de comportamientos a observar, limitándose la observación a identificar y a registrarlos. Entre los sistemas de registros categoriales más utilizados en la evaluación educativa se distinguen los siguientes:

- *Las escalas valorativas*, en las que el observador debe indicar la frecuencia o intensidad en la que se presenta(n) algún(os) comportamiento(s) que previamente ya han sido categorizados.
- *La lista de control* es semejante a la escala valorativa, sólo que únicamente debe señalarse la presencia o ausencia de determinada cualidad o característica; es decir, no es necesario señalar la frecuencia o intensidad.

El cuadro 7.1 recoge las distintas modalidades de observación y registro señaladas.

1.3. Planificación de la observación

La planificación de una observación va a depender del tipo de observación adecuado al propósito de la evaluación y en consecuencia de la modalidad de registro seleccionado. Sin embargo, Cabrera (2000a) señala que hay una serie de elementos que en mayor o menor medida aparecen en cualquier proceso observacional:

1. Establecer la finalidad y los objetivos de la observación.
2. Determinar los comportamientos que van a observarse.
3. Seleccionar el sistema de registro de las observaciones.
4. Identificar el objeto de la observación.

Cuadro 7.1. Modalidades de observación y registro utilizadas en evaluación

<i>Observación</i>	En función del nivel de sistematización del proceso.	<ul style="list-style-type: none"> • No sistemática. • Sistematizada. • Muy sistematizada.
	En función de la fuente de datos.	<ul style="list-style-type: none"> • Directa. • Indirecta.
	En función del rol del observador.	<ul style="list-style-type: none"> • Participante. • No participante.
	En función del agente evaluador.	<ul style="list-style-type: none"> • Autoobservación. • Heteroobservación.
<i>Registros</i>	Narrativos.	<ul style="list-style-type: none"> • Notas de campo. • Diarios. • Registros anecdóticos.
	Categoriales.	<ul style="list-style-type: none"> • Escalas valorativas. • Listas de control.

5. Temporalizar la observación.
6. Establecer las condiciones de la observación.
7. Formar al observador.
8. Analizar e interpretar los datos observados.

1.4. Ventajas y limitaciones de la observación

La observación presenta las ventajas siguientes:

- Permite obtener la información tal como sucede.
- Muchos comportamientos y situaciones que pasan desapercibidos para los evaluados, suelen ser percibidos por el observador.
- Puede utilizarse con sujetos cuya capacidad verbal es limitada.
- Requiere una participación pasiva de los sujetos evaluados.
- Proporciona información sobre aspectos de la vida escolar (o de cualquier otro escenario) que no pueden ser obtenidos por otros medios.

Sin embargo, también tiene una serie de limitaciones:

- Al no poder predecir cuándo se va a dar un suceso hace que muchas veces no se esté presente cuando sucede.
- Sucesos habituales pueden resultar difíciles de observar por la intervención de factores ocultos.
- La posibilidad práctica de realizar la observación está limitada a la duración de los sucesos o al contexto en el que tienen lugar.

- Muchas veces no se puede observar una conducta o fenómeno interesante debido a su innaccesibilidad (por ejemplo, si ya ha sucedido).
- La observación por sí misma proporciona únicamente una visión parcial de la conducta. Percibimos la conducta, pero no las razones.
- Aun el observador más experto filtra la realidad en función de su percepción.
- La reactividad ante la observación puede hacer que los escenarios y las conductas se conviertan en artificiales.
- La observación necesita de muchos recursos humanos y de tiempo para la recogida de información.

2. La entrevista

2.1. Concepto

La entrevista es una técnica de recogida de datos cuyo uso en la evaluación educativa está en plena expansión. Es considerada como complemento de la observación; sobre todo, cuando se hacen evaluaciones de corte cualitativo.

Hace ya varias décadas, Bingham y Moore (1924) definieron la entrevista como una «conversación con un propósito» (citado por Padilla Carmona, 2002, p. 125).

En la década de 1960, Matarazzo (1965) definió la entrevista como «un tipo de conversación verbal y no verbal con el propósito de cumplir un objetivo previamente definido».

Tomando como base estas y otras definiciones, podríamos aceptar como definición de entrevista en el campo de la evaluación educativa la siguiente: «una confrontación interpersonal, en la cual el entrevistador formula al entrevistado preguntas con el fin de conseguir respuestas relacionadas con los propósitos de la evaluación».

Las características de la entrevista son las siguientes:

- Constituye una relación entre dos o más personas.
- Es una vía de comunicación simbólica bidireccional; preferentemente, oral.
- Tiene lugar una interacción personal que la diferencia de las demás técnicas de encuesta.
- Tiene lugar una relación interpersonal asimétrica dado que requiere de unos objetivos prefijados, así como de una asignación de roles en donde el entrevistador controla la situación.
- Al darse una relación, es algo más que la aplicación mecánica de una técnica.
- Al ser una técnica, en ella se operativizan contenidos teóricos con la finalidad de estudiar situaciones.
- Es un proceso con diferentes fases.

2.2. Clasificación

Las entrevistas se pueden clasificar atendiendo a diferentes criterios:

1. *Según los objetivos*, hay autores que distinguen las entrevistas de investigación (que persiguen la recogida de información) y las que persiguen la intervención. Otros autores distinguen según este mismo criterio entre entrevistas de investigación, diagnósticas y terapéuticas.
2. *Según el marco teórico del entrevistador* se distingue entre entrevistas fenomenológicas, psicodinámicas y conductuales.
3. *Según el grado de estructuración* suele hablarse de tres tipos de entrevistas. De todas formas, aunque presentaremos estas tres formas, en función de la estructuración, debe entenderse que nos hallamos ante un continuo que va desde la nula estructuración a la máxima estructuración en la que caben diferentes tipos de entrevistas. En cualquier caso, la preparación y el entrenamiento del entrevistador es inversamente proporcional al grado de estructuración de aquélla. Hay que señalar también que esta clasificación es la más utilizada:
 - *Entrevistas estructuradas*, en las que las preguntas, así como su secuencia y en ocasiones incluso las respuestas (entre las que tiene que elegir el entrevistado) están establecidas de antemano.
 - Las *entrevistas semiestructuradas* son las más utilizadas en la evaluación educativa en la actualidad. En éstas, existe un esquema o pauta de entrevista previa con algunas preguntas que se plantearán al entrevistado. Sin embargo, en función de sus respuestas (tanto verbales como no verbales), el entrevistador planteará nuevas preguntas para profundizar en algunos aspectos o para abrir nuevas vías de indagación que resulten de interés a los propósitos de la evaluación.
 - En las *entrevistas no estructuradas* no existe un esquema o patrón previo para seguir, ni siquiera las preguntas están prefijadas, aunque la entrevista estará guiada por «esquemas implícitos» (Padilla Carmona, 2002).

2.3. Procedimiento

Del Rincón (1996) señala tres fases para planificar una entrevista:

FASE 1: OBJETIVOS DE LA ENTREVISTA

La planificación de la entrevista debe ser coherente con sus objetivos, que a su vez deben ajustarse a los objetivos de la evaluación.

FASE 2: MUESTREO DE LAS PERSONAS A ENTREVISTAR

El evaluador deberá seleccionar las personas susceptibles de ser entrevistadas. Esta selección puede realizarse siguiendo un muestreo aleatorio de toda la población de personas participantes en la evaluación; es decir, los sujetos se eligen al azar. También puede realizarse un muestreo significativo (utilizando la terminología de Santos Guerra, 1990) u opinático (con terminología de Del Rincón *et al.*, 1995), en el que la selección de los entrevistados se realiza debido a su condición de informante clave, ya sea por su experiencia con el tema, por la perspectiva que mantienen, por el cargo que ocupan, etc. Obviamente en las evaluaciones cuantitativas se opta generalmente por el muestreo aleatorio, mientras que el muestreo significativo es más utilizado en las evaluaciones cualitativas.

FASE 3: DESARROLLO DE LA ENTREVISTA

En esta fase el evaluador deberá atender a una serie de aspectos para llevar adelante la entrevista:

- *Contenido y naturaleza de las preguntas*; es decir, el entrevistador deberá plantearse las preguntas que hará, cómo las ordenará, qué nivel de detalle pedirá, qué duración tendrá la entrevista y cómo redactará las preguntas.
- *Organización y secuencia de las preguntas*. Toda entrevista debería comenzar con preguntas que contribuyan a alcanzar un clima agradable, cómodo y de confianza, para, de forma gradual, pasar en una segunda fase a preguntas más complejas y de mayor carga emocional. Por último, la tercera fase debería dedicarse a cuestiones más fáciles y relajadas que permitan terminar la entrevista de forma distendida.
- *Relación entrevistador-entrevistado*. El entrevistador debe tratar de conseguir un clima amistoso y desarrollar la empatía con el entrevistado.
- *Formulación de las preguntas*. El entrevistador debe utilizar un lenguaje claro y preciso en la formulación de las preguntas posibilitando así un nivel de motivación satisfactorio. Hay que tener en cuenta que la calidad de la respuesta está condicionada por la calidad de las preguntas. Por ello, a la hora de formular las preguntas deben tenerse en cuenta una serie de aspectos:
 - Formular preguntas realmente abiertas.
 - Formular preguntas con claridad.
 - Formular preguntas únicas.
 - Profundizar en las respuestas.
 - Apoyo y reconocimiento.
 - Neutralidad.
 - Simulación.
 - Control de la entrevista.

- *Registro de la información.* El entrevistador debe recoger tanto las respuestas orales del entrevistado como los demás sucesos de la entrevista de forma completa, no resumida. Este registro puede realizarse tomando notas o utilizando algún sistema tecnológico de grabación (grabadora de audio o grabadora de audio y vídeo).

2.4. Ventajas y limitaciones

Entre las ventajas que ofrece la entrevista podemos reseñar las siguientes:

- La relación interpersonal que implica.
- Posibilita al entrevistador adaptarse a las características del entrevistado debido a la retroalimentación que se da.
- Facilita observar al entrevistado tanto a nivel verbal como gestual.
- Posibilita registrar grandes cantidades de información de carácter subjetivo y de tipo biográfico.
- Permite recoger información de personas inaccesibles por otros medios (analfabetos, por ejemplo).

Algunas de las desventajas o limitaciones de la entrevista serían las siguientes:

- Su larga duración y el elevado coste de tiempo, esfuerzo y dinero que supone.
- La fiabilidad de la información recogida.
- La validez de la información recogida.
- La interferencia de sesgos de variada procedencia.
- La ausencia de anonimato.

3. El cuestionario

3.1. Concepto

El cuestionario es posiblemente la técnica más utilizada para la recogida de información en evaluación tanto educativa como de otros ámbitos. Como señala Hernández López (1996, p. 127) «expresado de forma sencilla, una encuesta consiste en la formulación de preguntas a un número determinado de individuos que representan a un colectivo más amplio en la búsqueda de informaciones que sean generalizables para un colectivo». En la misma línea, Cabrera (2000a, p. 210) señala que un cuestionario «consiste en un conjunto de preguntas que de forma sistemática y ordenada permite recoger información sobre las percepciones, actitudes, opiniones y características de las per-

sonas». Un cuestionario se utiliza cuando es necesario conocer la opinión acerca de un aspecto determinado que queremos evaluar y las respuestas se dan por escrito. Normalmente, además, esta técnica se utiliza cuando la población de personas sobre las que queremos tener opinión es muy amplia. Tiene una finalidad esencialmente exploratoria para contrastar opiniones y puntos de vista cuando el grupo de personas es muy amplio.

3.2. Clasificación

Al tratar de clasificar los cuestionarios nos encontramos que, atendiendo a diferentes criterios, se obtienen clasificaciones distintas. Padilla Carmona (2002, p. 149) citando a Matud (1993) señala cuatro criterios a partir de los cuales clasificar los cuestionarios:

- I. *Según el tipo de ítems*, se distinguen los cuestionarios cerrados, abiertos, mixtos, de elección múltiple, etc.
- II. *Según la escala de respuesta*, se distinguen los cuestionarios que buscan juicios categóricos y los que buscan juicios continuos.
- III. *Según el método de administración*, se distinguen los personales, por teléfono, por correo, por ordenador, etc.
- IV. *Según el tipo de información requerida*, se distinguen los cuestionarios de información general y los específicos.

3.3. Procedimiento

Teniendo en cuenta las propuestas de Cabrera (2000a) y Padilla Carmona (2002) se proponen las etapas siguientes en el proceso de construcción y utilización de un cuestionario en la evaluación:

1. Contextualizar el uso del cuestionario en el proceso de evaluación. Es decir, el cuestionario se empleará en la evaluación junto con otras técnicas y, por ello, deberá definirse con claridad cuál es el propósito de su uso.
2. Plantear y explicitar el marco teórico conceptual. Esta fase supone tres etapas:
 - a) Definir el constructo o constructos sobre los que vamos a recoger información.
 - b) Explicitar y definir las dimensiones de las que se compone ese constructo.
 - c) Enumerar y definir los indicadores que operativizan el significado de cada dimensión definida.

3. Redactar posibles preguntas a partir de los indicadores.
4. Seleccionar las mejores preguntas en función de su calidad técnica.
5. Confeccionar el primer borrador.
6. Estudio piloto del cuestionario.
 - a) Revisión por parte de especialistas.
 - b) Revisión por parte de los sujetos.
 - c) Aplicación piloto.
7. Elaboración de la versión definitiva del cuestionario.
8. Aplicación del cuestionario.
9. Análisis de las respuestas.
10. Redacción del informe.

3.4. Ventajas y limitaciones

Entre las ventajas del cuestionario podemos señalar las siguientes:

- Pueden ser aplicados a un conjunto amplio de sujetos.
- Pueden ser aplicables a multitud de variables y temas.
- El coste económico de su utilización no es excesivo.
- Ahorro de tiempo.
- No es necesario que los aplicadores del cuestionario conozcan a fondo la evaluación que se está llevando a cabo, por lo que puede utilizarse personal ajeno a la propia evaluación.
- Mayor normalización y estandarización de la situación.

Entre sus limitaciones podemos señalar las siguientes:

- La información recogida a través de los cuestionarios debe ser cotejada y contrastada con informaciones provenientes de otras técnicas de recogida de información, ya que existen diferentes fuentes de sesgos que pueden afectar a la calidad de la información recogida.
- Las preguntas deben ser planteadas previamente y deben tener un significado unívoco para los sujetos que contestan, aspecto difícil de conseguir en muchas ocasiones.
- La información obtenida es superficial; es decir, no permite la posibilidad de profundizar en las respuestas de los sujetos.
- Si el cuestionario es enviado por correo para su contestación, la pérdida de cuestionarios es muy importante.
- El porcentaje, en algunas ocasiones excesivamente elevado, de preguntas no contestadas.

4. Las pruebas de aprendizaje

Las pruebas de aprendizaje o de rendimiento de los alumnos tratan de recoger información acerca de lo que los alumnos han aprendido tras el proceso de enseñanza-aprendizaje. Entre las pruebas de aprendizaje se pueden distinguir:

4.1. Las pruebas objetivas

Las pruebas objetivas son instrumentos que analizan el dominio que logran los alumnos de los objetivos de aprendizaje específico de las diversas áreas curriculares. Para estimar el nivel de aprendizaje de los alumnos, se utilizan una serie de preguntas claras y breves cuya respuesta exige la utilización mínima de palabras permitiendo la corrección objetiva. Son instrumentos semejantes a los tests de rendimiento, pero contruidos con un formalismo menor (normalmente, por el profesor) y al igual que ellos pueden presentar formas muy diversas: evocar conocimientos, reconocer, solucionar problemas, comparar, observar, analizar, interpretar gráficos y mapas, etc. (véase cuadro 7.2).

Cuadro 7.2. Fases en la planificación de una prueba objetiva. (Adaptado de Del Rincón *et al.*, 1995, p. 142)

FASE 1: PLANTEAMIENTO DE LA PRUEBA

1. Identificación de los objetivos de la prueba.
 2. Selección y delimitación de contenidos.
 3. Elaboración de la tabla de especificaciones.
-

FASE 2: ELABORACIÓN DE LA PRUEBA

1. Redacción de ítems.
 2. Revisión de los ítems por expertos.
 3. Confección del original de la prueba.
 - 3.1. Instrucciones.
 - 3.2. Distribución de los ítems.
 - 3.3. Hoja de respuestas.
 4. Revisión del original.
 5. Normas de corrección.
 6. Estudio preliminar o ensayo.
 7. Impresión definitiva.
-

FASE 3: APLICACIÓN Y ORGANIZACIÓN DE LOS RESULTADOS

1. Presentación del instrumento.
 2. Administración de la prueba.
 3. Corrección.
 4. Confección de la tabla de resultados.
-

Cuadro 7.2. (continuación)

FASE 4: ESTUDIO EMPÍRICO DE LA PRUEBA

1. Análisis estadístico.
 2. Validez y fiabilidad de la prueba.
 3. Análisis de ítems.
 4. Baremos.
-

FASE 5: SÍNTESIS SOBRE EL ESTUDIO

1. Resultados más significativos.
 2. Optimización de la prueba elaborada.
 3. Sugerencias para posteriores aplicaciones.
-

4.2. Las pruebas de ensayo

En las pruebas de ensayo se pide al alumno que elabore y estructure una respuesta más o menos amplia o restringida. Su utilización es adecuada en las situaciones en las que el alumno tiene que (Cabrera, 2000a):

- Seleccionar y organizar la información.
- Realizar exposiciones utilizando los conceptos y razonamientos propios de la materia.
- Argumentar y valorar situaciones complejas.

En el cuadro 7.3 se presentan las normas de elaboración y corrección de una prueba tipo ensayo.

4.3. Las pruebas de ejecución

En estas pruebas se intenta que el alumno simule lo más posible una situación real que debe afrontar fuera de la escuela. Normalmente, se ponen en juego varias capacidades de la persona. La evaluación basada en ejecuciones requiere que los sujetos que están siendo evaluados «demuestren, construyan, desarrollen un producto o solución a partir de unas condiciones y estándares definidos» (García Jiménez *et al.*, 2002, p. 263). Entre los procedimientos de evaluación de ejecuciones, el portafolios es el que está más en auge (véase cuadro 7.4). Tal y como señalan García Jiménez *et al.* (2002, p. 263) «la carpeta o portafolios es un procedimiento de evaluación de ejecuciones, que se apoya en la recogida y almacenamiento de la información sobre los logros o adquisiciones hechas por una persona durante un periodo de formación». O, como señala Mateo (2000, p. 83) «consiste en una compilación de trabajos del alumno, recogidos a lo largo del tiempo, que aportan evidencias respecto a sus conocimientos, habilidades e incluso de su disposición para actuar de

Cuadro 7.3. Normas para la elaboración y corrección de una prueba de ensayo. (Tomado de Cabrera, 2000a, p. 186)

Normas para la ELABORACIÓN	Normas para la CORRECCIÓN
<ul style="list-style-type: none"> • Restringir su uso a resultados de aprendizaje que no puedan valorarse con prueba objetiva. • Formular cada pregunta de manera que exprese con claridad el tipo de conducta que debe orientar la respuesta. De aquí que siempre deban utilizarse verbos que expliciten la conducta: «explica...», «analiza comparativamente...», «describe...», «ilustra con un ejemplo...», «enumera...». • Hacer algunas precisiones sobre el contenido de la pregunta para indicar la profundidad y amplitud que se espera de la respuesta. • Indicar el límite aproximado de tiempo o espacio para las preguntas (importante en las pruebas de ensayo restringido) es un buen indicador del nivel de elaboración que se espera de la respuesta. • Las preguntas pueden tener igual o distintos valores en el conjunto de la prueba. Si es diferente, indicar su valor al lado de la pregunta es también una señal del nivel de respuesta que se espera. 	<ul style="list-style-type: none"> • Preparar un esbozo de la respuesta que cabe esperar en cada una de las preguntas. Se especifican los contenidos que se tendrán en cuenta para valorar las preguntas. • Determinar qué factores no directamente relacionados con la temática se tendrán o no en cuenta y el peso máximo que se le dará en la valoración global: presentación, nivel de redacción, legibilidad, etc. • Para calificar las preguntas pueden utilizarse dos procedimientos distintos: <ol style="list-style-type: none"> a) Analítico o por puntos: en el modelo de respuesta a la pregunta, se especifica la puntuación que se otorgará a cada contenido que debe aparecer en la respuesta; la puntuación global de la pregunta se obtiene sumando sus puntuaciones parciales. b) Global: se valora en general la respuesta, sin matizar puntuaciones por cada aspecto del contenido. Se establecen niveles de respuesta: por ejemplo, muy bueno, bueno, suficiente, flojo, muy flojo, que pueden traducirse o no a puntuaciones numéricas. • Para controlar factores extraños que pueden distorsionar la respuesta, se debe calificar de manera anónima y mezclar las pruebas para que no guarden el orden en que fueron entregados por las personas. • Si la decisión a tomar a partir de la calificación de la prueba es importante para las personas, se aconseja que sean calificadas por más de dos evaluadores a fin de asegurar la estabilidad y la consistencia de la calificación.

**Cuadro 7.4. Elementos a considerar en el diseño del portafolios.
(Tomado de García Jiménez *et al.*, 2002, p. 265)**

1. Propósitos del portafolios.
 2. Contenidos del portafolios:
 - 2.1. Las metas de aprendizaje.
 - 2.2. Las tareas de aprendizaje.
 - 2.3. Un modelo que estructure los contenidos.
 - 2.4. Los materiales o artefactos.
 - 2.5. Los criterios de puntuación.
 - 2.6. La validez del portafolios.
-

determinadas maneras». Mediante él, el profesor y el alumno podrán controlar el progreso de las actitudes, las habilidades y el conocimiento.

5. Los tests estandarizados

5.1. Concepto

La palabra «test» significa prueba, y como verbo significa comprobar, ensayar, probar. Por otra parte, esta palabra proviene de la raíz latina *testis* que en castellano ha derivado en testigo, atestiguar, testimonio, etc. La palabra test ha sido aceptada internacionalmente para designar un tipo de exámenes especiales muy utilizados en Psicología y Educación.

Un test es un reactivo que aplicado a un sujeto será el testigo de su nivel de instrucción, actitud o forma de ser. Por otra parte, Anastasi (1998) lo define como una medida tipificada y objetiva de una muestra del comportamiento. De la Orden (1989, p. 228) define el test como una estructura estimular para provocar respuestas que se aplica en condiciones determinadas para obtener información sobre un sujeto o un grupo de sujetos. Por su parte Mehrens y Lehmann (1982) señalan que los tests estandarizados ofrecen métodos para conseguir muestras de conducta mediante procedimientos uniformes, es decir, mediante la administración de una misma serie fija de preguntas con las mismas instrucciones y condiciones de tiempo y un procedimiento de corrección cuidadosamente definido y uniforme (Padilla Carmona, 2002). Por último, Forner y Latorre (1996, p. 140) y Lukas *et al.* (2000a) en el *Diccionario de Investigación Educativa* definen el test como «prueba estandarizada para medir un comportamiento de una persona, comparándolo con el de otras personas que se encuentran en la misma situación» (p. 239).

Como en cualquier otra ciencia, tomando una muestra, se hacen unas pruebas y en estas pruebas aparece el comportamiento del individuo. De esta forma, por ejemplo, un test de inteligencia será un conjunto de preguntas,

problemas o tareas cuya solución dependerá de la capacidad intelectual del individuo.

La primera condición que tienen que cumplir los tests es que las respuestas que se piden en los ítems o elementos nos tienen que dar la característica que pretendemos medir. Por ejemplo, si un profesor quiere medir el nivel de vocabulario de sus alumnos escogerá una muestra representativa de palabras y verá cómo responden sus alumnos. Verdaderamente estará en la longitud y en la naturaleza de la prueba lo que mida el comportamiento en el cual estamos interesados.

Todos los tests tienen un valor predictivo como indicadores del comportamiento. Decir que un niño conoce 30 palabras, en principio, no quiere decir nada, pero si existe una fuerte relación entre esas 30 palabras y el nivel de vocabulario del niño, entonces, ese test de sólo 30 palabras nos valdrá para tener información acerca del nivel de vocabulario de los niños.

Los elementos o ítems del test no tienen por qué parecerse al comportamiento que queremos medir, basta con que haya una correspondencia o relación empírica. Ahora bien, con la predicción se debe tener cuidado, puesto que sólo se puede utilizar cuando ha sido probado por medio de la experiencia que una muestra del comportamiento es el índice de otra. Una vez que esto haya sido probado se aplica el test a una muestra representativa de la población a la que está destinado el test. A este grupo o muestra se le denomina grupo normativo. Los resultados que da el grupo normativo, después de ser analizados estadísticamente, serán los que se utilizarán como normas o reglas. Cuando cualquier individuo de la población realice el test será comparado en relación al nivel del grupo normativo. Este proceso de establecer reglas o normas al test se conoce con el nombre de tipificación.

Este tipo de test estandarizado es el que ha prevalecido en la evaluación educativa durante largo tiempo y es conocido con el nombre de test normativo. Cuando utilizamos un test normativo de rendimiento para evaluar lo que un alumno conoce de una materia curricular en particular, lo que hacemos es comparar su puntuación con la del resto de los alumnos para conocer de esa manera la situación relativa que ocupa dentro de su grupo. No obstante, no miden lo que el alumno sabe en términos absolutos.

La alternativa a los tests estandarizados normativos son los tests de referencia criterial o tests criterioles sin más. Los tests criterioles permiten establecer el logro del alumno por referencia con un objetivo educativo específico. Las características más importantes de los test criterioles serían las siguientes (Mateo, 2000a, p. 87):

- La interpretación de la ejecución individual se obtiene fundamentalmente por contraste respecto a algún criterio fijado previamente y aceptado como valioso.
- De la ejecución individual nos interesa la composición analítica (los componentes) del contenido objeto de evaluación y la especificación de los procesos implicados en la ejecución.

- Los resultados así obtenidos permitirán posteriormente la puesta en marcha de estrategias de mejora a nivel individual, curricular o institucional.
- La selección de los elementos que componen las pruebas construidas de acuerdo al enfoque criterial se realiza sobre planteamientos de validez de contenido y no basadas en distribuciones teórico-matemáticas.

En el cuadro 7.5 pueden apreciarse las diferencias más significativas entre los tests normativos y los tests criterios.

Cuadro 7.5. Diferencias entre los tests normativos y los tests criterios

Tests normativos	Tests criterios
Selección de ítems en función de criterios estadísticos.	Selección de ítems en función de criterios de validez de contenido.
Ofrecen información sobre los aprendizajes en momentos específicos.	Ofrecen información sobre los cambios en los niveles de logro de los aprendizajes a lo largo del tiempo.
La puntuación individual se interpreta en función del nivel de rendimiento del grupo al que pertenece.	La puntuación individual se interpreta en contraste con un dominio de rendimiento claramente definido y delimitado.
Cubren un dominio extenso de objetivos de aprendizaje, aunque el número de ítems para cada objetivo es reducido.	Cubren un dominio de aprendizaje mucho más reducido, aunque el número de ítems para cada objetivo es amplio.
Predominan los ítems de dificultad media.	Los ítems tienen distintos niveles de dificultad en función de la tarea exigida.
Su interpretación requiere un grupo normativo.	Su interpretación requiere un dominio educativo claramente delimitado y definido.
Su uso está unido a la evaluación sumativa.	Su uso está unido a la evaluación formativa.

5.2. Clasificación de los tests

Los tests estandarizados se pueden clasificar atendiendo a diferentes criterios. Hoy en día no existe una clasificación universalmente aceptada que satisfaga a toda la comunidad científica. Es por ello por lo que se proponen diversas clasificaciones, si bien en la mayoría de los casos los criterios de ordenación utilizados son similares. En el cuadro 7.6 se han seleccionado y agrupado los criterios más utilizados en la clasificación de los tests.

Cuadro 7.6. Clasificación de los tests estandarizados

Criterio	Tipos de tests
En función del atributo a medir.	1. Tests de aptitudes , de rendimiento y de personalidad . 2. Tests de capacidad (aptitudes y rendimiento) y de personalidad . 3. Tests cognoscitivos y no cognoscitivos o afectivos .
En función de la regla general del método.	Tests psicométricos y proyectivos .
En función de la naturaleza de la medición.	Tests de realización máxima y de realización típica .
En función del modo en que se interpretan los resultados.	Tests normativos y criteriales .
En función de la administración.	Tests colectivos e individuales .
En función del tiempo.	Tests de potencia y de velocidad .
En función del examinador.	Tests personales e impersonales .
En función del material necesario.	Tests de papel y lápiz y manipulativos .
En función de la respuesta.	Tests de selección o elección y de facilitación o invención .

5.3. Procedimiento

El procedimiento de construcción de un test estandarizado supera con creces el propósito de este libro. Sin embargo, en los cuadros 7.7 y 7.8 se presentan las fases y etapas a seguir en la construcción de un test normativo y de un test criterial.

Cuadro 7.7. Etapas en la construcción de un test normativo.
(Tomado de Crocker y Algina, 1986, p. 66)

Etapas en la construcción de un TEST NORMATIVO
1. Identificar el fin principal para el que será utilizado el test.
2. Identificar las conductas que representan el constructo o definir el dominio.
3. Preparar un conjunto de especificaciones del test, delimitando la proporción de ítems que debería incluirse en cada tipo de conducta identificado en el paso anterior.
4. Construir un conjunto inicial de ítems.
5. Tener los ítems repasados y revisados.
6. Mantener las pruebas de ítems preliminares.
7. Probar los ítems en una población representativa de la población para la que el test está destinado.
8. Determinar las propiedades estadísticas de los ítems y, cuando sea apropiado, eliminar los ítems que no cumplan los criterios requeridos.
9. Elaborar los estudios de fiabilidad y validez para la forma final del test.
10. Desarrollar principios generales para la administración, corrección e interpretación de las puntuaciones del test (normas, puntos de corte, estándares de ejecución, etc.).

Cuadro 7.8. Pasos en la construcción de un test referido a criterio (Hambleton, 1990)

Etapas en la construcción de un TEST CRITERIAL

1. Consideraciones preliminares en la construcción de un test:
 - a) Declarar los propósitos del test.
 - b) Especificar las áreas de contenido en forma de un anteproyecto.
 - c) Identificar los grupos que se van a medir.
 - d) Especificar la cantidad de tiempo, dinero y expertos necesarios para realizar el proyecto.
2. Preparación y revisión de objetivos.
3. Planificación adicional del test:
 - a) Seleccionar los objetivos que se van a medir en el test.
 - b) Determinar la longitud aproximada del test y el número de ítems por objetivo.
 - c) Identificar a las personas que van a escribir los ítems.
4. Preparación de los ítems del test para medir los objetivos y edición inicial de los ítems.
5. Valoración de la fiabilidad y validez del contenido:
 - a) Uso de especialistas en medida y contenidos para valorar:
 1. la congruencia objetiva de los ítems,
 2. la calidad técnica de los ítems,
 3. la representatividad de los ítems,
 4. los sesgos de los ítems.
 - b) Examen de campo de los ítems para:
 1. detectar ítems defectuosos,
 2. comprobar o chequear sesgos,
 3. determinar niveles de dificultad e índices de discriminación del ítem,
 4. determinar la consistencia de las puntuaciones y/o decisiones.
6. Revisión de ítems del test.
7. Ensamblaje del test:
 - a) Determinar la longitud y el número de ítems del test por objetivo.
 - b) Seleccionar los ítems.
 - c) Valorar la representatividad de los ítems.
 - d) Preparar las preguntas para dirigir y administrar el test.
 - e) Preparar la composición y los cuadernos del test.
 - f) Preparar las claves de puntuación y las hojas de respuesta.
8. Selección e implementación de un método de posicionamiento del estándar o punto de corte.
9. Administración del test.
10. Valorar la fiabilidad y validez de la puntuación del test (es opcional la recopilación de normas sobre puntuaciones del test).
11. Preparación de un manual técnico y de un manual del administrador.
12. Recogida continua de información técnica adicional y actualización del manual técnico.

6. Otros procedimientos para la recogida de la información

Además de las cinco técnicas presentadas (observación, entrevista, cuestionario, pruebas de aprendizaje y tests estandarizados), que sin ningún género de dudas son las más utilizadas en la evaluación educativa, vamos a presentar otras técnicas de recogida de información de manera muy breve:

6.1. Las escalas de actitudes

Las escalas de actitudes¹ son instrumentos que se utilizan para determinar diferencias de grado o intensidad respecto a algún objeto actitudinal (Del Rincón *et al.*, 1995). A través de una serie de preguntas que se formulan al sujeto, éste queda situado dentro de un continuo de mayor a menor grado de favorabilidad respecto a la actitud que se está midiendo (Cabrera, 2000a). Mediante las escalas de actitudes es el propio individuo quien informa sobre sus opiniones y sentimientos y por ello debe garantizarse su sinceridad asegurando que se cumplen estos tres requisitos (Del Rincón *et al.*, 1995, p. 180):

1. Ser consciente de su actitud sobre el aspecto estudiado.
2. No estar afectado por la situación de prueba o por las posibles consecuencias de su revelación.
3. La escala debe especificar con detalle y objetividad el significado de los distintos grados.

Las escalas de actitudes más conocidas y utilizadas en la evaluación educativa son las siguientes:

- a) *Escalas diferenciales o escalas de Thurstone*: Son escalas que tienen entre 20 y 22 ítems que han sido posicionados en la escala mediante la clasificación llevada a cabo por jueces. Después se administra a los sujetos para que señalen uno o varios juicios con los que están de acuerdo, a partir de los cuales se obtiene su puntuación en la escala.
- b) *Escalas aditivas o escalas de Likert*: Son escalas que están formadas por una serie de afirmaciones que expresan juicios positivos y negativos sobre los cuales el sujeto debe mostrar su grado de acuerdo (desde totalmente de acuerdo hasta completamente en desacuerdo, pasando por los grados intermedios que se quieran intercalar siempre y cuando sea una gradación simétrica).
- c) *Escalas de diferencial semántico o escalas de Osgood*: Estas escalas se basan en la idea de que las personas utilizan adjetivos para expresar sus valoraciones. El procedimiento, por tanto, consiste en presentar al sujeto escalas bipolares de adjetivos antónimos (agradable-desagrada-

ble, valioso-inútil, etc.) en los que la persona señala su posición situándose más cerca de uno de los dos polos según califique al objeto en cuestión. Estas calificaciones muestran la imagen que tiene el sujeto del objeto en cuestión e indirectamente su actitud hacia él (Cabrera, 2000a).

6.2. La sociometría

La sociometría trata de medir las relaciones sociales a nivel científico. Puede definirse como el estudio de la evaluación de los grupos y de la posición que en ellos ocupan los individuos, prescindiendo del problema de la estructura interna de cada individuo.

Para llevar a cabo un sociograma, que es la aplicación práctica de la sociometría, se debe construir previamente un test sociométrico en el que se pregunta al sujeto por la elección o rechazo de entre los demás miembros de su grupo. A partir de estas elecciones o rechazos, los resultados se plasman, bien en un sociograma (lo más usual), o bien en un psicograma. El sociograma representa de manera simbólica (mediante números para identificar a los sujetos y flechas para representar las elecciones o rechazos entre ellos) la estructura interna del grupo.

Mediante la sociometría podemos conocer el nivel de aceptación o rechazo que posee un sujeto en su grupo, conocer el grado de cohesión del grupo, localizar los líderes y los sujetos rechazados, etc. No obstante, mediante la sociometría no podemos conocer las razones por las que los grupos están estructurados de esa manera.

6.3. El grupo de discusión

El grupo de discusión es un conjunto de personas que dialogan y debaten juntas sobre algún tema de interés aportando sus perspectivas, percepciones y puntos de vista específicos y particulares. Tal y como señala Cabrera (2000a), normalmente, llevan a cabo varias discusiones en grupo con participantes distintos para ver cuáles son las tendencias y opiniones sobre el tema en cuestión. El posterior análisis de las discusiones posibilita recoger información sobre cómo se percibe una situación y las expectativas o demandas que genera.

Krueger (1991) señala las características de los grupos de discusión:

- El número de individuos participantes en un grupo de discusión está entre 4 y 12.
- En función del objetivo de la evaluación, los individuos participantes son homogéneos respecto a un conjunto de características de interés.
- Es mejor que las personas participantes sean desconocidas entre sí.

- El propósito del grupo de discusión es recoger información.
- El tipo de información recogida es de naturaleza cualitativa.
- Los temas a discutir están elegidos y ordenados a priori.
- Además de moderar, la función del moderador es observar, escuchar y analizar de manera inductiva lo que está sucediendo en el grupo de discusión.

Las ventajas del grupo de discusión frente a otras técnicas son las siguientes (Cabrera, 2000a, p. 220):

- Coloca al participante en una situación social de interacción con los otros que se parece más a una situación real de la vida.
- Flexibilidad para explorar desviaciones inesperadas del discurso.
- Reducido coste de tiempo y personal.

6.4. El autoinforme

El autoinforme lleva implícito el concepto de autoobservación, dado que supone una introspección del sujeto sobre determinadas cuestiones que le son presentadas. El observado y el observador coinciden. El autoinforme puede ser de distinta índole, desde un planteamiento escrito en el que se muestran al sujeto una serie de cuestiones sobre su comportamiento hasta registros de tipo diario. El diario es quizás el tipo de autoinforme más utilizado. Como señalan Del Rincón *et al.* (1995, p. 344) «los diarios son relatos descriptivos escritos que recogen los comentarios reflexivos de los acontecimientos que tienen lugar en la vida de una persona de manera regular y continuada». Zabalza (1991) identifica tres tipos de diarios que pueden ser utilizados en el ámbito educativo:

- El diario como *organizador estructural de clase*, que está planteado como organizador secuenciador de la actividad docente.
- El diario como *descripción de las tareas*, que está planteado para plasmar en él las actividades que realizan tanto el profesor como el alumnado.
- El diario como *expresión de las características del alumnado y del profesorado*, que está planteado para plasmar en él los comportamientos, las conductas y las acciones de los protagonistas del proceso de enseñanza-aprendizaje.

6.5. Las historias de vida

Del Rincón *et al.* (1995) clasifican dentro de los documentos personales los relatos de vida y las historias de vida. Los relatos de vida (*life story*) son narraciones biográficas de una persona referidas a su vida y recitadas tal y como él las ha vivido. Las historias de vida (*life history*), sin embargo, además de recoger los relatos de vida, éstos son complementados con entrevistas a personas del entorno y cualquier otro documento o información que permite una reconstrucción más exhaustiva y objetiva de la biografía (Del Rincón, 1995).

6.6. Los documentos

Otra fuente de información importante utilizada en la evaluación educativa es el análisis de los documentos escritos por los sujetos. Woods (1987) señala que estos materiales pueden considerarse como instrumentos cuasiobservacionales. El análisis de documentos complementa a otras estrategias de recogida de información y en algunos casos es la única fuente de información debido a la inaccesibilidad por medio de otras formas.

El análisis de documentos es una actividad sistemática y planificada que consiste en examinar documentos escritos. Pretende obtener información útil y necesaria para dar respuesta a los objetivos identificados en el planteamiento de la investigación², como confirmar una hipótesis, categorizar un conjunto de eventos, reconstruir acontecimientos, contrastar y validar información, etc.

(Del Rincón *et al.*, 1995, p. 341)

Mediante el análisis de documentos podemos obtener información retrospectiva acerca de algún fenómeno, evento o institución, y es por ello por lo que es muy utilizado (casi imprescindible) al inicio de una evaluación de un programa o de un centro. Tras ese análisis se podrá conocer el propósito, la justificación y la historia del programa en cuestión (Del Rincón *et al.*, 1995).

6.7. La técnica Delphi

Mediante esta técnica, de reciente utilización en la evaluación de programas, se pretende obtener respuestas consensuadas de un grupo de expertos a preguntas específicas. Se considera una técnica cualitativa debido a la búsqueda de consenso entre las distintas perspectivas de los expertos participantes a través del conocimiento de las opiniones del resto de los expertos del grupo.

El procedimiento de esta técnica consiste básicamente en enviar a todos los expertos un primer cuestionario con las preguntas abiertas, de forma que

contesten de manera confidencial e individual. Tras recibir las respuestas, el evaluador las codificará y las categorizará y construirá un segundo cuestionario incluyendo esas respuestas, cuestionario que será nuevamente enviado a los expertos para que indiquen su grado de acuerdo con las respuestas surgidas. De esta manera y tras varias tentativas se intentará llegar a un consenso entre todos los expertos. Puede consultarse una buena revisión de esta técnica en Ballester Brage (2000).

7. La calidad de la información

La información recogida, independientemente del procedimiento utilizado es, como señala Tójar (2001, p. 141), «un segmento sesgado de la realidad». Por ello, su análisis puede reportarnos conclusiones no veraces y de ahí que debamos asegurar la calidad de la información recogida.

Los criterios de calidad de los datos cuantitativos son su fiabilidad y su validez, que son propiedades de la medida y no de los instrumentos. Tras la aplicación de un instrumento de medida a un fenómeno o evento para obtener unas medidas es posible estimar su calidad a partir del análisis de su fiabilidad y validez.

Preguntarse por la **fiabilidad de la medida** es preguntarse si ésta es consistente, estable y precisa. Es decir, se plantea hasta qué punto la medida no es atribuible a factores del azar sino al factor concreto que se está midiendo. A nivel operativo diríamos que la fiabilidad es el coeficiente de correlación entre dos conjuntos de puntuaciones obtenidos independientemente, en formas paralelas de la prueba, para un mismo grupo.

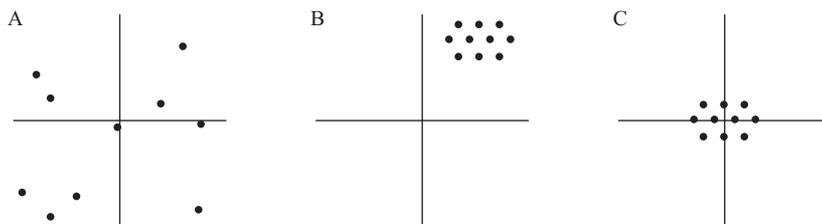
Por **validez** entendemos «el grado en que la medida refleja con exactitud el rasgo, característica o dimensión que se pretende medir» (Del Rincón *et al.*, 1995, p. 74). Es decir, debemos preguntarnos si realmente el instrumento mide lo que dice medir. Esto puede parecer una perogrullada si hablamos de mediciones físicas tales como la talla o el peso, en cuyo caso esa pregunta no tendría sentido si utilizáramos un metro para medir la talla y una báscula para medir el peso. Pero, sin embargo, en Ciencias Humanas la misma pregunta toma una dimensión mayor debido a la «distancia» entre el objeto a medir y el instrumento. Cualquier instrumento de medida representa una muestra de comportamiento a partir del cual se pretenden hacer ciertas inferencias. La validez se refiere a la adecuación, significación y utilidad de las inferencias específicas hechas a partir de las medidas obtenidas por los instrumentos de medida utilizados; es decir, el conjunto de pruebas que han de recogerse para garantizar la pertinencia de esas inferencias. Más que el instrumento de medida en sí, lo que se validan son las inferencias. Convencionalmente, la validez ha sido definida como el grado en que un test mide lo que pretende medir.

Tradicionalmente se han distinguido tres tipos de validez:

- *La validez de contenido*, que hace referencia a la representatividad y relevancia de la muestra de conductas recogidas por la medida.
- *La validez empírica o de criterio* hace referencia a la relación que existe entre las medidas obtenidas a través del instrumento utilizado y alguna otra conducta criterio que previamente ya ha sido identificada como la conducta que el instrumento pretende medir. Es decir, cuando se trata de validar un test, por ejemplo, el criterio se refiere a una medida directa e independiente de lo que está destinado a medir el test.
- *La validez de constructo o de concepto* hace referencia a la determinación del grado en que la prueba mide el rasgo psicológico o intelectual teórico al que se refiere. Se corresponde con el nivel máximo de validación que puede alcanzar una prueba y recoge el estudio de cualquier otro tipo de validez.

En cuanto a la relación entre la fiabilidad y la validez, se debe señalar que la primera (la fiabilidad) es una condición necesaria, pero no suficiente de la segunda (la validez). O sea, el hecho de medir algo con precisión no garantiza que la medida sea exacta, que recoja lo que pretendemos medir. Suen y Ary (1989) presentan el caso de un testigo en un juicio. El testimonio del testigo ha de ser fiable y los hechos serán fiables si son confirmados igualmente por el resto de los testigos, aunque dicha fiabilidad no garantiza la exactitud o validez, ya que los testigos podrían haberse puesto de acuerdo de forma deliberada para no relatar los hechos con veracidad. Por otra parte, Tójar presenta otro ejemplo en la figura 7.1 extraído de la *Teoría del tiro* que ilustra la relación entre la precisión (fiabilidad) y exactitud (validez).

Figura 7.1. Precisión y exactitud en el tiro (Tójar, 2001, p. 142)



Los tres casos de la figura 7.1 representan las diferencias entre precisión y exactitud en los resultados de los impactos de proyectiles. En el caso A, los tiros no han sido ni precisos ni exactos. En el caso B, al estar los impactos agrupados existe precisión, pero no exactitud. Sin embargo, en el caso C, los impactos se agrupan en la diana por lo que tendríamos precisión y además exactitud.

Al tratar de la calidad de la información cualitativa recolectada, algunos autores han propuesto aplicar los mismos conceptos de fiabilidad y validez, aunque la mayoría propone un cambio de perspectiva. En este sentido, la propuesta de criterios de calidad para los datos cualitativos que más aceptación ha tenido (Tójar, 2001) en la investigación y evaluación educativas ha sido la realizada por Guba y Lincoln (1994). El criterio por excelencia para la investigación y evaluación naturalistas sería la veracidad que se articula en las siguientes características:

1. *Credibilidad*, que muestra que la evaluación se ha realizado de forma pertinente, garantizando que la realidad fue descrita e identificada con exactitud. La indagación, como señalan Del Rincón *et al.* (1995, p. 255), debe ser «creíble para los constructores de realidades originales múltiples».
2. *Transferibilidad*. El evaluador naturalista evita las generalizaciones basándose en el principio de que todos los fenómenos sociales o de conducta dependen del contexto. Por ello, aunque es un concepto cercano al de aplicabilidad, trata de formular hipótesis que puedan aplicarse o transferirse en otras situaciones parecidas (Tójar, 2001).
3. *Dependencia*. Hace referencia a la consistencia o estabilidad de determinados resultados, pese a que se tienen en cuenta las condiciones cambiantes en el fenómeno. Utilizando métodos de recogida de datos que se complementen o incluso se solapen se podrá estimar si determinados elementos se muestran estables.
4. *Confirmabilidad*. Hace referencia al concepto tradicional de objetividad que consiste en confirmar la información, la interpretación de los significados y la generación de conclusiones. El evaluador debe actuar con neutralidad alejando cualquier atisbo de prejuicio o expectativa previa.

No obstante, como señala Tójar (2001), esta clasificación de criterios de calidad de la información cualitativa ha sido criticada por haber sido construida con una lógica propia de las perspectivas empírico-analíticas. Ante esta situación, Guba y Lincoln (1989) y Lincoln (1995) han propuesto nuevos conceptos para garantizar la calidad de la información cualitativa que giran alrededor del concepto central de autenticidad. Son los siguientes (Tójar, 2001, p. 145):

- *Imparcialidad*, como oportunidad de presentación de los diferentes puntos de vista sobre una misma información.
- *Autenticidad ontológica*, respeto en la maduración, elaboración y evaluación de las construcciones, de las concepciones que tienen las personas.
- *Autenticidad educativa*, supone un paso más allá de las propias elaboraciones y la posibilidad de contrastarlas con otras para destacar cómo sistemas de valores alternativos que proponen soluciones diversas.

- *Autenticidad catalizada*, se refiere al grado en que una acción es promovida y agilizada por procesos de evaluación educativa.
- *Autenticidad táctica*, como potencial o capacitación de las personas implicadas para la acción.

La triangulación es la estrategia más utilizada para dotar de autenticidad o veracidad a la información cualitativa.

8. Recursos informáticos para la recogida de la información

Desde los primeros pasos en el desarrollo de la informática ha habido intentos de aplicación de sus recursos al ámbito de la investigación y también al de la evaluación y el diagnóstico. La utilización de las nuevas tecnologías de la información en el ámbito de la evaluación ha abierto nuevas perspectivas que hasta hace bien poco eran insospechables.

Los usos de la informática en la evaluación se dan en los siguientes ámbitos (Jornet, 2001).

8.1. La edición, corrección y proceso de información de pruebas de evaluación

En este punto entrarían los lectores ópticos que con su *software* específico permiten la corrección automatizada de las pruebas de evaluación o diagnóstico. Además, algunas casas comerciales editoras de pruebas (TEA, CEPE, COSPA, EOS, etc.) han creado *software* ad hoc para emitir informes de resultados de diversos tipos. Aun cuando el juicio profesional del educador será siempre importante y nunca podrá ser obviado, las ventajas que ofrece el ordenador en la corrección mecanizada de tests y en la edición de informes son palpables:

- Los errores en la corrección y baremación de las pruebas son casi nulos.
- Los datos son procesados en menor tiempo.
- Facilita la combinación de resultados en las pruebas compuestas por subpruebas.
- Facilita el proceso estadístico.

8.2. El diseño y presentación de ítems y tests

Los nuevos procesadores de información ofrecen nuevas posibilidades en la creación de ítems y de tests integrados por diferentes tipos de información (textos, gráficos, movimiento de gráficos, etc.) que pueden incluso ser aplicados mediante el ordenador.

8.3. Sistemas de evaluación dinámica y adaptada a las personas

Sin duda, una de las mayores posibilidades que brinda la informática en los procesos diagnóstico-evaluativos de las personas es la individualización de los tests según el nivel de desempeño mostrado. Sobre esto hablaremos más adelante en el apartado dedicado a la recogida de datos.

8.4. Evaluación y diagnóstico de personas con necesidades especiales

Se están haciendo esfuerzos de aplicaciones informáticas y telemáticas a fin de mejorar la evaluación y el diagnóstico de personas con necesidades educativas especiales (discapacitados, fracaso escolar, sobredotación intelectual, etc.) (Jornet, 2001). De entre ellos, las personas con discapacidad motora o verbal, es decir, con graves problemas de comunicación en muchos casos, son los que han suscitado un mayor interés y hacia los que se van creando aplicaciones informáticas que mejoren la comunicación.

8.5. Recogida de datos

Junto con el análisis de datos, es en la recogida de información en donde la informática ha supuesto un gran avance. Como señala Jornet (2001) citando a Belloch (2000), «los cambios en la tecnología de la información y la comunicación, unidos a los ocurridos en la medición educativa y en las aportaciones de la psicología cognitiva y las teorías del procesamiento de la información, han impactado de forma decisiva en la incorporación de estos avances en el ámbito de la evaluación y el diagnóstico, tanto en Psicología como en Educación» (p. 307).

El uso de la informática tanto en la construcción como en la administración y/o gestión de pruebas, además de posibilitar la operativización de los procesos hasta límites insospechados, ha constituido también un cambio cualitativo que permite recoger información procesual, de tal forma que se asume que a través de los tests se puede apreciar el desempeño cognitivo de los sujetos.

En el cuadro 7.9 se presentan los tipos de instrumentos informáticos desarrollados para la recogida de información.

De entre los instrumentos citados en el cuadro los TAI, los TADI y los TME son los que están suscitando un mayor interés. Los tres se basan en la idea de adaptar la prueba al nivel de la persona examinada (aunque la forma de hacerlo es diferente en cada uno de ellos) de tal forma que se mejora la precisión de la medida y se ahorra en el tiempo de aplicación (Muñiz y

Cuadro 7.9. Síntesis de instrumentos informatizados para la recogida de información. (Adaptado de Jornet, 2001, pp. 312-317)

Denominación de la aplicación	Siglas	Descripción
<i>Tests Convencionales Informatizados</i>	TCI	Se trata de aplicaciones informáticas que administran tests convencionales, originalmente diseñados para su aplicación en formato de lápiz y papel.
<i>Tests Procesuales Informatizados</i>	TPI	Basados en planteamientos derivados de la Psicología Cognitiva pretenden recabar información acerca del proceso subyacente a la resolución de problemas. No obstante, se basan en un mismo formato de aplicación igual para todos los sujetos.
<i>Sistemas Integrados de Evaluación Informatizada</i>	SIEI	Se basan en el desarrollo de Bancos de Ítems calibrados mediante TRI (teoría de respuesta a los ítems). Sobre el SIEI se sustentan todas las operaciones técnicas de conformación de las pruebas, que están sujetas a determinantes técnicos complejos.
<i>Generación Automática Informatizada de Ítems</i>	GAI	Se trata de aplicaciones informáticas que gestionan formatos de ítems preestablecidos, que facilitan la generación automática de éstos. Se basan en aplicaciones de Reglas de Generación Automática, como son los <i>Ítem For</i> de Hively o el <i>Diseño de Facetas</i> de Guttman.
<i>Tests Óptimos</i>	TO	Son aquellos que contienen los ítems más adecuados para un determinado objetivo.
<i>Tests Adaptativos Informatizados</i>	TAI	Son aquellos que se adaptan al nivel en la habilidad mostrada por el evaluando. Se aplican, por tanto, diferentes ítems para cada sujeto, presentándose en cada ocasión aquellos que resultan más efectivos para estimar el nivel de habilidad.
<i>Tests Autoadaptados Informatizados</i>	TADI	En la TADI es el evaluando el que selecciona el nivel de dificultad de los ítems, presentándose el número de ítems necesarios para alcanzar una determinada precisión en la estimación de la medida.
<i>Tests Multietápicas</i>	TME	Estos tests realizan el proceso de medición en varias etapas. En la primera etapa se realiza la estimación inicial del nivel del evaluando por medio de un conjunto de ítems. En las etapas siguientes se le administran los ítems que se ajusten a su nivel de habilidad.

Hambleton, 1999). Sin embargo, para el funcionamiento de una herramienta de este tipo es necesario disponer de (Olea y Ponsoda, 1996):

- Un banco de ítems calibrado siguiendo los procedimientos de la teoría de respuesta a los ítems (TRI).
- Un procedimiento matemático de estimación de los niveles de habilidad.
- Una estrategia para la selección sucesiva de los ítems.

García, Gil y Rodríguez (1999a y b) señalan los criterios que deberían seguir (véase cuadro 7.10) los tests adaptativos informatizados (TAI) (aunque también valdrían la mayoría de las indicaciones para los TADI y los TME) en el proceso de enseñanza/aprendizaje.

Cuadro 7.10. Orientaciones para la evaluación de Tests Adaptativos Informatizados según su vinculación al proceso de enseñanza-aprendizaje. (Tomado de García, Gil y Rodríguez, 1999b)

Dimensiones	Orientaciones
<i>Estrategias de aprendizaje</i>	<ol style="list-style-type: none"> 1. Deberían describir diferentes estrategias de aprendizaje utilizadas por el estudiante: exploratorias, basadas en explicaciones, inductivas, deductivas, guiadas. 2. Es necesario que los TAI hagan posible diseñar microestrategias (de repetición o de elaboración) y macroestrategias de aprendizaje (de organización, de regulación y afectivo-emocionales). 3. Los resultados de los TAI deberían ofrecer información sobre las ideas o concepciones erróneas que sostienen los alumnos en un determinado dominio o ámbito de conocimiento.
<i>Ambientes de aprendizaje</i>	<ol style="list-style-type: none"> 4. Los TAI deberían favorecer la creación de ambientes educativos próximos a la capacidad del estudiante. 5. El profesor está en disposición de obtener información sobre el nivel de dominio que cada uno de sus alumnos/clientes posee sobre determinadas habilidades y, gracias a ella, puede rediseñar su enseñanza/terapia. 6. El alumno puede conocer qué nuevas habilidades debe aprender. 7. El tiempo invertido en la evaluación no supone interrumpir la secuencia instruccional ni el tipo de organización que habitualmente se sigue en la clase. El TAI está integrado dentro del proceso de enseñanza. 8. El TAI puede facilitar, con las herramientas instruccionales/técnicas oportunas, una mejor monitorización de la enseñanza y un tratamiento más personalizado del alumno.

Cuadro 7.10. (continuación)

Dimensiones	Orientaciones
	9. El TAI no dificulta la interacción entre los estudiantes o de éstos con el profesor. Por el contrario, gracias al desarrollo de sistemas tutoriales adecuados, el test aumenta la frecuencia y calidad de dichas interacciones.
<i>Transferencia del aprendizaje</i>	<p>10. Los TAI deberían permitir a los profesores conocer en qué grado el estudiante posee las claves o indicios (elementos) que resultan básicos para adquirir y utilizar la capacidad medida.</p> <p>11. Los profesores deberían tener acceso a las claves o indicios en los que apoyarse para enseñar al estudiante en relación con la capacidad medida.</p> <p>12. Los profesores deberían conocer en qué medida las habilidades adquiridas configuran una capacidad de uso habitual en el estudiante y no un mero repertorio de habilidades muertas.</p> <p>13. El profesional de la educación tras la ejecución del alumno debería tener acceso a informaciones relativas a las nuevas habilidades que está en condiciones de adquirir ese alumno sin ayuda (eficiencia del aprendizaje).</p> <p>14. Un TAI debería ofrecer información sobre la posibilidad de que el estudiante pueda aplicar las habilidades adquiridas a diversas situaciones no necesariamente recogidas en el TAI (virtualidad de la transferencia).</p> <p>15. Tras las ejecuciones de un estudiante, el profesor debería tener información sobre el nivel en que un estudiante comprende las reglas y principios asociados a la capacidad medida por el TAI (ejecución de la transferencia).</p> <p>16. El profesor debería tener información sobre el nivel en que la capacidad adquirida resulta un aprendizaje transversal (extensión de la transferencia).</p>
<i>Diagnóstico de habilidades</i>	<p>17. Los parámetros de los ítems deberían apoyarse en la estructura cognitiva exigida a los individuos en cada uno de los reactivos.</p> <p>18. Las estimaciones de la TRI deberían incorporar medidas sobre las habilidades cognitivas de los sujetos (tiempo de reacción, tiempo de recuperación de la información, etc.).</p> <p>19. Las estrategias de selección de los ítems deberían tener en cuenta el tipo de aprendizaje/conocimiento objeto de medida.</p> <p>20. Para seleccionar un ítem en un TAI debería considerarse el nivel de competencia que exige al sujeto, en función del tipo de aprendizaje o conocimiento requerido.</p>

Por último, en el cuadro 7.11, Hontangas (1999) presenta una revisión de los programas informáticos existentes para la evaluación junto con sus características más generales.

Cuadro 7.11. Características generales de los programas para la evaluación informatizada. (Adaptado de Hontangas, 1999, y Jornet, 2001)

Programa	Sistema	Construcción			Administración					Análisis
		Ítem	Banco	Test	TC	TCI	TAI	TRL	TIN	
ADTEST (Específico para la aplicación de TAI; Ponsada, Olea y Revuelta, 1994)	DOS						✓			
Álgebra WG (Especializado en la generación automática de ítems; Biergaard, 1990)	DOS	✓		✓	✓					
BANKIT (Prestaciones suficientes; Molina, 1997)	MAC	✓	✓	✓	✓					2
CAT Soft.System (Paquete completo, altas prestaciones; Computer Adaptative Technologies, 1994)	TODOS	✓	✓	✓	✓	✓	✓	✓	✓	3
CONTEST (Especializado en Tests Óptimos; Timminga, Van der Linden y Scheweider, 1996)	DOS	✓	✓	✓	✓		✓			
C-Quest (Paquete completo, altas prestaciones; Cogent Computing Corporation, 1997)	Windows	✓	✓	✓	✓	✓		✓	✓	1
DEMOTAC (Específico para la aplicación de TAI; Renom, 1993)	DOS						✓			
ESATEST (Paquete completo, altas prestaciones; Engineering Software Associate, 1998)	Windows	✓	✓	✓	✓	✓		✓	✓	1
FASTEST (Paquete con altas prestaciones; Assessment; Systems Corporation, 1997)	Windows	✓	✓	✓	✓	✓		✓		
Interactive Assesor (Prestaciones suficientes; EQL International, 1996)	Windows	✓	✓	✓		✓		✓		
JBC (Especializado en la administración por internet; Holmes, 1997)	Windows	✓	✓	✓					✓	
LXRTest (Prestaciones suficientes; Logix eXtension Resources, 1994)	Windows MAC	✓	✓	✓	✓	✓				
MicroCAT (Paquete completo, altas prestaciones; Assessment Systems Corporation, 1995)	DOS	✓	✓	✓	✓	✓	✓	✓		3
MicroTest Pro (Paquete completo, altas prestaciones; Chariot Software Group, 1998)	Windows MAC	✓	✓	✓	✓	✓		✓	✓	

Cuadro 7.11. (continuación)

Programa	Sistema	Construcción			Administración					Análisis
		Ítem	Banco	Test	TC	TCI	TAI	TRL	TIN	
Perception (Paquete completo, altas prestaciones; Question Mark Corporation, 1998)	Windows	✓	✓	✓			✓	✓	✓	1
Quiz Factory (Prestaciones suficientes; Learning Ware, 1997)	Windows MAC	✓	✓	✓	✓	✓		✓		
TECA (Realiza el cálculo de índices estadísticos y métricos; Sanmartín y Ruiz, 1986)	MAC			✓						2
Test Builder (Prestaciones suficientes; TopGrade Software, 1997)	Windows	✓	✓	✓	✓	✓				
Test Generator (Prestaciones suficientes; Fain and Company, 1997)	Windows	✓	✓	✓	✓	✓		✓		
Test Genie (Especializado en GAI; Sleek Software Corporation, 1997)	DOS	✓		✓	✓					
Test Maestro (Especializado en la gestión de bancos de ítems; Stevens, 1998)	Windows	✓	✓	✓	✓					
The Examiner (Paquete completo, altas prestaciones; Examiner Corporation, 1998)	Windows	✓	✓	✓	✓	✓		✓		1
UCAT (Estima parámetros de los ítems según modelos TRI; Linacre, 1987)	DOS							✓		3
WinAsk Pro (Prestaciones suficientes; SmartLite Software, 1997)	Windows	✓	✓	✓	✓	✓				
WinWTE (Prestaciones suficientes; WICAT, 1997)	Windows	✓	✓	✓	✓	✓				

CONSTRUCCIÓN:

- **Ítem:** Construcción de ítems, bien manualmente o generación automática.
- **Banco:** Elaboración de bancos de ítems y gestión de éstos.
- **Test:** Elaboración de los tests por selección manual o automática de los ítems.

ADMINISTRACIÓN:

- **TC:** Tests convencionales de lápiz y papel.
- **TI:** Tests convencionales informatizados y administrados por ordenador.
- **TAI:** Tests Adaptativos Informatizados.
- **TRL:** Tests administrados por red local.
- **TIN:** Tests administrados por la red internet.

ANÁLISIS: Incluye cálculo de índices estadísticos, métricos y elaboración de informes.

1. Ofrece estadísticos descriptivos básicos de ítems y sujetos, y algunos, la posibilidad de gráficos, tablas e informes.
2. Calculan índices métricos con dificultad del ítem, discriminación del ítem y fiabilidad del test.
3. Permiten realizar análisis complejos.

Preguntas y ejercicios

1. Cuando se va a realizar una evaluación, ¿en qué nos basamos para decidir la utilización de una u otra técnica de recogida de datos?
2. Elabora el esquema de los distintos tipos de observación y registro de la información.
3. Realiza el esquema de los distintos tipos de entrevista.
4. ¿Qué ventajas y limitaciones tiene la observación frente a la entrevista?
5. Señala las principales ventajas y limitaciones del cuestionario como procedimiento para la recogida de información.
6. Teniendo en cuenta las normas que aparecen en el cuadro 7.3 para la confección de una prueba de ensayo, prepara 3 o 4 ítems y señala las normas para su corrección.
7. ¿Qué ventajas tienen los tests criterioles frente a los tests normativos en la evaluación educativa?
8. ¿Qué importancia tiene la calidad de los datos en el proceso de evaluación?
9. ¿Cuáles son los criterios de calidad de la información cuantitativa? ¿Y los de la información cualitativa?

Notas

1. «Una actitud es una disposición que condiciona al sujeto a percibir y a reaccionar de un modo determinado» (Del Rincón *et al.*, 1995, p. 179).
2. O de la evaluación.

8. Procedimientos para el análisis de la información

En el proceso lógico —aunque como se verá más adelante no lineal en todos los casos— tras la recogida de la información pertinente, ésta deberá ser analizada para llegar a conclusiones que contribuyan a emitir los juicios de valor que a su vez ayudarán a tomar las decisiones más adecuadas para de esta manera cerrar el ciclo evaluativo.

El primer paso en el análisis de la información recogida será almacenarla utilizando algún sistema simple informatizado (un procesador de textos, una base de datos o una hoja de cálculo) que posibilite un uso polivalente de los datos, de tal forma que si en el futuro queremos emplear una técnica de análisis distinta, no debamos volver a introducir los datos en el ordenador (Tójar, 2001). Este almacenamiento debe posibilitar la confidencialidad de los datos personales salvaguardando así la intimidad de las personas involucradas en la evaluación.

Analizar los datos supone llegar a conclusiones que sean razonables y a generalizaciones que estén basadas en esos datos. Worthen y Sanders (1987, p. 328) afirman que «el propósito del análisis de datos es reducir y sintetizar la información para que “tome sentido” fuera de sí y pueda inferirse a la población». En la misma línea, Mayor Ruiz (1998, p. 376) dice que el análisis de datos «es un proceso a través del cual los datos brutos se convierten en datos manejables para ser interpretados y valorados». Lizasoain y Joaristi (2000, p. 358), refiriéndose concretamente al análisis de datos en la evaluación de programas, aseveran que analizar los datos supone «proporcionar evidencia suficiente o respuesta a las hipótesis y preguntas planteadas mediante

técnicas analíticas (habitualmente, estadísticas en el paradigma cuantitativo) transformando los datos en resultados». Por último, Gil y Perera (2001, p. 13) refiriéndose al análisis de datos cualitativos, pero igualmente válido para datos cuantitativos, señalan que el análisis de datos es el «examen e interpretación de informaciones diversas».

Como ya se ha señalado previamente, los datos recogidos pueden ser cuantitativos o cualitativos, aunque ello no presupone los análisis que se vayan a realizar después. Es decir, caben cuatro posibilidades, tal y como se muestra en el cuadro 8.1.

Cuadro 8.1. Tipos de análisis en función de la naturaleza de los datos

	Datos cuantitativos	Datos cualitativos
<i>Análisis cuantitativos</i>	Análisis cuantitativos de datos cuantitativos (estadística) (1)	Análisis cuantitativos de datos cualitativos (estadística) (2)
<i>Análisis cualitativos</i>	Análisis cualitativos de datos cuantitativos (interpretación de índices, frecuencias...) (3)	Análisis cualitativos de datos cualitativos (análisis comprensivo...) (4)

Los casos 1 y 4 tienen un evidente campo epistemológico de aplicación claramente definido, mientras que el 2 y el 3 pueden inducir a errores. En el caso 2, se ha demostrado que es posible analizar de manera cuantitativa datos categoriales o cualitativos, y en el caso 3, dentro de un interés interpretativo, pueden analizarse datos cuantitativos de forma cualitativa (Tójar, 2001).

Hechas estas precisiones terminológicas vamos a presentar algunas características inherentes al análisis de datos tanto cualitativos como cuantitativos en la evaluación educativa (Lizasoain y Joaristi, 2000):

1. Como ya se vio en el capítulo 1, la finalidad de la evaluación no es la producción de conocimiento, sino el empleo del conocimiento para la acción, el cambio y la mejora. De ahí que los análisis que se realicen deben estar claramente orientados a proporcionar respuesta y evidencia a las preguntas previamente planteadas. Cuando nos referimos a los análisis estadísticos en concreto no basta con la significación estadística, sino que ha de tener **significación sustantiva**, en palabras de Botella y Barriopedro (1995, p. 198), o **significación práctica**, en palabras de Bickman y Rog (1998, p. XI).
2. La evaluación de programas se caracteriza por la **complementariedad**, de forma que el análisis de datos debe llevarse a cabo utilizando de forma complementaria o simultánea múltiples enfoques, métodos y técnicas.
3. La decisión de las técnicas a utilizar en el análisis de datos va a **estar condicionada por las decisiones que previamente se han tomado**

en el proceso evaluativo: el enfoque cuantitativo, cualitativo o mixto dado a la evaluación, las preguntas o cuestiones planteadas, diseño utilizado, que tipo de información y mediante qué instrumento se ha recabado, etc.

4. Si bien tras el planteamiento de las preguntas de evaluación vendrá la recogida de información y tras ésta el análisis de los datos para dar respuesta a las preguntas, este proceso no debe ser tomado al pie de la letra, ni siquiera en una evaluación cuantitativa.

1. Análisis cuantitativos

Los análisis cuantitativos hacen referencia a análisis estadísticos de variables, tanto cuantitativas como cualitativas. En este sentido, las técnicas estadísticas han tenido una gran difusión y uso en la investigación social desde que Galton lo hiciera por primera vez a finales del siglo XIX. La extrapolación del uso de la estadística del campo de la investigación social al campo de la evaluación educativa lleva implícita, como hemos mencionado previamente, la pérdida de significación de la significatividad estadística en favor de la significación sustantiva o práctica. Además, las conclusiones y estimaciones obtenidas tras la aplicación de las técnicas estadísticas no podrá ir acompañada de una completa certeza, sino con una confianza limitada (Botella y Barriopedro, 1995).

No obstante, el uso de la estadística en la evaluación educativa no ha decaído y su validez está de sobra demostrada sobre todo en ciertas situaciones en las que la información recogida es cuantitativa debido a que las cuestiones o preguntas de la evaluación así lo exigen.

Botella y Barriopedro (1995) señalan las cuatro funciones principales del análisis de datos en la valoración de los resultados de un programa:

- a) Describir de forma comprensiva datos brutos.
- b) Ayudar a decidir si alguna relación entre los datos puede admitirse con confianza.
- c) Estimar magnitudes; especialmente, cantidades de cambio o diferencias.
- d) Determinar si ciertos efectos que aparecen en la información alcanzan un nivel relevante en el contexto del programa.

La primera de las funciones mencionadas se resuelve mediante la estadística descriptiva (gráficos, porcentajes, índices y coeficientes, etc.), mientras que las tres restantes exigen el uso de la estadística inferencial (estudios de relaciones mediante técnicas bivariadas como la correlación y el chi cuadrado o multivariadas como los *clusters*, análisis discriminante, análisis causal, análisis factorial, modelos jerárquico-lineales, regresión, etc.).

Botella y Barriopedro presentan una guía orientativa de los pasos que podrían seguirse en el análisis estadístico de los datos (véase cuadro 8.2).

**Cuadro 8.2. Pasos en el análisis cuantitativo de los datos.
(Tomado de Botella y Barriopedro, 1995, pp. 202-203)**

- a) Establecer el sistema de codificación de las variables cualitativas y ordinales y la métrica de las cuantitativas.
 1. Obtener la fiabilidad entre codificadores con submuestras aleatorias de los datos.
 - b) Preparar la matriz de unidades por variables.
 - c) Obtener un primer análisis descriptivo de cada variable y de combinaciones seleccionadas de dos variables, consiguiendo una primera versión de los instrumentos descriptivos que hemos detallado.
 1. Detectar valores o combinaciones de valores imposibles.
 2. Identificar valores atípicos y decidir sobre ellos.
 3. Recodificar mediante agregados o subdivisiones de categorías o crear nuevas variables cuantitativas mediante la suma o ponderación de otras variables.
 4. Volver a obtener los instrumentos descriptivos tras las eventuales modificaciones derivadas de los puntos anteriores.
 - d) Analizar la presencia y aleatoriedad de la distribución de valores ausentes mediante las técnicas apropiadas.
 - e) Aplicar técnicas inferenciales para determinar la significación de los efectos aparentes en los datos.
 - f) Estimar la magnitud del impacto en los distintos indicadores.
 - g) Aplicar técnicas inferenciales para decidir sobre la consecución de las magnitudes criterios establecidas como objetivo del programa.
 - h) Aplicar técnicas meta-analíticas, cuando proceda, para mostrar y comparar tamaños del efecto diferencial.
-

Por último, estos mismos autores señalan cuatro principios que debieran presidir el proceso de análisis estadístico de los datos. A saber:

1. El análisis de datos no debe reducirse a una aplicación mecánica de las técnicas estadísticas reflejadas en el diseño, sino que debe ser un proceso más flexible e interactivo. Nunca deben olvidarse las preguntas que guiaron tanto el diseño como la recogida de la información.
2. En caso de duda entre dos técnicas de análisis deben utilizarse ambas, ya que en caso de arrojar resultados semejantes la duda estará disipada.
3. Aunque es conveniente explotar los datos al máximo, una buena descripción preliminar es el mejor primer paso para que los análisis sean comprensibles, dado que en muchos casos los receptores de los informes no están preparados para entender las sutilezas de los análisis más sofisticados.

4. La imparcialidad con respecto a los resultados y por ende la credibilidad aumentará si se asume el compromiso de informar de todas las limitaciones del análisis de datos (pérdida de sujetos, etc.).

2. Análisis cualitativos

En función de los objetivos de la evaluación, el evaluador ha podido decantarse por la utilización de instrumentos de recogida de datos que ofrecen información cualitativa. Éste es el caso, por ejemplo, de la utilización de la observación participante, las entrevistas semiestructuradas, el análisis de documentos, etc. Al igual que sucede con los análisis cuantitativos, si pretendemos realizar análisis cualitativos de esa información, se deberán establecer en todo el proceso indagativo mecanismos que garanticen el rigor y la transparencia para asegurar la validez o credibilidad de los resultados obtenidos.

Antes de explicar el procedimiento para llevar a cabo el análisis cualitativo debemos tener en cuenta una serie de principios que caracterizan la evaluación cualitativa (Gil y Perera, 2001, pp. 15-18):

- Los estudios cualitativos se ocupan de procesos antes que de productos.
- Los estudios cualitativos estudian objetos acotados en el espacio.
- Los estudios cualitativos adoptan una visión holística de la realidad.
- En los estudios cualitativos se pretende penetrar en la subjetividad de los sujetos y de los grupos.
- En los estudios cualitativos se sigue una vía inductiva en la construcción del conocimiento.
- En los estudios cualitativos, el evaluador trata de no alterar la realidad; la estudia y evalúa en el modo natural en que se manifiesta.
- En los estudios cualitativos se subraya el papel del investigador (o evaluador) como instrumento de investigación (o evaluación).

Como consecuencia de estos principios y debido también al propio proceso de la evaluación cualitativa, hay una serie de aspectos relativos al evaluador cualitativo y a los datos a tener en cuenta.

En primer lugar, mientras que las técnicas estadísticas permiten el análisis de los datos cuantitativos incluso de manera automatizada, los datos y los análisis cualitativos exigen que sea el evaluador quien atribuya significados, elabore resultados y extraiga conclusiones. De tal forma que el evaluador cuantitativo que conozca la estadística tiene el camino marcado, mientras que el evaluador cualitativo no dispone de procedimientos claramente definidos que le indiquen el camino a seguir. Por ello, muchas veces cada evaluador crea su propia forma de analizar los datos y en demasiadas ocasiones se ha

declarado que el análisis de datos cualitativos tiene un gran componente artístico. De ahí, es decir, de esa supuesta subjetividad del evaluador, han llovido muchas de las críticas hacia la investigación y evaluación cualitativas.

En segundo lugar, el evaluador cualitativo se va a encontrar con datos que están expresados en lenguaje verbal en la mayoría de los casos. El registro de los datos puede hacerse de manera diversa, pero al final el evaluador cualitativo transcribirá esa información a textos escritos que serán con los que tendrá que trabajar posteriormente.

En tercer lugar, el evaluador se va a encontrar con el problema de la gran cantidad de información con la que va a tener que trabajar (transcripciones de entrevistas, notas de campo de observaciones...). Otro problema unido a esa voluminosidad va a ser la heterogeneidad de la información en parte debido al intento de captar holísticamente las situaciones estudiadas y en parte debido a la propia riqueza expresiva del lenguaje verbal. El último problema va a ser que al no estar la información expresada numéricamente, no es posible el tratamiento estadístico de ella.

En cuarto lugar, en la evaluación cualitativa el análisis de datos no es inmediatamente posterior a la recogida de la información. No debe entenderse como un proceso lineal, sino como un proceso cíclico, a veces simultáneo; en todo caso interactivo.

En quinto lugar, el análisis de datos cualitativos puede hacerse siguiendo un procedimiento más o menos común a la mayor parte de los análisis desarrollados en la investigación cualitativa. Es decir, se pueden extraer determinadas operaciones básicas que están presentes generalmente en los análisis cualitativos y, por tanto, pueden ser organizados en un esquema general de trabajo. Las tareas, trabajos u operaciones comunes serían los siguientes (Gil, 1994; Gil, García y Rodríguez, 1997; Cabrera, 2000a; Gil y Perera, 2001):

FASE 1: REDUCCIÓN DE LA INFORMACIÓN

Como se ha señalado previamente, uno de los problemas principales con los que se va a encontrar el evaluador va a ser la cantidad de información a analizar, y es por ello por lo que para superar ese obstáculo debe reducir la información. En esta primera fase podrían distinguirse las subfases siguientes:

- a) *Selección e identificación de los documentos a analizar.*
- b) *Determinación de los objetivos del análisis.*
- c) *Establecimiento del sistema de categorías.* La categorización supone transformar los datos brutos de una comunicación en un sistema de categorías que lo represente. Es decir, consiste en identificar el tema o categoría del significado a la que hace referencia determinado fragmento del texto escrito. Como señalan Gil y Perera (2001), las categorías pueden referirse a opiniones, actitudes, sentimientos, valoracio-

nes realizadas por los sujetos, actitudes, procesos, conductas, interacciones entre individuos, lugares, momentos, características de las personas, etc. La categorización no es más que simplificar la información a un número menor de clases o categorías. Para llevar a cabo la categorización, previamente se deberá decidir la selección de la unidad de registro que puede ir desde la unidad de registro más pequeña, que es la palabra, a la unidad de registro más grande, que es el propio documento. Normalmente, la unidad de registro utilizada suele ser el tema que se refiere a «un fragmento cualquiera de la comunicación que tenga un sentido completo para los objetivos del análisis» (Cabrera, 2000a, p. 226). Posteriormente, deberemos establecer el sistema de categorías para poder clasificar las unidades de registro. Este sistema puede ser determinado a priori (partiendo de la teoría previa, de los utilizados por otros autores, etc.) o a posteriori (utilizando una vía puramente inductiva). Normalmente suelen utilizarse procedimientos mixtos; es decir, se parte de un sistema de categorías abierto y flexible que se irá acomodando y adaptando a la información recogida. Cabrera (2000a) señala las características exigibles a un sistema de categorías: adecuadas y pertinentes, homogéneas, exhaustivas, objetivas, exclusivas. Por último, se deberá asignar un código (numérico o verbal) para cada categoría identificada.

- d) *Estudio piloto del sistema de categorías.* Se trata de comprobar mediante un pequeño estudio la fiabilidad y la validez del sistema de categorías. Para el estudio de dicha fiabilidad, se deberá asegurar que en caso de más de un analista, haya unanimidad en la identificación de las unidades de registro y en la asignación del código correspondiente. Para ello (Cabrera, 2000a), propone la fórmula siguiente en la que la fiabilidad quedará asegurada si se obtienen puntuaciones superiores al 85%. La validez del sistema de categorías deberá ser comprobada con el fin de asegurar su validez a los objetivos del análisis.

$$\% \text{ Acuerdo} = \frac{(a \times 100)}{b}$$

Donde, a es el número de unidades codificadas idénticamente, y b es el número de unidades codificadas.

- e) *Codificación.* Mediante la codificación asignamos a cada fragmento del texto un código o etiqueta que representa a una categoría para señalar la pertinencia de dicho fragmento a esa categoría.
- f) *Desecho de datos irrelevantes.* En todo estudio cualitativo se producen informaciones irrelevantes para sus objetivos. Esta información debe ser desechada de tal forma que se contribuya también de esta manera a la reducción de los datos.

FASE 2: ORGANIZACIÓN Y PRESENTACIÓN DE LA INFORMACIÓN

Tras la categorización y codificación de la información, el paso siguiente consistirá en agrupar la información de tal manera que facilite su análisis y su interpretación. La forma más sencilla de agrupar los datos es la presentación agrupada de todos los fragmentos del texto que hacen alusión a una misma categoría. De esta manera se rompe la estructura lineal del texto y se reorganiza en función del contenido temático posibilitando llegar a conclusiones sobre el contenido de la información de cada categoría (Gil y Perera, 2001).

Otra forma interesante de reorganizar la información es mediante la utilización de matrices textuales (Gil y Perera, 2001). Una matriz textual es una tabla de doble entrada en la que las filas las constituyen las diferentes categorías, y las columnas, las diferentes personas, escenarios, fuentes de datos, etc. El cruce de cada fila y columna recoge la información correspondiente a la intersección entre ambas. Mediante estas matrices se puede llegar a conclusiones, bien a través de la comparación entre filas, o bien a través de la comparación entre columnas.

Además de las dos formas de organizar y presentar la información señaladas, caben otras posibilidades como los diagramas, los mapas conceptuales, etc.

FASE 3: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

Esta tercera y última fase hace referencia a la extracción de conclusiones que constituyen en definitiva los resultados finales del estudio (véase cuadro 8.3).

La extracción de conclusiones es quizás el aspecto más oscuro del proceso de análisis. No obstante, para clarificar este proceso señalan una serie de aspectos a tener en cuenta:

- a) Las conclusiones comienzan a tomar forma en la mente del evaluador desde el inicio mismo de la recogida de la información. Tanto al realizar observaciones como entrevistas, al evaluador le surgen sospechas, intuiciones, etc., de carácter provisional que deberá ir focalizando en las recogidas de datos posteriores para ir confirmando o rechazándolas.
- b) Los sistemas de categorías obtenidos a posteriori, es decir, tras un proceso de codificación inductivo, suponen también conclusiones del análisis, puesto que recogen el conjunto de significados presentes en los datos. La descripción de las categorías, su definición, su delimitación frente a otras son procesos que también van a generar resultados.
- c) La comparación será otro procedimiento importante para extraer conclusiones:
 - Comparaciones entre filas y columnas si hemos utilizado matrices textuales.
 - Contrastar los hallazgos con el posible marco teórico de partida.
 - Contrastar nuestros resultados con otros trabajos similares.

Cuadro 8.3. Pasos en el análisis cualitativo de los datos

FASE 1: REDUCCIÓN DE LA INFORMACIÓN.

- A. Selección e identificación de la información a analizar.
- B. Determinación de los objetivos del análisis.
- C. Establecimiento de un sistema de categorías.
 - Selección de la unidad de registro.
 - Establecer el sistema de categorías.
 - Identificar códigos para cada categoría.
- D. Estudio piloto del sistema de categorías.
- E. Codificación.
- F. Desechar datos irrelevantes.

FASE 2: ORGANIZACIÓN Y PRESENTACIÓN DE LA INFORMACIÓN.

FASE 3: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS.

Por último, para terminar con la tercera fase del análisis cualitativo de datos, el evaluador debe asegurarse de convencer a las audiencias de la validez de sus conclusiones. Para ello, las estrategias más utilizadas son:

1. *La triangulación*, que consiste en llegar a los mismos resultados por diferentes caminos. Algunos autores incluyen la triangulación como una técnica más de recogida de información, ya que tras ella se obtiene nueva información; sin embargo, lo más usual suele ser considerarla dentro del análisis de la información. La triangulación puede llevarse a cabo a nivel de técnicas de recogida de datos (tanto de datos del cuestionario como los de la observación, como los de las entrevistas son coincidentes, por ejemplo), de fuentes (profesores, alumnos, equipo directivo, por ejemplo), de evaluadores (diferentes personas que analizan los mismos datos llegan a las mismas conclusiones, por ejemplo), etcétera.
2. *El contraste de nuestras conclusiones y su posterior aprobación por parte de los participantes en la evaluación*. Si mediante la evaluación cualitativa pretendemos identificar las perspectivas de los sujetos evaluados, habrá que contrastar con ellos si realmente nuestro propósito se ha cumplido. Para llevar adelante este proceso, se envía un informe preliminar a los evaluados para que en una posible reunión posterior (evaluadores-evaluados) muestren sus acuerdos, matizaciones o discrepancias.
3. *La presencia prolongada en el campo* es también una garantía de comprensión de la realidad.

3. Recursos informáticos para el análisis de la información

El campo en el que sin duda se ha dado un mayor avance en relación con las aplicaciones informáticas en evaluación es el relativo al análisis de la información. Como señala Jornet (2001), este avance ha afectado al desarrollo de la ciencia en general en cuanto que incrementa la precisión y la rapidez en la investigación. Aplicaciones informáticas de uso casi exclusivo a la investigación están hoy en día disponibles para el análisis de datos en la evaluación educativa.

Además, estos programas informáticos que han ido desarrollándose se han ido perfeccionando de tal manera que además de mejorar los procesos de gestión y análisis de la información se han constituido bajo entornos (como el Windows, por ejemplo) que han conseguido facilitar en gran medida su utilización. Es decir, los conocimientos informáticos necesarios para su uso son mínimos, y ello amplía las posibilidades de su empleo.

No obstante, se debe aclarar que el uso de estas aplicaciones informáticas en el análisis de datos no constituye per se una garantía de rigor y validez del proceso evaluativo o de credibilidad de los resultados obtenidos. Las aplicaciones informáticas facilitan la labor del evaluador como herramientas de trabajo que son, pero no garantizan la calidad de la evaluación (Gil y Perera, 2001).

El avance de la informática vinculada al análisis de datos en evaluación se ha dado en tres ámbitos:

- a) Aplicaciones para el análisis cuantitativo de datos que sin duda es la que ofrece mayores posibilidades dada por una parte su tradición y, por otra, la facilidad de informatización que suponen estos procesos.
- b) Aplicaciones para el análisis cualitativo de datos que, aunque no tiene el desarrollo de las aplicaciones anteriores, sí ha experimentado un gran avance en los últimos años.
- c) Aplicaciones de ayuda para la construcción de instrumentos.

Vamos a ver las aplicaciones asociadas a cada uno de estos tres ámbitos de manera más detenida.

3.1. Aplicaciones para el análisis cuantitativo de datos

Las notas características de las aplicaciones informáticas para el análisis cuantitativo de la información evaluativa serían las siguientes (Jornet, 2001):

- Las aplicaciones informáticas para el análisis cuantitativo es el área de mayor desarrollo en cuanto a aplicaciones disponibles.
- Tanto para la síntesis de la información como para el estudio de la información, la evaluación necesita de aplicaciones informáticas para realizar el análisis de datos.

8. Procedimientos para el análisis de la información

- El concepto de paquete estadístico (que integra un amplio abanico de opciones de gestión y análisis de datos) facilita enormemente su utilización; es decir, los conocimientos informáticos necesarios para su uso son mínimos.
- La no necesidad de grandes conocimientos informáticos debido a la facilidad de su manejo está relacionada con la unificación de los interfaces de usuario con el de mayor difusión que no es otro que el Windows de Microsoft (los paquetes estadísticos SPSS, BMDP o SYSTAT utilizan el mencionado sistema).
- De manera simultánea, se han ido creando programas específicos para el análisis cuantitativo de datos procedentes de instrumentos de recogida de información variados (The Observer, CAMERA y VTLOGANL para datos provenientes de la observación; RepGrid para datos provenientes de entrevistas, etc.).
- La mayor parte de los paquetes estadísticos estandarizados tienen la ventaja de estar en entorno Windows con lo que la traslación de los resultados a programas de tratamiento de textos (como el Microsoft Word o el Corel Word Perfect) o a hojas de cálculo (como Excel) es muy factible, facilitándose la comunicación enormemente.

En el cuadro 8.4 pueden apreciarse los recursos informáticos más usuales para el análisis cuantitativo de la información.

Cuadro 8.4. Recursos informáticos usuales para el análisis cuantitativo de la información. (Tomado de Jornet, 2001, p. 281)

Tipo/Ámbito de uso	Recursos tecnológicos
<i>Paquetes estadísticos estandarizados</i>	<ul style="list-style-type: none"> • BMDP (Dixon, 1983, 1990) • SAS (SAS Institute, 1988) • SPAD.N y SPAD.T (Lebart <i>et al.</i>, 1994a, 1994b) • SPSS (Norusis y SPSS Inc., 1990, 1994) • STATVIEW (Martín Tabernero y Delgado, 1993) • SYSTAT (Ato y López, 1994)
<i>Hojas de cálculo</i>	<ul style="list-style-type: none"> • EXCEL • LOTUS • WORKS
<i>Gráficos</i>	<ul style="list-style-type: none"> • Harvard Graphics
<i>Observación</i>	<ul style="list-style-type: none"> • CAMERA (Vandel Vlugt <i>et al.</i>, 1992) • The Observer (Noldus, 1991) • VTLOGANL (Roberts <i>et al.</i>, 1993)
<i>Entrevista</i>	<ul style="list-style-type: none"> • RepGrid 2.0 (Freixas <i>et al.</i>, 1993)
<i>Cuestionarios/ encuestas</i>	<ul style="list-style-type: none"> • BARBRO • EAF (Jornet, Suárez y Belloch, 1998)

3.2. Aplicaciones para el análisis cualitativo de datos

Las aplicaciones para el análisis cualitativo de datos han tenido un gran desarrollo en estos últimos veinte años, de tal manera que han venido a facilitar los procesos de recogida y de análisis de información textual haciendo más manejables y asequibles las tareas implicadas en ellos.

Sin embargo, hay una serie de características, en algunos casos diferenciadoras de las aplicaciones informáticas para el análisis de datos cuantitativos, que queremos reseñar (Gil y Perera, 2001):

- En el análisis cualitativo de datos textuales deben distinguirse las tareas conceptuales (las que implican decisiones e intervenciones del evaluador para llegar a resultados) de las tareas mecánicas (las que conllevan manipular físicamente los datos y los resultados de sus análisis). Las aplicaciones informáticas únicamente facilitan la labor de las tareas mecánicas no la de las conceptuales que será el evaluador quien las deberá llevar a cabo. Cuando no se disponía (o no se dispone) de aplicación informática, el analista utilizaba estrategias tales como usar lápices de colores para subrayar todos los fragmentos pertenecientes a una misma categoría o fotocopiaba y cortaba todos los fragmentos de una misma categoría. Hoy en día, todo este tipo de tareas las ejecuta el ordenador ya que contamos en la actualidad con programas diseñados específicamente para ayudar al analista en este tipo de trabajos.
- Las aplicaciones estadísticas permiten introducir matrices de datos en el ordenador y tras introducir las instrucciones pertinentes, pedirle que nos haga los análisis obteniendo en pocos segundos todos los resultados. Con los análisis cualitativos, no obstante, el ordenador no proporciona al evaluador los resultados del análisis como respuesta a la introducción de información textual, simplemente le auxilia en la tarea de realizar los análisis. El ordenador, pues, ejecutará las tareas mecánicas, pero las conceptuales las tendrá que realizar el evaluador.
- En los apartados anteriores nos hemos referido al análisis cualitativo de datos textuales. Sin embargo, también se han desarrollado programas específicos para analizar cuantitativamente datos cualitativos. En éstos, se cuantifica la información cualitativa mediante recuentos, se construyen matrices de datos cuantitativas y, posteriormente, pueden aplicarse técnicas estadísticas.

En los cuadros 8.5 y 8.6 se muestran algunos programas informáticos para el análisis cualitativo de datos textuales y para el análisis cuantitativo de datos textuales.

Cuadro 8.5. Programas para el análisis de datos cualitativos.
(Adaptado de González Such, 2000)

Nombre	Autor	Entorno
Aquad Five 5.7	Günter L. Huber	Windows
ATLAS,ti	Thomas Muhr	Windows
Code-A-Text 6.2	Alan Cartwright	Windows
CISAID , Code-A-Text Integrated System for the Analysis of Interviews and Dialogues	Alan Cartwright	Windows
EZ-TEXT 3.06C	Conwal Incorporated	Windows
Hyperqual 3.0 versión 1.1	Raymond V. Padilla	MAC
HyperResearch 2.0	S. Hesse-Biber, P. Dupuis, T.S. Kinder	Windows MAC
Kwalitan 5.0	V. Peters	Windows DOS
Martin 2.0	R. Schuster	Windows
MECA , Map Extraction Comparision and Analysis	K.M. Carley	DOS Unix
QED 1.1.005 , Qualitative Enhanced Data-analysis	M. Berg, Erasmus University Rotterdam	Windows
QMA , Qualitative Media Analysis	Carl Verner Skou	Windows
QSR NUD.IST , Non-numerical Unstructured Data Indexing Searching and Theorizing	Lyn and Tom Richards	Windows
QSR Nvivo 1.1	Lyn and Tom Richards	Windows
Tatoe 0.987 , Text Analysis Tool with Object Encoding	Lothar Rostek, Melina Alexa	Windows
Textbase alpha	Bo Sommerlund	DOS
Textbase beta	Bo Sommerlund	DOS
The Ethnograph 5.04	J.V. Seidel	Windows

Cuadro 8.6. Programas para el análisis cuantitativo de datos cualitativos. (Adaptado de González Such, 2000)

Nombre	Autor	Entorno
DICTION 5.0	Roderick F. Hart	Windows
Hamlet	Alan Brier	DOS Windows
SALT 6.1 , Systematic Analysis of Language Transcripts	R.S. Chapman, H.F. Miller	Windows MAC
SPAD-T	CISIA-CERESTA	Windows
SWIFT 3.1 , Structured Word Information Frequency Tables	R.B. Heady	DOS
TEXTPACK 7.0 , TextPACKage	Zuma Mannheim	Windows

3.3. Aplicaciones de ayuda para la construcción de instrumentos

Un campo específico de las aplicaciones informáticas en el análisis de datos de la evaluación educativa ha sido el desarrollado para la ayuda en la construcción y el análisis de pruebas objetivas y tests estandarizados (sean de capacidad, personalidad o rendimiento, tanto para tests criteriosales como sobre todo para tests normativos). En este sentido se han dado dos líneas de desarrollo (Jornet, 2001):

- El desarrollo de paquetes específicos para este menester, como, por ejemplo, el programa ITEMAN.
- La incorporación de recursos en paquetes estadísticos estandarizados ya existentes, como, por ejemplo, la posibilidad de calcular la fiabilidad que introdujo el SPSS.

Como señala Jornet (2001), estos recursos disponibles son interesantes, ya que además de posibilitar la mejora en la construcción de pruebas, acercan este tipo de recursos a los profesionales del ámbito aplicado.

En los cuadros 8.7 y 8.8 se presentan programas para el análisis de pruebas objetivas y tests estandarizados y los principales programas informáticos para el análisis según la teoría de respuesta a los ítems (TRI).

Cuadro 8.7. Programas para el análisis de tests. (Adaptado de González Such, 2000)

Programa	Autor	Finalidad	Resultados
<i>BMDP</i>	BMDP, 1990	<ul style="list-style-type: none"> • Fiabilidad. • Validez. • Transformación de las puntuaciones. 	Teoría clásica de los tests (TCT). Teoría de la generalizabilidad (TG). Regresión (simple, múltiple, logística), correlación, análisis factorial exploratorio.
<i>DEMOCAT</i>	Renom y Martínez Bazán, 1993	Elaboración de ítems, aplicación y análisis.	Teoría de respuesta a los ítems (TRI).
<i>EQUATE</i>	Baker, 1993	Equiparación de puntuaciones.	Cualquier problema de equiparación.
<i>GENOVA</i>	Crick y Brennan, 1982, 1983, 1992	Fiabilidad.	TG hasta cinco facetas.
<i>ITEMS</i>	García Cueto, 1991	Fiabilidad.	Guttman, Alpha para dicotómicos. Descriptivos del test.
<i>MicroCAT</i>	Assessment Systems Corporation, 1987, 1994	<ul style="list-style-type: none"> • Construcción del test, aplicación y análisis del mismo. • Fiabilidad como consistencia interna. • Transformación de puntuaciones. 	TCT. TRI. Likert.
<i>PSIEL</i>	Rodríguez Soler y García Cueto, 1993	<ul style="list-style-type: none"> • Fiabilidad. • Transformación de las puntuaciones. • Análisis de ítems. 	Guttman y Alfa para cualquier ítem. Likert. Descriptivos del test. Bondad de ajuste.
<i>SAS</i>	Sas Institute, 1988	<ul style="list-style-type: none"> • Fiabilidad. • Validez. • Análisis de ítems. • Transformación de las puntuaciones. 	TCT. Análisis factorial. Correlación. Regresión.
<i>SPSS</i>	Norusis, 1993	<ul style="list-style-type: none"> • Fiabilidad. • Validez. • Análisis de ítems. • Transformación de las puntuaciones. 	TCT, TG. Regresión (simple, múltiple, logística), correlación, análisis factorial exploratorio. Descriptivos del test, correlación ítem-total, Alfa sin el ítem.
<i>SYSTAT</i>	Wilkinson, 1986	<ul style="list-style-type: none"> • Fiabilidad. • Validez. • Análisis de ítems. • Transformación de las puntuaciones. 	TCT, TRI, TG. Regresión (simple, múltiple, logística), correlación, análisis factorial exploratorio.

Cuadro 8.8. Principales programas para TRI. (Adaptado de Hambleton, Swaminathan y Rogers, 1991, pp. 48-50)

Programa	Autor	Método	Número de parámetros
<i>ANCILLES</i>	Urry, 1978	Heurístico	3 parámetros
<i>ASCAL</i>	Assessment Systems Corporation, 1986	Bayesiano	1, 2 y 3 parámetros
<i>BIG-SCALE</i>	Wright <i>et al.</i> , 1989	Máxima verosimilitud incondicional	1 parámetro
<i>BILOG</i>	Mislevy y Bock, 1984	Máxima verosimilitud marginal y bayesiano	1, 2 y 3 parámetros
<i>LOGIST</i>	Wingersky, 1983	Máxima verosimilitud incondicional	1, 2 y 3 parámetros
<i>MICRO-SCALE</i>	Mediex Interactive Technologies, 1986	Máxima verosimilitud incondicional	1 parámetro
<i>MIRTE</i>	Carlson, 1987	Máxima verosimilitud incondicional	1, 2 y 3 parámetros
<i>MULTILOG</i>	Thissen, 1986	Extensión del BILOG para ítems de elección múltiple	1, 2 y 3 parámetros
<i>NOHARM</i>	Fraser y McDonald, 1988	Mínimos cuadrados	1, 2 y 3 parámetros
<i>PML</i>	Gustafson, 1980	Máxima verosimilitud condicional	1 parámetro
<i>RASCAL</i>	Assessment Systems Corporation, 1986	Máxima verosimilitud incondicional	1 parámetro
<i>RIDA</i>	Glas, 1990	Máxima verosimilitud incondicional o marginal	1 parámetro

Preguntas y ejercicios

1. ¿Qué tipos de análisis pueden realizarse con los datos cuantitativos? ¿Y con los cualitativos?
2. ¿Cuáles son los pasos a seguir para efectuar los análisis cuantitativos de los datos?
3. ¿Qué consideraciones ha de tener en cuenta el evaluador cuando se enfrenta al análisis de información cualitativa?
4. ¿Cuáles son los pasos a seguir para realizar los análisis cualitativos de los datos?

9. Actividades prácticas

A continuación, se propone una batería de actividades que servirán para realizar prácticas relacionadas con la evaluación educativa (muchas de ellas también son válidas para el campo de la investigación educativa). Como hemos señalado en la introducción del libro, estas prácticas pretenden completar la formación teórica, realizando actividades en las que los estudiantes tengan que participar de manera activa. Las actividades prácticas propuestas se han clasificado atendiendo a los capítulos teóricos presentados a lo largo del libro. Como se puede comprobar, estas actividades van desde la realización de un trabajo monográfico o la composición de un mapa conceptual hasta el visionado de un vídeo o lectura de un artículo y su posterior reflexión y debate, pasando por la construcción de instrumentos de medida o el análisis de éstos.

En el cuadro 9.1 se presenta la clasificación de las actividades prácticas propuestas en función de los capítulos teóricos presentados previamente. No obstante, hemos de señalar que en algunos casos, una misma actividad práctica podría pertenecer a más de un capítulo, y en otros, la realización de una práctica implica necesariamente a más de un capítulo.

La mayoría de las prácticas están diseñadas para que sean realizadas en pequeños grupos (3 o 4 estudiantes), aunque también pueden llevarse a cabo individualmente. En algunos casos se propone más tarde la puesta en común en el grupo clase.

Cuadro 9.1. Batería de actividades prácticas de evaluación educativa

Capítulos	Actividades prácticas
CAPÍTULO 1: Naturaleza de la investigación y evaluación en educación.	Práctica 1 Práctica 2 Práctica 3 Práctica 4 Práctica 5 Práctica 6
CAPÍTULO 2: Historia de la evaluación educativa.	Práctica 1 Práctica 2
CAPÍTULO 3: Concepto, componentes y fases de la evaluación.	Práctica 1 Práctica 2 Práctica 3 Práctica 4 Práctica 5 Práctica 6 Práctica 7
CAPÍTULO 4: Evaluación de programas.	Práctica 1 Práctica 2 Práctica 3 Práctica 4 Práctica 5 Práctica 6
CAPÍTULO 5: Ámbitos de la evaluación educativa.	Práctica 1 Práctica 2 Práctica 3 Práctica 4 Práctica 5 Práctica 6 Práctica 7 Práctica 8
CAPÍTULO 6: Diseños de investigación aplicados a la evaluación.	Práctica 1 Práctica 2 Práctica 3 Práctica 4 Práctica 5 Práctica 6
CAPÍTULO 7: Procedimientos para la recogida de la información.	Práctica 1 Práctica 2 Práctica 3 Práctica 4 Práctica 5 Práctica 6 Práctica 7 Práctica 8
CAPÍTULO 8: Procedimientos para el análisis de la información.	Práctica 1 Práctica 2 Práctica 3 Práctica 4 Práctica 5 Práctica 6

CAPÍTULO 1: Naturaleza de la investigación y evaluación en educación

PRÁCTICA 1

Definición de conceptos básicos.

Definición de los siguientes conceptos utilizando tus propias palabras: *Ciencia, método científico, investigación, investigación educativa, paradigma, paradigma positivista, paradigma interpretativo, paradigma crítico, método cuantitativo, método cualitativo.*

PRÁCTICA 2

Role Playing sobre los paradigmas de investigación educativa.

Elección de uno de los paradigmas tradicionales para su posterior defensa. Preparación de la defensa del paradigma elegido (señalando los puntos fuertes) y las críticas a los otros dos paradigmas (señalando los puntos débiles). Posteriormente, puede ser defendido el paradigma elegido debatiendo con otras personas que hayan elegido paradigmas diferentes.

PRÁCTICA 3

Esbozo de un proyecto de investigación y evaluación.

Elección de un tema educativo que resulte de interés. Sobre ese tema se debe realizar un esbozo de proyecto de investigación básica. De la misma manera, sobre ese mismo tema habría que pensar en un programa educativo y la evaluación que se haría de él. De esta manera se irán viendo las semejanzas y las diferencias entre la investigación y la evaluación. Después, en el grupo clase, cada uno puede presentar el proyecto de investigación y el de evaluación y el resto señalar los aciertos y los aspectos criticables de cada proyecto.

PRÁCTICA 4

Análisis de un informe de evaluación

Analiza y describe la opción metodológica adoptada en el estudio del Ararteko (2006) *Convivencia y conflictos en los centros educativos*.

Dirección electrónica del informe: <http://www.ararteko.net/webs/iextras/conflictos-ceneduc2006/conflictos-ceneduc2006E.pdf>

- ¿Por qué se puede hablar de complementariedad en este estudio?
- En este estudio, ¿cuáles son los indicadores de que se ha utilizado metodología cuantitativa?

- En este estudio, ¿cuáles son los indicadores de que se ha utilizado metodología cualitativa?
- ¿Qué tipo de relación se establece entre las dos metodologías?
- ¿Qué ofrece cada tipo de resultados (cuantitativos cualitativos) al estudio?

PRÁCTICA 5

Análisis de un informe de evaluación

Analiza y describe la opción metodológica adoptada en el estudio de Lukas, J. F.; Santiago, K. y Lizasoain, L. (2005). *Vitoria-Gasteiz como ciudad educadora: Evaluación de los programas del Departamento Municipal de Educación*.

Dirección electrónica del informe: http://www.sc.ehu.es/plwllumuj/ebaluazio_pedagogia/praktikak/INFORME-vitoria.pdf

- ¿Por qué se puede hablar de complementariedad en este estudio?
- En este estudio, ¿cuáles son los indicadores de que se ha utilizado metodología cuantitativa?
- En este estudio, ¿cuáles son los indicadores de que se ha utilizado metodología cualitativa?
- ¿Qué tipo de relación se establece entre las dos metodologías?
- ¿Qué ofrece cada tipo de resultados (cuantitativos cualitativos) al estudio?

PRÁCTICA 6

Análisis de un informe de evaluación

Analiza y describe la opción metodológica adoptada en el estudio de Etxeberria, J.; Lukas, J. F. y Santiago, K. (2008). *IKASYS programaren aplikazio pilotuaren ebaluazioa*.

Dirección electrónica del informe: http://www.sc.ehu.es/plwllumuj/ebaluazio_pedagogia/praktikak/INFORME-vitoria.pdf

- ¿Por qué se puede hablar de complementariedad en este estudio?
- En este estudio, ¿cuáles son los indicadores de que se ha utilizado metodología cuantitativa?
- En este estudio, ¿cuáles son los indicadores de que se ha utilizado metodología cualitativa?
- ¿Qué tipo de relación se establece entre las dos metodologías?
- ¿Qué ofrece cada tipo de resultados (cuantitativos cualitativos) al estudio?

CAPÍTULO 2: Historia de la evaluación educativa

PRÁCTICA 1

A partir de la lectura del capítulo 2, elaboración de un esquema o diagrama en el que se resalten, por un lado, los diferentes autores relacionados con el desarrollo y evolución de la evaluación educativa y, por otro, las aportaciones y contribuciones realizadas por dichos autores.

PRÁCTICA 2

Lectura y análisis de un artículo.

Escudero Escorza, Tomás (2003): «Desde los tests hasta la investigación evaluativa actual. Un siglo, el xx, de intenso desarrollo de la evaluación en educación». *Revista ELección de Investigación y EValuación Educativa*, v.9, n.1. http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm.

Lectura individual del artículo y respuesta a las preguntas siguientes. Posteriormente, en el grupo pequeño se pueden cotejar y analizar las respuestas para que al final se haga un debate en el grupo clase.

1. ¿Cuáles son las características más importantes de cada una de las épocas de la historia de la evaluación educativa señaladas por el autor?
2. ¿Cómo ha evolucionado el concepto de evaluación educativa a lo largo de la historia?

CAPÍTULO 3: Concepto, componentes y fases de la evaluación

PRÁCTICA 1

Aproximación al concepto de evaluación.

Individualmente, anotar los términos asociados con el concepto de evaluación. Más tarde, en el pequeño grupo se pueden reunir todas las palabras y se da una definición de evaluación educativa. En el grupo clase, con ayuda de la pizarra buscar los elementos comunes a todas las definiciones e intentar llegar a la definición actual de evaluación.

PRÁCTICA 2

Definición de evaluación educativa.

A partir de las diferentes definiciones de evaluación educativa presentadas en el anexo 1, elaboración de una definición propia. Posteriormente, las definiciones de cada grupo pueden ser cotejadas y analizadas en el grupo clase.

PRÁCTICA 3

Análisis de los casos prácticos sobre las normas de evaluación.

Lectura de casos ilustrativos de las normas de evaluación (Joint Committee on Standards for Educational Evaluation, 1988) B2 (Viabilidad política) y C3 (Exposición total y franca). Análisis de los errores cometidos por los evaluadores y propuesta de soluciones o pautas de actuación para posteriores evaluaciones. En el grupo clase se puede hacer una puesta en común de los errores detectados y de las soluciones planteadas. También pueden ser analizadas otras normas.

PRÁCTICA 4

Búsqueda en internet de información sobre evaluación educativa.

Búsqueda de información a través de internet sobre evaluación educativa. Entre otros, búsqueda de informes de evaluación, referencias bibliográficas, agencias evaluadoras tanto autonómicas, estatales como internacionales y asociaciones o redes de evaluación, anotando la dirección y un breve resumen de cada página.

PRÁCTICA 5

Búsqueda en internet de bases de datos en donde encontrar bibliografía sobre evaluación educativa.

Búsqueda en diferentes bases de datos (ERIC, PSYCLIT, etc.) de información sobre distintos tópicos de la evaluación educativa. De la misma manera, se puede ir elaborando una base de datos propia sobre evaluación utilizando algún programa informático (File Maker, Open Access, etc.).

PRÁCTICA 6

Búsqueda de información en la biblioteca.

Visita alguna biblioteca e identificación de las distintas fuentes de información o de consulta que pueden ser útiles para una evaluación que se desee realizar. Hacer una lista de los hallazgos valorando las ventajas y las limitaciones de cada una de las fuentes encontradas.

PRÁCTICA 7

Elaboración de un mapa conceptual sobre la evaluación educativa.

Elaboración de un mapa conceptual que represente la visión del panorama actual de la evaluación educativa desde un punto de vista científico. Posteriormente, pueden ser pasados a transparencias y presentados al grupo mediante el retroproyector.

CAPÍTULO 4: Evaluación de programas

PRÁCTICA 1

Análisis de modelos de evaluación de programas.

A partir de la lectura de los distintos modelos de evaluación propuestos, señala las similitudes y las diferencias a nivel teórico y metodológico de los distintos modelos así como las ventajas y los inconvenientes de cada uno de ellos.

PRÁCTICA 2

Evaluabilidad de algún programa educativo.

Mediante esta práctica se trata de juzgar la evaluabilidad de un programa según la calidad del proceso de planificación e implementación llevado a cabo y de las limitaciones que se podrían encontrar al realizar la evaluación. El programa a evaluar podría ser cualquier asignatura y el cliente podría ser el propio profesor de la misma. Estudio y aplicación del instrumento «listado de cuestiones relevantes en evaluación de programas» (LCREP) de Fernández-Ballesteros y Hernández (1989) que aparece citado en Fernández-Ballesteros (1995d, pp. 339-352). Tras la corrección del instrumento se realiza un informe sobre las posibilidades de evaluar el programa de la asignatura y las posibles limitaciones.

PRÁCTICA 3

Elección del modelo de evaluación.

Elección de uno de los modelos de evaluación presentados en el capítulo y describir el contexto de explicación, alguna finalidad y algunos apuntes metodológicos relacionados con las técnicas e instrumentos a emplear. Posteriormente, se podrían poner en común los distintos modelos elegidos.

PRÁCTICA 4

Elección del modelo para la evaluación de un programa sobre educación en valores.

Ante el supuesto de que un director de Renovación Pedagógica de la Administración Educativa desea introducir un nuevo programa sobre *educación en valores* en el currículum de la Educación Secundaria Obligatoria. El director nos pide ayuda sobre qué tipo de modelo de evaluación sería el más adecuado para evaluar dicho programa. Se trata de elegir el modelo de evaluación más adecuado según las circunstancias señaladas. Posteriormente, se podría presentar en grupo la elección hecha y los argumentos esgrimidos para ello, pudiéndose además criticar esa elección aportando argumentos en contra.

PRÁCTICA 5

Visionado del vídeo «Evaluación de proyectos y programas de formación en la empresa. Estrategias y técnicas» (López-Barajas Zayas, 1999, VÍDEO. Madrid: UNED).

Tras la visión del vídeo mencionado se propone una guía para el debate posterior en grupo:

- Importancia de la evaluación de programas.
- Los pros y los contras de las distintas técnicas y procedimientos a utilizar en la evaluación.
- Aplicación práctica de los análisis de datos realizados.
- Controversia entre la evaluación cualitativa y cuantitativa.

PRÁCTICA 6

Análisis de distintos modelos de evaluación de programas educativos.

En pequeños grupos (o de forma individual) recopilar informes de evaluación, artículos de revista, actas de congreso, etc., donde aparezcan al menos dos perspectivas distintas de llevar a cabo un proceso de evaluación. Tras la búsqueda, elaborar un cuadro de doble entrada en donde aparezcan los siguientes datos:

FILAS: autor(es), breve descripción del programa, finalidad de la evaluación, metodología (diseño, instrumentos, dimensiones analizadas, etc.), análisis realizados, resultados obtenidos y decisiones adoptadas.

COLUMNAS: Modelo aplicado, tipo de evaluación (interna/externa) perspectiva cuantitativa/cualitativa.

Junto con el cuadro redactar un informe señalando las semejanzas y las diferencias entre los informes de evaluación analizados.

CAPÍTULO 5: Ámbitos de la evaluación educativa

PRÁCTICA 1

Análisis de la evaluación tradicional y alternativa del alumnado.

Teniendo en cuenta las características de la evaluación tradicional y de la evaluación alternativa, por un lado, y, por otro, las diferentes etapas educativas (Educación Infantil, Educación Primaria, Educación Secundaria Obligatoria y Postobligatoria y Educación Superior), analizar desde el punto de vista propio qué sentido tienen las prácticas evaluativas de la evaluación tradicional y alternativa en cada una de las etapas señaladas.

PRÁCTICA 2

Visionado del vídeo «En clave de reforma: La evaluación» (S. Fajardo e I. Gonzalo, 1994, VÍDEO. Madrid: SM).

Tras la visión del vídeo mencionado se propone una guía para el debate posterior en el grupo:

1. La integración de la evaluación en el proceso de enseñanza-aprendizaje.
2. La integración de los criterios y los instrumentos de evaluación en la programación.
3. La necesidad de la evaluación inicial.
4. Las funciones formativa y sumativa de la evaluación.
5. Los informes de evaluación de los alumnos.

PRÁCTICA 3

Lectura y análisis de un artículo.

Fernández Sierra, J. (1996): «¿Evaluación? No, gracias, calificación». *Cuadernos de Pedagogía*. n.º 243, pp. 92-97.

Lectura individual del artículo y responder a las siguientes preguntas. Posteriormente, se podrían cotejar y analizar las respuestas en grupo.

1. Para que la actual reforma suponga un cambio real en cuanto a la evaluación, ¿qué factores dentro del mundo educativo deben cambiar?
2. ¿Cómo refleja el autor la relación entre disciplina y evaluación?
3. Desde tu experiencia como estudiante, señala los puntos en los que estás de acuerdo y aquellos en los que estás en desacuerdo en relación a las ideas que plantea el autor.

PRÁCTICA 4

Análisis de los casos prácticos sobre los estándares de evaluación del personal.

Lectura de los casos ilustrativos de los estándares de evaluación de personal (Joint Committee on Standards for Educational Evaluation, 2001) n.º 1 de P1 (Orientación hacia el servicio) y n.º 2 de PR2 (Entorno de trabajo). Análisis de los errores cometidos por los evaluadores y propuesta de soluciones o pautas de actuación para posteriores evaluaciones. En el grupo se puede hacer una puesta en común de los errores detectados y de las soluciones planteadas. También pueden ser analizadas otras normas.

PRÁCTICA 5

Aplicación y corrección de una prueba estandarizada para evaluar la actuación docente del profesor (puede utilizarse cualquiera de los cuestionarios que habitualmente se emplean en la práctica totalidad de las universidades españolas).

Aplicación de la mencionada prueba en algún grupo. Mediante esta práctica se pretende observar, por una parte, la percepción individual sobre la actuación docente del profesor y, por otra, la obtención de los resultados medios para todo un grupo. Posteriormente, se puede analizar y valorar la adecuación, utilidad, factibilidad y probidad de la evaluación realizada mediante esta técnica.

PRÁCTICA 6

Elaboración de un proyecto de evaluación de un centro educativo.

Elaboración de una lista de las dimensiones, subdimensiones y aspectos que se evaluarían en un centro educativo. En cada caso se debe señalar el instrumento o la técnica que se utilizaría para obtener la información, así como la fuente de información que se emplearía. Posteriormente, se podría presentar en grupo el resultado obtenido.

PRÁCTICA 7

Lectura y análisis de un artículo.

Lukas, J. F.; Santiago, C. (2004): «Evaluación de centros de secundaria del País Vasco». *Revista Electrónica de Investigación Educativa*, v 6, n 2. <http://redie.ens.uabc.mx/vol6no1>.

Lectura individual del artículo y responder a las siguientes preguntas. Posteriormente, se pueden cotejar y analizar en grupo las respuestas dadas.

1. ¿Cuáles han sido las razones de que la evaluación de centros tenga el auge que tiene en la actualidad?
2. Teniendo en cuenta las dimensiones propuestas por Tiana y De Miguel, ubica en ellos el modelo de evaluación de REDES.
3. Señala las características más importantes del modelo de evaluación de centros propuesto.
4. ¿Cuáles son las ventajas y los inconvenientes del modelo de evaluación de REDES?

PRÁCTICA 8

Análisis de algún informe de evaluación del sistema educativo.

Lectura de algún informe de evaluación del sistema educativo efectuado por alguna administración educativa (por ejemplo, evaluación de la educación

primaria realizada por el INCE en 1996: *Lo que aprenden los alumnos de 12 años. Evaluación de la Educación Primaria. Datos básicos*. Madrid: Centro de Publicaciones del Ministerio de Educación y Cultura). Elaborar un resumen identificando los objetivos de la evaluación, señalando el modelo de evaluación utilizado, los resultados obtenidos en la evaluación y las acciones de mejora que podrían sugerirse como consecuencia.

CAPÍTULO 6: Diseños de investigación aplicados a la evaluación

PRÁCTICA 1

Lectura y análisis de un artículo.

De La Orden Hoz, A. (2004): «Nuevos horizontes en la investigación educativa». *Bordón*, v 56, n 1, pp. 117-128.

Tras la lectura del artículo responder a las siguientes preguntas. Posteriormente, se pueden cotejar y analizar en grupo las respuestas dadas.

1. ¿Cuáles son los diseños de investigación y evaluación que supuestamente van a prevalecer en los próximos años según el autor?
2. ¿En qué se basa para realizar este planteamiento?
3. ¿Qué ventajas plantean este tipo de diseños? ¿Cuáles son sus limitaciones?
4. ¿Cuál es tu opinión con respecto a las ideas del autor?

PRÁCTICA 2

Elección del diseño de investigación/evaluación.

Ante los siguientes supuestos presentados a continuación, identificar el diseño que se considera más adecuado para cada caso. Posteriormente, se pueden poner en común las coincidencias y las divergencias (los tres primeros ejemplos están tomados de Rodríguez Conde, 1999, p. 360).

Ejemplo 1. En un centro se detectan ciertos problemas de relación y comunicación entre profesores y alumnos. Se aplica un programa de formación de profesorado que mejore el problema. Queremos saber si el programa ha logrado sus objetivos.

Ejemplo 2: Se está interesado en evaluar un programa de desintoxicación de drogadictos. Se debe justificar el diseño elegido y proponer un esquema de aplicación.

Ejemplo 3: Se está interesado en evaluar un programa de animación socio-cultural en una residencia de ancianos. Se debe justificar el diseño elegido y proponer un esquema de aplicación.

Ejemplo 4: Se está interesado en probar la siguiente hipótesis: «*El método de lectura natural-global es más efectivo que el método silábico*». Se debe justificar el diseño elegido y proponer un esquema de aplicación.

Ejemplo 5: Se está interesado en conocer la opinión del alumnado universitario con respecto a un nuevo plan de tutorización de las prácticas externas puesto en marcha. Se debe justificar el diseño elegido y proponer un esquema de aplicación.

Ejemplo 6: Se está interesado en analizar si el rendimiento académico de los estudiantes universitarios está relacionado con sus antecedentes socioeconómico-culturales. Se debe justificar el diseño elegido y proponer un esquema de aplicación.

PRÁCTICA 3

Acceso al escenario.

Diseñar un esquema de actuación para iniciar un trabajo de campo. Junto con ello, redactar un escrito dirigido a una autoridad académica para solicitar permiso formal para acceder a una institución o contexto educativo determinado.

PRÁCTICA 4

Análisis de un informe de evaluación.

Teniendo en cuenta el estudio del Ararteko (2006) *Convivencia y conflictos en los centros educativos*. Vitoria-Gasteiz: Ararteko.

Dirección electrónica del informe: <http://www.ararteko.net/webs/iextras/conflictos-ceneduc2006/conflictos-ceneduc2006E.pdf>

Contesta a las cuestiones siguientes:

- a) Describe las características del diseño utilizado en la parte cuantitativa del estudio.
- b) Describe las características del diseño empleado en la parte cualitativa del estudio.
- c) Analiza las ventajas y los inconvenientes de cada uno de los planteamientos.

PRÁCTICA 5

Análisis de un informe de evaluación.

Teniendo en cuenta el estudio: Lukas, J. F.; Santiago, K. y Lizasoain, L. (2005). *Vitoria-Gasteiz como ciudad educadora: Evaluación de los programas del Departamento Municipal de Educación*. Vitoria-Gasteiz: DME.

Dirección electrónica del informe: http://www.sc.ehu.es/plwlumuj/ebalua-zioa_pedagogia/praktikak/INFORME-vitoria.pdf

Contesta a las cuestiones siguientes:

- a) Describe las características del diseño utilizado en la parte cuantitativa del estudio.
- b) Describe las características del diseño empleado en la parte cualitativa del estudio.
- c) Analiza las ventajas y los inconvenientes de cada uno de los planteamientos.

PRÁCTICA 6

Análisis de un informe de evaluación.

Teniendo en cuenta el estudio: Etxeberria, J.; Lukas, J. F. y Santiago, K. (2008). *IKASYS programaren aplikazio pilotuaren ebaluazioa*.

Dirección electrónica del informe: http://www.sc.ehu.es/plwlumuj/ebaluazio_pedagogia/praktikak/EBALUAZIO%20TXOSTENA.pdf

Contesta a las cuestiones siguientes:

- a) Describe las características del diseño empleado en la parte cuantitativa del estudio.
- b) Describe las características del diseño utilizado en la parte cualitativa del estudio.
- c) Analiza las ventajas y los inconvenientes de cada uno de los planteamientos.

CAPÍTULO 7: Procedimientos para la recogida de la información

PRÁCTICA 1

Elaboración de un cuestionario.

Diseño de un cuestionario como instrumento de recogida de datos para evaluar el proceso de enseñanza-aprendizaje de alguna asignatura. El diseño del cuestionario constará de los elementos siguientes:

- Objetivos del cuestionario.
- Listado de los aspectos sobre los que se va a preguntar.
- Redacción de las preguntas.
- Primer borrador del cuestionario (título, presentación, instrucción, preguntas, etc.).
- Correcciones desde el punto de vista lógico-racional.
- Cuestionario definitivo.
- Definir la población/muestra.

PRÁCTICA 2

Preparación y aplicación de una entrevista.

Preparar una entrevista para las personas implicadas en alguna asignatura acerca del proceso de enseñanza-aprendizaje desarrollado. Para preparar la entrevista se seguirá el esquema siguiente:

- Definir los objetivos de la entrevista en consonancia con los objetivos de la evaluación.
- Muestreo de las personas a entrevistar.
- Contenido y naturaleza de las preguntas.
- Organización y secuencia de las preguntas.
- Registro de la información.

A partir de la preparación, realizar una entrevista. Al final se deberá presentar un informe en donde aparezcan:

- Los pasos seguidos en la preparación de la entrevista y las preguntas preestablecidas.
- Análisis de las respuestas recibidas.
- Propuesta de nuevas preguntas para realizar entrevistas posteriores.

PRÁCTICA 3

Panel sobre técnicas de recogida de información.

Para esta práctica, es necesario un grupo amplio de personas. Se hacen grupos de 10 personas. A cada miembro de ese grupo se le propone un tema a investigar:

1. La entrevista estructurada.
2. El cuestionario.
3. Los tests estandarizados y las pruebas objetivas.
4. El autoinforme.
5. La observación estructurada.
6. La entrevista abierta.
7. El grupo de discusión.
8. La observación participante.
9. La técnica Delfi.
10. El socioanálisis.

En un primer momento, todas las personas que tienen el mismo número deben buscar bibliografía sobre el tema, estudiarla y elaborar en grupo un pequeño documento sobre la cuestión. Posteriormente, se vuelven a juntar los grupos originales compuestos por las 10 personas. En este grupo, cada uno de sus miembros debe explicar al resto de sus compañeros la técnica de recogida de información que ha estudiado.

PRÁCTICA 4

Aplicación y corrección de la prueba estandarizada «Escala de Clima Social de Clase» de Moos (TEA).

Aplicación de una prueba estandarizada a un grupo de personas que asisten al mismo curso en una misma clase. Mediante esta práctica se pretende observar la percepción individual del clima de clase de cada alumno con respecto a sus compañeros y por otra obtener los resultados medios para toda la clase. Posteriormente, se hace un comentario en grupo para valorar la adecuación, utilidad, factibilidad y probidad de la evaluación realizada mediante esta técnica. También podría utilizarse cualquier otro test estandarizado para realizar la misma práctica.

PRÁCTICA 5

Construcción de un protocolo de observación.

Construir un protocolo para la observación sistemática de la actuación docente del profesor universitario para analizar las estrategias utilizadas por los profesores para fomentar la participación activa de los alumnos en el desarrollo de las clases.

PRÁCTICA 6

Análisis de un informe de evaluación

Describe los procedimientos de recogida de información utilizados en el estudio del Ararteko (2006) *Convivencia y conflictos en los centros educativos*.

Dirección electrónica del informe: <http://www.ararteko.net/webs/iextras/conflictos-ceneduc2006/conflictos-ceneduc2006E.pdf>

PRÁCTICA 7

Análisis de un informe de evaluación

Describe los procedimientos de recogida de información utilizados en el estudio de Lukas, J. F.; Santiago, K. y Lizasoain, L. (2005). *Vitoria-Gasteiz como ciudad educadora: Evaluación de los programas del Departamento Municipal de Educación*.

Dirección electrónica del informe: http://www.sc.ehu.es/plwlumuj/ebalua-zioa_pedagogia/praktikak/INFORME-vitoria.pdf

PRÁCTICA 8

Análisis de un informe de evaluación

Describe los procedimientos de recogida de información utilizados en el estudio de Etxeberria, J.; Lukas, J. F. y Santiago, K. (2008). *IKASYS programaren aplikazio pilotuaren ebaluazioa*.

Dirección electrónica del informe: http://www.sc.ehu.es/plwlumuj/ebaluazioa_pedagogia/praktikak/EBALUAZIO%20TXOSTENA.pdf

CAPÍTULO 8: Procedimientos para el análisis de la información

PRÁCTICA 1

Análisis de datos cuantitativos y cualitativos.

Búsqueda de ejemplos reales o invención de ejemplos ficticios en los que se den cada una de las situaciones siguientes (véase cuadro 8.1):

1. Análisis cuantitativos de datos cualitativos.
2. Análisis cuantitativos de datos cuantitativos.
3. Análisis cualitativos de datos cualitativos.
4. Análisis cualitativos de datos cuantitativos.

PRÁCTICA 2

Comparación entre los análisis cuantitativos y los cualitativos.

Señalar las similitudes y las diferencias entre el análisis cuantitativo y el análisis cualitativo.

PRÁCTICA 3

Análisis de informes de evaluación.

Partiendo de informes de evaluación recogidos en la biblioteca o documentación personal, extraer el esquema desarrollado en dicho informe de evaluación y valorar hasta qué punto la información presentada recoge todos los aspectos necesarios para poder llevar a cabo una toma de decisiones sobre el programa.

PRÁCTICA 4

Análisis de un informe de evaluación

Describe los procedimientos de análisis de información utilizados en el estudio del Ararteko (2006) *Convivencia y conflictos en los centros educativos*.

Dirección electrónica del informe: <http://www.ararteko.net/webs/iextras/conflictos-ceneduc2006/conflictos-ceneduc2006E.pdf>

PRÁCTICA 5

Análisis de un informe de evaluación

Describe los procedimientos de análisis de información utilizados en el estudio de Lukas, J. F.; Santiago, K. y Lizasoain, L. (2005). *Vitoria-Gasteiz como ciudad educadora: Evaluación de los programas del Departamento Municipal de Educación*.

Dirección electrónica del informe:

http://www.sc.ehu.es/plwlumuj/ebaluazioa_pedagogia/praktikak/INFORME-vitoria.pdf

PRÁCTICA 6

Análisis de un informe de evaluación

Describe los procedimientos de análisis de información utilizados en el estudio de Etxeberria, J.; Lukas, J. F. y Santiago, K. (2008). *IKASYS programaren aplikazio pilotuaren ebaluazioa*.

Dirección electrónica del informe:

http://www.sc.ehu.es/plwlumuj/ebaluazioa_pedagogia/praktikak/EBALUAZIO%20TXOSTENA.pdf

Anexos

Anexo 1: Definiciones de evaluación educativa¹

Tyler (1950, p. 69):

Proceso para determinar en qué medida los objetivos educativos han sido alcanzados.

Cronbach (1963, p. 672):

Proceso de recopilación y utilización de la información para tomar decisiones.

Suchman (1967):

Proceso de emitir juicios de valor.

Scriven (1967; 1991a, p. 139):

Proceso por el que se determina el mérito o valor de alguna cosa.

Scriven (1967):

Actividad metodológica que consiste simplemente en la recopilación y combinación de datos de trabajo mediante la definición de unas metas que proporcionen escalas comparativas o numéricas, con el fin de justificar los instrumentos de recopilación de datos, las valoraciones y la selección de las metas.

¹. Las definiciones están ordenadas cronológicamente.

Stake (1967):

Un proceso sistemático de emisión de juicios basado en una descripción de las discrepancias entre las ejecuciones de estudiantes y profesores y los estándares del aprendizaje y de la enseñanza previamente definidos.

Alkin (1969):

Proceso de reconocer las áreas de decisión importantes, seleccionar la información apropiada y reconocer y analizar la información para transmitir datos que ayudan a tomar decisiones para tomar alternativas.

Weiss (1970):

La evaluación analiza la relación programa/necesidad social, midiendo el grado, y si es posible, la profundidad, en que sus fines son logrados, y por lo tanto, las necesidades cubiertas, o los problemas solucionados.

Franco (1971):

Fijar el valor de una cosa; para hacerlo se requiere efectuar un procedimiento mediante el cual se compara aquello a evaluar respecto de un criterio o patrón determinado.

Provus (1971, p. 242):

Comparación de las ejecuciones o comportamientos de los alumnos con ciertas normas para determinar si se debe llevar a cabo, se continúa o se concluye en el proceso de enseñanza.

Stanley, Glock y Wardeberg (1972):

Proceso en el que un maestro suele emplear información derivada de muchos orígenes para formular un juicio de valor.

Stufflebeam (1973):

El proceso de delimitar, obtener y proveer de información útil para juzgar entre alternativas de decisión.

Worthen y Sanders (1973):

Determinación del valor de una cosa. Incluye la obtención de información para juzgar el valor de un programa, un procedimiento, un producto, un objetivo o la utilidad potencial de aproximaciones alternativas para alcanzar determinados objetivos.

Bloom *et al.* (1975, p. 23):

Reunión sistemática de evidencias a fin de determinar si en realidad se producen ciertos cambios en los alumnos y establecer el grado de cambio en cada estudiante.

Levine (1975):

Es el examen de los efectos, resultados u outputs de un programa.

Mager (1975, p. 20):

Acto de comparar una medida con un estándar y emitir un juicio basado en la comparación.

Tuckman (1975, p. 12):

Proceso mediante el cual las partes, los procesos o resultados de un programa se examinan para ver si son satisfactorios con referencia a los objetivos establecidos, a nuestras propias expectativas o nuestros estándares de excelencia.

Beeby (1977):

La recogida e interpretación sistemática de la evidencia, que conduce, como parte del proceso, a un juicio de valor con vistas a la acción.

De la Orden (1977):

Un medio que permite observar y describir con mayor precisión los aspectos cuantitativos y cualitativos de la estructura, el proceso y el producto de la educación. Su finalidad es facilitar una predicción y un control lo más exacto posible del proceso educativo.

Lafourcade (1977, p. 21):

Etapa del proceso educacional que tiene como fin controlar de modo sistemático en qué medida se han logrado los resultados previstos en los objetivos que se hubieran especificado con antelación.

Kaufman y English (1979):

Consiste en analizar las discrepancias entre lo que es y lo que debe ser; es decir, la disparidad entre una situación deseada o esperada y otra existente o real.

Unesco (1979):

Proceso encaminado a determinar sistemática y objetivamente la pertinencia, eficiencia, eficacia e impacto de todas las actividades a la luz de sus objetivos. Se trata de un proceso organizativo para mejorar las actividades que se hallan aún en marcha y ayudar en la planificación, programación y decisiones futuras.

Forns (1980):

La palabra evaluación nos induce inmediatamente a pensar en la posibilidad de efectuar cierto número de estimaciones cuantitativas y cualitativas acerca del valor, la importancia o la incidencia de determinados objetos, personas o hechos.

Livas González (1980, p. 14):

Proceso que consiste en obtener información sistemática y objetiva acerca de un fenómeno y en interpretar dicha información a fin de seleccionar entre distintas alternativas de elección.

Popham (1980, p. 16):

Consiste en un juicio formal del valor de los fenómenos educativos.

Berk (1981, p. 4):

Es la aplicación del método científico para recoger la información fiable y válida necesaria para tomar decisiones sobre un programa educativo.

Perrenoud (1981):

Criterio social de desviación o conformidad a una norma de buen comportamiento escolar.

OMS (1981):

Un medio sistemático de aprender empíricamente y de utilizar las lecciones aprendidas para el mejoramiento de las actividades en curso y para el fomento de una planificación más satisfactoria mediante una selección rigurosa entre las distintas posibilidades de acción futura.

Rossi (1981, p. 221):

La aplicación del conocimiento de las ciencias sociales y los métodos de investigación para la orientación de programas sociales.

De la Orden (1982, p. 22):

Proceso de recogida y análisis de información relevante para describir cualquier faceta de la realidad educativa y formular un juicio sobre su adecuación a un patrón o criterio previamente establecido, como base para la toma de decisiones.

Pérez Gómez (1983, p. 431):

Proceso de recogida y provisión de evidencias sobre el funcionamiento y evolución de la vida en el aula, en base a las que se forman decisiones sobre la posibilidad, efectividad y valor educativo del currículum.

Rotger Amengual (1983, p. 603):

Actividad sistemática y continua, integrada dentro del proceso educativo, que tiene por objeto proporcionar la máxima información para mejorar este proceso, reajustando sus objetivos, revisando críticamente planes y programas, métodos y recursos, y facilitando la máxima ayuda y orientación a los alumnos.

Taba (1983):

Proceso intrincado y complejo que comienza con la formulación de objetivos, que involucra decisiones sobre los medios para asegurar la evidencia de su cumplimiento, los procesos de interpretación para llegar al significado de esta evidencia y los juicios sobre las seguridades y las deficiencias de los estudiantes y que finaliza con las decisiones acerca de los cambios y las mejoras que necesitan el currículum y la enseñanza.

Asociación Nacional de Universidades e Instituciones de Educación Superior (1984, pp. 22-23):

Un proceso continuo, integral y participativo que permite identificar una problemática, analizarla y explicarla mediante información relevante, y que como resultado proporciona juicios de valor que sustentan la consecuente toma de decisiones.

Ruttman (1984, p. 12):

Es la utilización del método científico con la intención de tomar decisiones sobre la implementación de un programa y medir sus resultados.

Tenbrink (1984, p. 19):

Proceso de obtener información y usarla para formar juicios que a su vez se utilizarán en la toma de decisiones.

Alvira Martín (1985):

La acumulación de información sobre una intervención —programa—, sobre su funcionamiento y sobre sus efectos y consecuencias.

Carreño (1985):

Acción de juzgar, de inferir juicios a partir de cierta información desprendida directa o indirectamente de la realidad evaluada, o bien, atribuir o negar calidades o cualidades al objeto evaluado, o, finalmente, establecer reales valoraciones en relación con lo enjuiciado.

De la Orden (1985a, p. 133):

Un componente intrínseco del proceso educativo en su conjunto, en la medida en que la educación sea una actividad sistemática intencional. Supone un proceso o conjunto de procesos para la obtención y análisis de la información relevante en que apoyar un juicio de valor, en que esencialmente consiste, como soporte a una decisión. Dicho juicio de valor puede recaer sobre un objeto, fenómeno o proceso o acontecimiento.

Eisner (1985):

Proceso que, en parte, nos ayuda a determinar si lo que hacemos en las escuelas está contribuyendo a conseguir los fines valiosos o si es antitético a estos fines.

Gimeno Sacristán (1985, p. 215):

La comprobación de la validez de las estrategias didácticas, configuradas por las opciones que se han tomado de las numerosas dimensiones de los elementos didácticos, en orden a la consecución de los objetivos didácticos.

Gronlund (1985, p. 5):

Proceso sistemático de recopilar, analizar e interpretar información para determinar el grado en que los alumnos han logrado los objetivos instructivos.

Cabrera Rodríguez (1986, p. 77):

Proceso sistemático de obtener información para describir, sea algún elemento, sea el proceso, sean los resultados educativos, con objeto de emitir un juicio sobre ellos, según su adecuación a unos criterios previamente establecidos, y todo ello en orden a servir de base para una eventual decisión educativa.

Fernández Pérez (1986):

Actividad sistemática integrada en el proceso educativo, cuya finalidad es el mejoramiento, lo más posible, del alumno en todos los aspectos de su personalidad, y de una información ajustada sobre el proceso educativo y sobre los factores personales y ambientales que en éste inciden.

Kemmis (1986):

Proceso de proyectar, obtener, proveer y organizar informaciones y argumentos que permitan a las personas y grupos interesados participar en el debate crítico sobre un programa específico.

Mateo (1986, p. 135):

Proceso contextualizado de aplicar procedimientos científicos para recoger información válida y fiable, orientada a la toma de decisiones, acerca de un sujeto determinado (assessment) o un programa o intervención educacional (investigación evaluativa).

Pérez Juste (1986, p. 31; 1995e, p. 23):

Acto de valorar una realidad, que forma parte de un proceso cuyos momentos previos son los de fijación de las características de la realidad a valorar, y de recogida de información sobre las mismas, y cuyas etapas posteriores son la información y la toma de decisiones en función del juicio de valor emitido.

Cabrera (1987):

Proceso de recogida y análisis de información relevante en que apoyar un juicio de valor sobre la entidad evaluada que sirva de base para una eventual toma de decisiones.

Pineault y Daveluy (1987):

Proceso que consiste en determinar y en aplicar criterios y normas con el fin de realizar un juicio sobre los diferentes componentes del programa, tanto en el estadio de su concepción como de su ejecución, así como sobre las etapas del proceso de planificación que son previas a la programación.

Rodríguez Lajo (1987, p. 182):

Proceso sistemático de obtener información relevante acerca de un objeto educativo, con el propósito de enjuiciar su adecuación a unos criterios, previamente establecidos, que pueden servir de base a unas posibles decisiones educativas.

Stufflebeam y Shinkfield (1987):

Proceso sistemático planificado, dirigido y realizado con el fin de ayudar a un grupo de clientes a juzgar y/o perfeccionar el valor y/o mérito de algún objeto.

Stufflebeam y Shinkfield (1987, p. 183):

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

Joint Committee (1988, p. 37; 1998, p. 31):

Investigación sistemática del valor o mérito de algún objeto.

García Ramos (1989, p. 20):

Actividad o proceso de identificación, recogida y tratamiento de datos sobre elementos y hechos educativos con el objetivo de valorarlos primero y, sobre dicha valoración, tomar decisiones.

Alfaro (1990):

Establecer una comparación entre lo deseado y lo realizado. Comparar lo que nos proponemos en los objetivos con lo que hemos sido capaces de realizar.

Escudero (1990, p. 25):

El proceso de identificación, remodelación y tratamiento de datos, seguido para obtener una información que justifique una determinada decisión.

Rosales (1990, p. 14):

Función característica del profesor que consiste básicamente en una actividad de reflexión sobre la enseñanza. Esta reflexión se entiende como evaluación si se sitúa en el origen de decisiones para el perfeccionamiento.

Sancho (1990, p. 8):

Emisión de un juicio sobre el valor de algo proyectado o realizado por algún individuo o grupo y presupone un proceso de recogida de información sistemática y relevante que garantice la calidad del juicio formulado.

Walberg y Haertel (1990, p. XVII):

Examen cuidadoso, realizado con rigor, de un currículum educativo, de un programa, de una institución, de una variable organizativa o de una política educativa.

Alvira (1991, p. 10):

Evaluar es emitir juicios de valor, adjudicar valor o mérito a un programa/intervención, basándose en información empírica recogida sistemática y rigurosamente.

Chadwick y Rivera (1991, p. 38):

Reunión sistemática de evidencia, a fin de determinar si en realidad se producen ciertos cambios (aprendizajes) en los alumnos, y controlar, también, el estadio del aprendizaje en cada estudiante.

De la Orden (1991):

Conjunto de procesos sistemáticos de recogida, análisis e interpretación de información relevante para medir o describir cualquier faceta de la realidad educativa y sobre esta descripción formular un juicio de valor por su comparación con un criterio o patrón, todo ello como base para tomar decisiones.

Secretaría de Educación Pública (1991, pp. 41-42):

Un proceso permanente que permite mejorar de manera gradual la calidad académica. En consecuencia, debe incorporar un análisis a lo largo del tiempo de los avances y logros, identificar obstáculos y promover acciones de mejoramiento académico.

Weiss (1991):

Medir los efectos de un programa por comparación con las metas que se propuso alcanzar; a fin de contribuir a la toma de decisiones sobre el programa y para mejorar la programación futura.

Arnal, Del Rincón y Latorre (1992, p. 213):

Un proceso, o conjunto de procesos, para la obtención y análisis de información significativa en que apoyar juicios de valor sobre un objeto, fenómeno, proceso o acontecimiento, como soporte de una eventual decisión sobre el mismo.

Pérez Juste (1992):

Proceso sistemático, diseñado intencional y técnicamente, de recogida de información rigurosa —valiosa, válida y fiable—, orientado a valorar la calidad de un programa, como base para la posterior toma de decisiones de mejora tanto del programa como del personal implicado y, de modo indirecto, del cuerpo social en que se encuentra inmerso.

Real Academia Española (1992):

Señalar el valor de una cosa. Estimar, apreciar, calcular el valor de una cosa. Estimar los conocimientos, las aptitudes y el rendimiento de los alumnos.

Ferrández (1993, p. 11):

Emitir un juicio valorativo sobre una realidad educativa en función de unos datos y con el propósito de tomar decisiones al respecto.

Rossi y Freeman (1993):

La aplicación sistemática de los procedimientos de investigación social en la valoración de la conceptualización y diseño de un programa, de la implementación y de la utilidad de los programas de intervención social.

Santos Guerra (1993a, p. 73):

Evaluar es atribuir valor a las cosas, es afirmar algo sobre su mérito. Se evalúa para comprender y, en definitiva, para cambiar y mejorar.

Aguilar y Ander-Egg (1994, p. 8):

Una forma de valoración sistemática que se basa en el uso de procedimientos que, apoyados en el uso del método, sirven para identificar, obtener y proporcionar la información pertinente y enjuiciar el mérito y el valor de algo, de manera injustificable.

Aguilar y Ander-Egg (1994, p. 18):

Es una forma de investigación social aplicada, sistemática, planificada y dirigida; encaminada a identificar, obtener y proporcionar, de manera válida y fiable, datos e información suficiente y relevante, en que apoyar un juicio acerca del mérito y el valor de los diferentes componentes de un programa (tanto en la fase de diagnóstico, programación o ejecución), o de un conjunto de actividades específicas que se realizan, han realizado o realizarán, con el propósito de producir efectos y resultados concretos; comprobando la extensión y el grado en que dichos logros se han dado, de forma tal, que sirva de base o guía para una toma de decisiones racional e inteligente entre cursos de actuación, o para solucionar problemas y promover el conocimiento y la comprensión de los factores asociados al éxito o al fracaso de sus resultados.

Darder y Mestres (1994, p. 424):

Un proceso sistemático de carácter valorativo, decisorio y prospectivo; recogida de información de forma selectiva y orientada, una vez elaborada, a facilitar la toma de decisiones y la emisión de juicios y sugerencias respecto al futuro.

Medina (1994, p. 89):

Es el juicio de valor, que realizamos sobre la concepción y práctica educativa a fin de conocerla en profundidad, actuar en consecuencia y buscar las acciones más adecuadas para mejorar los procesos de enseñanza-aprendizaje, la formación de los alumnos y la profesionalización del docente.

Nieto Gil (1994):

Actividad valorativa e investigadora, que facilita el cambio educativo y el desarrollo profesional de los docentes. Su finalidad es adecuar o reajustar permanentemente el sistema escolar a las demandas sociales y educativas. Su ámbito de aplicación abarca no sólo a los alumnos —en especial sus adquisiciones instructivas—, sino también a los profesores y los centros educativos, a los servicios de apoyo y a la misma administración educativa.

Casanova (1995a, p. 54):

La evaluación aplicada a la enseñanza y aprendizaje consiste en un proceso sistemático y riguroso de recogida de datos, incorporado al proceso educativo desde su comienzo, de manera que sea posible disponer de información continua y significativa para conocer la situación, formar juicios de valor con respecto a ella y tomar las decisiones adecuadas para proseguir la actividad educativa mejorándola progresivamente.

Casanova (1995a, p. 55):

Recogida de información rigurosa y sistemática para obtener datos válidos y fiables acerca de una situación con objeto de formar y emitir un juicio de valor con respecto a ella. Estas valoraciones permitirán tomar las decisiones consecuentes en orden a corregir o mejorar la situación evaluada.

Doval (1995, p. 211):

Proceso o resultado de la indagación de una realidad educativa a examen, que permite describirla, compararla con un estándar y emitir juicios ordenados a la toma de decisiones.

Doval (1995, p. 213):

Aquella actividad técnico-profesional que se ocupa de justipreciar los efectos o grado de eficiencia de una secuencia de acciones pedagógicas o programas (sistemáticamente ordenados a producirlos) y por comparación con los criterios de satisfacción en ellos establecidos, al objeto de tomar decisiones de promoción o en su caso de reconsideración y ajuste del propio programa.

García Llamas (1995, p. 46):

Proceso de identificación, recolección y tratamiento de datos para obtener una información que justifique una determinada decisión.

Pérez Juste (1995a, p. 82; 2000, p. 269; 2002, p. 61):

La valoración, a partir de criterios y referencias preespecificados, de la información técnicamente diseñada y sistemáticamente recogida y organizada, sobre cuantos factores relevantes integran los procesos educativos para facilitar la toma de decisiones de mejora.

Rul Gargallo (1995, p. 77):

Es la expresión de la capacidad humana de construir valores ajustados a los hechos a partir de la consideración rigurosa de los datos e informaciones objetivadas.

Amezcu Viedma (1996, p. 2):

Un proceso que compara lo observado con lo estándar o esperado.

Forner y Latorre (1996, p. 70):

Proceso sistemático de obtención y análisis de información significativa en que se basan juicios de valor sobre un fenómeno.

García Garrido (1996, p. 256):

La evaluación viene hoy referida sobre todo como control de la calidad, tanto en lo que se refiere a las tareas de aprendizaje como a los resultados más generales de las instituciones e incluso del sistema educativo.

Jiménez Lara (1996, p. 91):

Consiste en un examen crítico del desarrollo y de los efectos actuales o potenciales de un programa, y se orienta a aumentar la eficacia del programa en curso y a proporcionar una base orientadora para el diseño e implantación de programas futuros.

Nieto Gil (1996, p. 8):

Evaluar es establecer criterios y aplicar instrumentos de medida, tanto de rasgos psíquicos como de conductas o procesos, así como también de productos educativos. En estas tres vertientes, tanto del educador como de los educandos, esta actividad siempre será realizada para incidir en los procesos, de forma que se mejoren los resultados o productos, tanto de tipo cognitivo como afectivo de los alumnos.

Nevo (1997, p. 22):

La recogida sistemática de información referente a la naturaleza y calidad de los objetos educativos.

Mateo (1998):

Proceso de recogida de información orientado a la emisión de juicios de mérito o de valor respecto de algún sujeto, objeto o intervención con relevancia educativa.

Mayor Ruiz (1998, p. 24):

Un proceso que parte de la descripción de una realidad, más o menos concreta —un centro, un programa o un profesor—; que contrasta la información recogida; que reflexionando sobre ella es capaz de interpretarla, analizarla y valorarla; que la contextualiza y la difunde a los interesados; y que proporciona argumentos para tomar decisiones de mejora.

Bordas Alsina (1999, p. 238):

Detectar cómo es una realidad educativa —alumno, grupo clase, centro formativo o educativo, planificación curricular, sistema educativo, etc.—, sea desde una perspectiva cualitativa, cuantitativa o mixta, con el fin de tomar decisiones.

Jiménez Jiménez (1999b, p. 21):

Proceso ordenado continuo y sistemático de recogida de información cuantitativa y cualitativa, que responde a ciertas exigencias (válida, creíble, dependiente, fiable, útil...) obtenida a través de ciertas técnicas e instrumentos, que tras ser cotejada o comparada con criterios establecidos nos permite emitir juicios de valor fundamentados que facilitan la toma de decisiones que afectan al objeto evaluado.

López Mojarro (1999, p. 22):

La evaluación es la búsqueda de un conocimiento de la realidad. Este conocimiento, para ser fiable y eficaz, debe basarse en un análisis sistemático de dicha realidad. El conocimiento de la realidad incluye los hechos, los procesos que los han generado y los factores que los han condicionado. La evaluación es un acto profesional con intención de mejora.

Rodríguez Conde (1999, p. 149):

Conjunto de procesos sistemáticos de recogida, análisis e interpretación de información válida y fiable, que en comparación con una referencia nos permita llegar a una decisión optimizante.

Cabrera (2000a, p. 17):

La evaluación es un proceso sistemático de obtener información objetiva y útil en la que apoyar un juicio de valor sobre el diseño, la ejecución y los resultados de la formación con el fin de servir de base para la toma de decisiones pertinentes y para promover el conocimiento y la comprensión de las razones de los éxitos y los fracasos de la formación.

De Miguel (2000a, p. 290):

Proceso sistemático de búsqueda de evidencias para formular juicios de valor que orienten la toma de decisiones.

Lukas, et al. (2000a, p. 76):

Proceso sistemático de obtención y análisis de información significativa en que se basan juicios de valor sobre un fenómeno.

Mateo (2000a, p. 13):

Proceso de recogida y análisis de información relevante con el fin de describir cualquier realidad educativa de manera que facilite la formulación de juicios sobre la adecuación a un patrón o criterio de calidad debatido previamente, como base para la toma de decisiones.

Mateo (2000b):

Proceso de reflexión sistemática, orientado sobre todo a la mejora de la calidad de las acciones de los sujetos, de las intervenciones de los profesionales, del funcionamiento institucional o de las aplicaciones a la actividad educativa.

Carrión Carranza (2001, p. 37):

Proceso de comunicación de una comunidad escolar cuyo objeto es la identificación de los atributos del sistema de educación institucional, que constituyen la esencia, la naturaleza y la estructura de la institución respectiva, para calificarlos y determinar sus virtudes y defectos; su fin último es el mejoramiento de la formación de los individuos.

Caride Gómez (2002, p. 145):

Proceso por el cual una sociedad aprende de sí misma.

Salinas (2002, pp. 21-22):

Un conjunto de experiencias y vivencias de profesores y alumnos que tienden a tratar de evidenciar o constatar determinados aprendizajes del alumno (unos y no otros) con la finalidad tanto de mejorarlos como de juzgarlos desde un marco institucional que regula, distribuye y valora el conocimiento que dichos juicios toman como referencia.

Perales Montolío (2002b, p. 11):

Proceso sistemático de indagación y comprensión de la realidad educativa que pretende la emisión de un juicio de valor sobre la misma, orientado a la toma de decisiones y la mejora.

Anexo 2: Recursos informáticos para la evaluación educativa

En este anexo vamos a presentar algunos recursos y direcciones electrónicas que pueden ser interesantes para la investigación y la evaluación en Educación. Vamos a presentar una selección de las que consideramos más interesantes en el momento de escribir estas páginas, pues es de sobra conocida la rápida pérdida de novedad en este tipo de cuestiones. Puede accederse a todos estos recursos a través de la siguiente dirección electrónica: <http://www.sc.ehu.es/plwsaecc/webguneak.htm>.

Bases de datos documentales

Las bases de datos documentales constituyen una de las fuentes de información más importantes, ya que permiten en un espacio relativamente corto de tiempo acceder a gran cantidad de documentos y en algunos casos incluso almacenarlos y recuperarlos. Desde las bases de datos en CD-ROM que comenzaron a producirse a mediados de la década de 1980 hasta el día de hoy en que hay bases de datos accesibles a través de internet, la proliferación de las bases de datos que potencialmente interesan a los evaluadores e investigadores en el campo educativo ha sido muy elevada, por lo que la presentación de todas ellas sería imposible. A continuación, mencionamos solamente algunas de las más representativas:

BIBLIOMECA. Está producida por el Centro de Información y Documentación Educativa (CIDE) del Ministerio de Educación y Cultura y recoge los fondos bibliográficos de la biblioteca del CIDE.

CD THÈSSES. Es el catálogo de tesis doctorales francesas. Incluye cerca de 200.000 títulos.

DISSERTATION ABSTRACTS. Creado por University Microfilms International. Es la mayor base de datos que existe en el mundo sobre tesis, tanto doctorales como de licenciatura o máster. Fundamentalmente recoge las tesis de Estados Unidos y Canadá. Se distribuye con CD-ROM, y con restricciones, en internet. Está compuesta por las siguientes bases de datos:

- Comprehensive Dissertation Index (CDI). Tesis doctorales norteamericanas aprobadas desde 1861. Son más de medio millón de entradas.
- Dissertation Abstracts International (DAI). Relación de tesis doctorales más actuales, con un resumen de los contenidos.
- Master Abstracts International (MAI): Tesis de máster o licenciatura incluyendo resúmenes de los contenidos.
- American Doctoral Dissertation (ADD): Tesis doctorales aprobadas en EE.UU. cada nuevo curso.

DOC-E (Documentos de Educación). Es una base de datos constituida por unos 60.000 registros y organizada en las siguientes 10 bases de datos:

- Documentación Educativa.
- Revistas de Educación.
- Legislación Educativa.
- Convocatorias y Actividades de Formación y Actualización del Profesorado.
- Cultura y Actualidad.
- Vídeos Didácticos.
- *Software* Educativo.
- CD-ROM.
- Literatura Infantil/Juvenil.
- Libros de Conocimiento y Librería.

Las primeras cinco bases de datos son acerca de documentación e información educativa y las otras cinco sobre recursos didácticos.

Educational Research Abstracts (ERA). Base de datos que contiene resúmenes de investigaciones educativas internacionales recogidas en diversas publicaciones y enlaces al texto completo de los artículos. Es necesario suscribirse, pero permite hacer una suscripción gratuita de 30 días.

ERIC (Educational Resources Information Center). Es una base de datos creada por el National Institute of Education, organismo del Ministerio de

Educación de Estados Unidos. Es una red de centros especializados destinados a recoger y difundir información sobre diversos temas educativos. Su principal base de datos es conocida como ERIC, que está dividida en dos bases de datos, el CIJE (*Current Index of Journal in Education*) que vacía revistas y RIE (*Resources in Education*) que versa sobre la conocida como «literatura gris» o de escasa difusión como: comunicaciones, informes administrativos, tesis, etc. Es la fuente de información sobre educación más amplia que existe, ya que además del amplio volumen de referencias que contiene, cuenta con un servicio bien organizado para conseguir copias de los documentos referenciados.

EUDISED R & Data Base. Producida por el Consejo de Europa, contiene información sobre proyectos de investigación educativa, tanto las que están en curso como las ya finalizadas de 34 países europeos. Se corresponde con el repertorio *eudised European Educational Research Yearbook*.

FRANCIS (*Fichier de Recherches Automatisées sur les Nouveautés, la Communication et l'Information en Sciences Sociales et Humaines*). Base de datos sobre Ciencias Sociales y Humanas con un apartado dedicado a Educación y que está producida por el Institut de l'Information Scientifique et Technique. Es la principal base de datos científica en lengua francesa, que recoge más de un millón de referencias. Contiene información acerca de 250 revistas escritas en francés y en inglés. Incluye también un pequeño número de otros tipos de documentos como informes de investigación, trabajos presentados a encuentros científicos y libros. Durante mucho tiempo ha estado correspondido con el repertorio impreso *Francis. Bulletin Signalétique 520 Sciences de l'Education*.

ISOC. PSEDISOC. Es una base de datos realizada por el Centro e Información y Documentación Científica (CINDOC) del CSIC, que ha elaborado diferentes bases de datos sobre publicaciones españolas. La más relacionada con nuestro ámbito de trabajo es sin duda la PSEDISOC. La versión escrita de esta base es el Índice Español de Ciencias Sociales. Serie Psicología y Ciencias de la Educación. Contiene referencias de aproximadamente 221 revistas españolas de Ciencias de la Educación y Psicología. Están indexados artículos publicados desde 1976.

PsycINFO. Creada por la APA (American Psychological Association) es la base de datos más amplia en el campo de la Psicología. Incluye información procedente de varias bases como el *Psychological Abstracts* y el *Dissertation Abstracts International*. Incluye información acerca de 1.300 revistas y algo de literatura gris. También es conocido con el nombre de PsycLIT (nombre comercial de la versión en CD-ROM).

REDINET. Bases de datos de información educativa sobre investigación, innovación y recursos didácticos. Está producida por el *CIDE* y diferentes Consejerías de Educación de las comunidades autónomas.

REDUC. Base de datos que recopila la documentación sobre Educación de América Latina y el Caribe. Contiene resúmenes de investigaciones, experiencias y estudios sobre Educación que son recopilados y procesados por el Centro de Investigación y Desarrollo de la Educación y algunos centros asociados. El formato en papel corresponde al repertorio *Resúmenes Analíticos de Educación*.

TESEO. Es una base de datos multidisciplinar y no una base de datos específica de Educación. Fue creada en 1976 por el Ministerio de Educación y Ciencia español y contiene las referencias bibliográficas y resúmenes de las tesis doctorales leídas en las universidades del Estado español desde 1976. Está estructurado con la clasificación de la Unesco, por universidades y año académico.

Bibliotecas, organizaciones, asociaciones y centros de investigación y evaluación nacionales

Las direcciones que presentamos a continuación en este apartado y los siguientes se refieren a bibliotecas, organizaciones, asociaciones, centros de evaluación e investigación junto con una relación de recursos sobre *assessment*, evaluación y medición, tanto a nivel nacional como internacional:

AIDIPE. Asociación Interuniversitaria de Investigación Pedagógica.
<http://www.uv.es/aidipe/>

Biblioteca de la UNED. Consulta a la base de datos de tesis doctorales de la Universidad Nacional de Educación a Distancia.

Bibliotecas españolas y del mundo. Mapa de bibliotecas españolas y del mundo accesibles desde internet.
<http://exlibris.usal.es/bibesp/>

BNE. Biblioteca Nacional. <http://www.bne.es>

CIDE. Centro de Investigación y Documentación Educativa del Ministerio de Educación, Cultura y Deporte.

CSDA. Cataluña: Consell Superior d'Avaluació del Sistema Educatiu.
<http://www.gencat.net/ense/csda/>

CSIC. Consejo Superior de Investigaciones Científicas.
<http://www.csic.es>

ICEC. Instituto Canario de Evaluación y Calidad Educativa.
<http://www.gobiernodecanarias.org/educacion/icec/scripts/default.asp>

IDEA. Instituto de Evaluación y Asesoramiento Educativo.
<http://www.idea-educa.com/>

IEDCYT. Instituto de Estudios Documentales sobre Ciencia y Tecnología (antiguo CINDOC)
<http://www.cindoc.csic.es/info/infobjetivos.html>

Instituto de Evaluación
<http://www.institutodeevaluacion.mec.es/>

IVECE. Instituto Valenciano de Evaluación y Calidad Educativa.
<http://www.edu.gva.es/eva/docs/calidad/publicaciones/es/cd/iframe.html>

IVEI/ISEI. Instituto Vasco de Evaluación e Investigación Educativa.
<http://www.isei-ivei.net>

REBIUN. Red de Bibliotecas Universitarias Españolas.
<http://www.csic.es/cbic/webuni.htm>

RedIRIS. Red Académica y de Investigación Española.
<http://www.rediris.es>

SEP. Sociedad Española de Pedagogía.
<http://www.uv.es/soespe/>

Bibliotecas, organizaciones, asociaciones y centros de investigación y evaluación internacionales

ACER. The Australian Council for Educational Research. Australia.
<http://www.acer.edu.au>

AERA. American Educational Research Association. <http://aera.net>

AIR. Association for Institutional Research. <http://www.airweb.org>

American Evaluation Association.
<http://www.eval.org/ListsLinks/EvaluationLinks/htm>

BERA. British Educational Research Association.
<http://www.bera.ac.uk/>

Bibliothèque Nationale de France. <http://www.bnf.fr/>

CEBIAE. Centro Boliviano de Investigación y Acción Educativas.
<http://www.cebiae.edu.bo/>

Center for Research and Documentation of O.L.M.E. Centro de investigación y documentación educativa. <http://www.otenet.gr/kemete/>

Centro de Investigação em Educação. Centro portugués de carácter interdisciplinar que, dentro del Instituto de Educação e Psicologia de la Universidade do Minho (Portugal), se dirige a coordinar investigación científica en educación y dominios afines. <http://www.iep.uminho.pt/cied/>

Centro de Psicopedagogía. Centro de investigación de la Faculdade de Psicologia e Ciências da Educação de la Universidade de Coimbra (Portugal) dedicado a estimular la investigación en las áreas de la Psicología y la Pedagogía, realizar estudios normativos, colaborar con otros centros de investigación, etc. <http://www.uc.pt/psicop/>

CERI. Center for Educational Research and Innovation centro de investigación e innovación educativas de la OCDE. http://www.oecd.org/department/0,3355,en_2649_35845581_1_1_1_1_1,00.html

CNDP. Centre National de Documentation Pedagogique. Francia. Documentación pedagógica de Francia. <http://www.cndp.fr/accueil.htm>

CNE. Comité National d'Évaluation (Francia). <http://www.cne-evaluation.fr/>

COMIE. Consejo Mexicano de Investigación Educativa. <http://www.comie.org.mx/v1/sitio/portal.php>

CREDE. Center for Research on Education, Diversity & Excellence. Enseñanza de la lengua y la atención a la diversidad. <http://www.cal.org/crede/>

CREP. Center for Research in Educational Policy. Centro de investigación sobre políticas educativas. http://www.people.memphis.edu/~coe_crep/

CRESST. National Center for Research on Evaluation, Standards and Student Testing. <http://www.cse.ucla.edu/>

CSTEEP. Center for the Study of Testing, Evaluation and Educational Policy. <http://www.bc.edu/research/csteep/>

DEP. Direction de l'Évaluation et de la Prospective (Francia). <http://www.education.gouv.fr/stateval/default.htm>

DiNIECE. Dirección Nacional de Información y Evaluación de la Calidad Educativa de Argentina. <http://diniece.me.gov.ar/>

EERA. European Educational Research Association. <http://www.eera.ac.uk>

ERIC. Educational Resources Information Center. <http://www.eric.ed.gov>

ESERA. European Science Education Research Association.

<http://www.naturfagsenteret.no/esera/>

ETS. Educational Testing Service.

<http://www.ets.org>

EVA. The Danish Evaluation Institute.

<http://www.eva.dk/>

EU-AEI. European network for educational research on Assessment, Effectiveness and Innovation. <http://www.pjb.co.uk/npl/bp9.htm>

EURYDICE. Red Europea de Información sobre la Educación en Europa.

<http://www.eurydice.org/>

GAVE. Gabinete de Avaliação Educacional (Portugal).

<http://www.gave.pt/>

GRADE. Grupo de Trabajo de Estándares y Evaluación que coordina el Grupo de Análisis para el Desarrollo.

http://www.gradeworkinggroup.org/_ES/index.htm

HCéé. Haut Conseil de l'Évaluation de l'École (Francia).

<http://cisad.adc.education.fr/hcee/>

IBM. Personal learning System. <http://www.es.ibm.com/>

IEA. International Association for the Evaluation of Educational Achievement. <http://www.iea.nl/>

IIEP. International Institute for Educational Planning.

<http://www.unesco.org/iiep/index.htm>

INEE. Instituto Nacional para la Evaluación de la Educación (México).

<http://www.inee.edu.mx/>

INRP. Institut National de Recherche Pédagogique.

<http://www.inrp.fr/>

Institut Romand de Recherches et de Documentation Pédagogiques.

<http://www.unine.ch/irdp/>

Institute of Education.

<http://ioe.websserver.ioe.ac.uk/ioe/index.html>

INValSI. Istituto Nazionale per la Valutazione del Sistema dell'Istruzione (Italia). <http://www.invalsi.it/>

LLECE. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. <http://llece.unesco.cl/esp/>

NCES. National Center for Education Statistics (EE.UU.).

<http://nces.ed.gov/>

NFER. National Foundation for Educational Research. Organismo independiente que lleva a cabo investigación educativa. Desarrolla proyectos en todos los sectores del sistema educativo público del Reino Unido.
<http://www.nfer.ac.uk/>

NZCER. New Zealand Council for Educational Research. Consejo neozelandés sin ánimo de lucro dedicado a la investigación educativa.
<http://www.nzcer.org.nz/>

CARE. Centre for Applied Research on Education.
<http://research.edu.uea.ac.uk/care>

OERI. Office of Educational Research and Improvement.
<http://www.ed.gov/offices/OERI/index.html>

PISA Project. El proyecto PISA.
<http://www.pisa.oecd.org/>

Red Europea de Responsables de Evaluación de los Sistemas Educativos.
<http://www.reva-education.eu/>

TIMMS. Estudio de Tendencias en Matemáticas y Ciencias.
<http://nces.ed.gov/TIMSS/>

Unesco. Centro de información e investigación sobre educación comparada de la Unesco. <http://www.ibe.unesco.org/>

Recursos informáticos

Recursos sobre Assessment

CARLA: Centre for Advanced Research on Language Acquisition.
<http://www.carla.umn.edu/index.html?go>

Educational Assessment.
<http://bubl.ac.uk/link/e/educationalassessment.htm>

Educational Assessment in Wisconsin.
<http://www.dpi.state.wi.us/dpi/oea/assessmt.html>

Educational Assessment Project (EAP).
<http://www.meoserrc.org/eap/>

Educational Technology Assessment Office.
<http://education.boisestate.edu/itce/>

ERIC Clearinghouse on Assessment and Evaluation.
<http://ericae.net/>

Index to ERIC Sites.

<http://www.ed.gov/help/site/expsearch/eric.html>

Individual Development Educational Assessment.

<http://www.idea.ksu.edu/>

International Association for Educational Assessment (IAEA).

<http://www.iaea.info/>

Internet Resources for Higher Education Outcomes Assessment.

<http://www2.acs.ncsu.edu/UPA/assmt/resource.htm>

Internet Resources for Institutional Research: Assessment.

<http://apollo.gmu.edu/~jmilam/air95/assess.html>

Maine Educational Assessment (MEA).

<http://www.maine.gov/portal/education/k12.html>

Michigan Educational Assessment Program (MEAP)

http://www.michigan.gov/mde/0,1607,7-140-22709_31168---,00.htm

Neurodevelopmental & Educational Assessment Center.

<http://www.neuroed.com/>

Northwest Regional Educational Laboratory. Assessment & Accountability:

<http://www.nwrel.org/assessment>

Office of Educational Assessment. University of Washington.

<http://www.washington.edu/oea/>

OFSTED: Office for Standards in Education

<http://www.ofsted.gov.uk/>

Regional Educational Assessment & Consultation Team.

<http://www.crec.org/tabs/brokering/react.php>

Standards for Teacher Competence in Educational Assessment of Students.

<http://www3.interscience.wiley.com/journal/119380747/abstract?CRET RY=1&SRETRY=0>

Teacher Quality: A Report on Teacher Preparation and Qualifications of Public School Teachers.

<http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=1999080>

Recursos sobre Evaluación

American Evaluation Association. <http://www.eval.org/>

Counseling Education Research and Evaluation in the Behavioral Sciences.

<http://www.ohiou.edu/che/Ce-res.html>

Course Listings: Education-Research and Evaluation.
<http://www.ohiou.edu/gcatalog/97-99/courses/edre.htm>

Développement des Méthodologies d'Evaluation en Education en Europe
(Homepage de l'ADMEE)
<http://www.unine.ch/irdp/admee/welcome.html>.

Education-Right of Independent Education Evaluation (IEE)
http://www.ucp.org/ucp_channel.doc.cfm/1/12/69/69-69/705

Education Evaluation and Employment Credential Services. American
Evaluation Institute Our Evaluators.
<http://www.educationevaluations.com/>

Education Evaluation, Research and Statistics.
<http://home.universalclass.com/education/educationevaluation/>

Educator's Guide to Evaluating the Use of Technology in.
<http://www.ed.gov/pubs/EdTechGuide/>

Evaluation Research - Special Education.
<http://www.wested.org/cs/we/view/feat/27>

Evaluation Sites. <http://gsociology.icaap.org/methods/redother.htm>

EvNet: Network for the Evaluation of Education and Training.
<http://www.socsci.mcmaster.ca/srnet/evnet.htm>

Extension Education Evaluation.
<http://danr.ucop.edu/eee-aea/aeapaper.html>

Higher Education Center: Home Research and Evaluation of Social
Norms Campaigns.
<http://edc.higheredcenter.org/socialnorms/research.html>

Impact Evaluation of Education Projects: Decentralization and Privatiza-
tion Issues.
<http://www.worldbank.org/html/dec/Publications/Abstracts97/02labor/labor2.html>

Institute for Educational Research. <http://www.jyu.fi>

International Education Research Foundation.
<http://www.ierf.org/aboutierf.asp>

MH School California Social Studies Research.
<http://mhcalifornia.com/socialstudies/research/contreras.html>

National Centre for Vocational Education Research. The University of
Western Australia.
<http://www.acs.uwa.edu.au/research/ops/sum/grants/deetyaantarac.html>

Office of Education and Research Evaluation.
<http://www.aecom.yu.edu/desm/epbstat/edu.htm>

Planning and Evaluation Service Home Page.
<http://www.ed.gov/offices/OUS/PES/>

Research, Evaluation & Education-Papers.
<http://www.nnah.org.uk/membership/subforums/research/articles/papers/default.asp>

Scottish Council for Research in Education (SCRE).
<http://www.scre.ac.uk/Homepage.html>

Special Education: EVALUATION.
<http://www.fape.org>

Test and Evaluation Research and Education Center (TEREC).
<http://www.terec.gatech.edu/>

The Department of Education Psychology. The Research, Evaluation, Measurement, and Statistics (REMS) Program.
<http://www.coe.uga.edu/edpsych/remis/>

The Evaluation Center at Human Services Research Institute (HSRI).
<http://www.hsri.org/>

The evaluation of education systems.
<http://www.ibe.unesco.org>

University of Virginia Library.
<http://www.lib.virginia.edu/index.html>

Recursos sobre Medición Educativa

BUROS Institute of Mental Measurement. <http://www.unl.edu/buros>

CITO: OPML one parameter logistic model.
http://www.cito.com/research_and_development/pyschometrics/psychometric_software/oplm.aspx

Code of Professional Responsibilities in Educational Measurement.
http://www.natd.org/Code_of_Professional_Responsibilities.html

Comparison of PIAT-R,K-TEA,WRMT-R.
<http://www.agsnet.com/news/edachcompare.asp>

Computer Adaptive Testing Project.
<http://www.carla.umn.edu/assessment/CAT.html>

Constructing written test questions for the basic and clinical sciences.
<http://www.nbme.org/publications/index.html>

CRESST-UCLA National Center for Research on Evaluation, Standards, and Student Testing.

<http://www.cse.ucla.edu/>

EdMeas Page Links a Medición Educativa.

<http://www.angelfire.com/me4/andrews/EdMeas.html>

Educational Measurement and Statistics. University of Iowa.

<http://www.education.uiowa.edu/measstat/>

EDUTEST. <http://www.edutest.com/>

ETS Net: Educational Testing Service Network. <http://www.ets.org/>

Fair Test: The National Center for Fair and Open Testing.

<http://www.fairtest.org/>

Hardcourt. <http://www.harcourt.com/>

HyperStat online. <http://davidmlane.com/hyperstat/index.html>

Resources in Language Testing.

<http://www.le.ac.uk/education/testing/ltr.html>

Item Response Theory for Unfolding.

<http://www.psychology.gatech.edu/Unfolding/>

IUPUI Testing Center, Indianapolis.

<http://assessment.iupui.edu/testing/>

LXR-Test Home Page <http://www.lxrtest.com/>

Michigan Educational Assessment Program (MEAP).

http://www.michigan.gov/mde/0,1607,7-140-22709_31168---,00.html

National Center for Research on Evaluation, Standards, and Student Testing. <http://www.cse.ucla.edu>

National Council on Measurement in Education. <http://www.ncme.org/>

NTC Educational Measurement Center.

http://www.cblohm.com/news/NTC/NTC_080124/

On line personality test. <http://www.keirsey.com/>

Online IQ test. <http://www.psychtests.com/iq.html>

Princeton Review online. <http://www.review.com/about/privacy.cfm>

Research and Evaluation, Department of Educational Psychology.

<http://www.uwm.edu/Dept/EdPsych/repq.html>

Riverside Publishing Products.

<http://www.riverpub.com/products/index.html>

School of Educational Studies.

http://www.stou.ac.th/Eng/Courses/Edu_stu/bach_eva.html

StatLib. <http://lib.stat.cmu.edu/>

STATOO-Statistics, Datamining, Stockmarket analysis.

<http://www.statoo.com/>

Tests. <http://www.queendom.com/>

The Statistical Laboratory for Educational and Psychological Measurement. University of Illinois.

<http://www.stat.uiuc.edu/psychometrics/index.html>

Umeå University. The Faculty of Social Sciences Department of Educational Measurement. Universidad de Umeå, Suecia.

http://www8.umu.se/edmeas/index_eng.html

USF Educational Measurement and Research.

<http://www.ugs.usf.edu/catalogs/9900/EDMEAS.HTM>

Bibliografía

- Abalde Paz, E.; Muñoz Cantero, J. M. y Ríos de Deus, M. P. (2002): «Evaluación docente vs. evaluación de la calidad», ponencia presentada en el Tercer Congreso Virtual de AIDIPE. <http://www.uv.es/aidipe/CongVirtual3/Ponencia3.htm>.
- Aguilar Idáñez, M. J. y Ander-Egg, E. (1994): *Evaluación de servicios y programas sociales*, Buenos Aires, Lumen.
- Alfaro, I. J. (1994): «Diagnóstico y evaluación», *Revista de Investigación Educativa*, n.º 23, pp. 577-583.
- Alfaro, M. E. (1990): «Aspectos prácticos del proceso de programación y evaluación», *Documentación Social*, n.º 81, pp. 65-80.
- Aliaga Abad, F. (200): *Bases epistemológicas y proceso de investigación psicoeducativa*, Valencia, CSV.
- Alkin, M. C. (1969): «Evaluation theory development», *Evaluation comment*, n.º 2, p. 1.
- Álvarez González, M. y Bisquerra, R. (coord.) (1996): *Manual de orientación y tutoría*, Barcelona, Praxis.
- Álvarez Méndez, J. M. (2001): *Evaluar para conocer, examinar para excluir*, Madrid, Morata.
- Álvaro Page, M. et al. (1990): *Hacia un modelo causal del rendimiento académico*, Madrid, Centro de Publicaciones del MEC-CIDE.
- Alvira Martín, F. (1982): «Las perspectivas cualitativa y cuantitativa en las investigaciones sociales», *Estudios de Psicología*, n.º 11, pp. 34-36.
- (1985): «La investigación evaluativa: una perspectiva experimentalista», *Revista Española de Investigaciones Sociológicas*, n.º 29.
- (1991): *Metodología de la evaluación de programas*, Madrid, Centro de Investigaciones Sociológicas.

- Amezcu Viedma, C. (1996): «Introducción a la evaluación», en C. Amezcu Viedma y A. Jiménez Lara, *Evaluación de programas sociales*, Madrid, Díaz Santos, pp. 1-14.
- Anastasi, A. y Urbina, S. (1998): *Tests psicológicos*, México, Prentice Hall.
- Anguera, M. T. (1982): *Metodología de la observación en las ciencias humanas*, Madrid, Cátedra.
- (1989): *Metodología de la observación en las Ciencias Humanas*, Madrid, Cátedra.
- (1995): «Diseños», en Fernández-Ballesteros, R. (ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*, pp. 149-172.
- Arnal, J.; Del Rincón, D. y Latorre, A. (1992): *Investigación educativa. Fundamentos y metodología*, Barcelona, Labor.
- Arnau, J. (1982): *Psicología experimental. Un enfoque metodológico*, México D.F., Trillas.
- Ary, D.; Jacobs, L. Ch. y Razauvieh, A. (1987): *Introducción a la investigación pedagógica*, México, Interamericana.
- (1996): *Introduction to research in education*, Nueva York, Harcourt Brace College Publishers.
- Asociación Nacional de Universidades e Instituciones de Educación Superior. (1984): *La evaluación de la educación superior en México*, México, Asociación Nacional de Universidades e Instituciones de Educación Superior.
- Ballester Brage, L. (2000): «Delphi taulak: hezitzaileentzako ikerkuntzametodologia bat», *Tantak*, n.º 24, pp. 65-94.
- Barbier, R. (1996): *La Recherche Action*, París, Anthropos.
- Bartolomé, M. (1994): «La investigación cooperativa», en V. García Hoz (dir.), *Problemas y métodos de investigación en enseñanza personalizada*, Madrid, Rialp, pp. 376-403.
- Batanaz Palomares, L. (1996): *Investigación y diagnóstico en educación: una perspectiva psicopedagógica*, Marazana, Granada, Aljibe.
- Beeby, C. E. (1977): «The meaning of evaluation», *Current Issues in Education. n.º 4: Evaluation*, Wellington, Department of Education, pp. 68-78.
- Bélair, L. M. (2000): *La evaluación en la acción. El dossier progresivo de los alumnos*, Sevilla, Díada.
- Belloch, C. (2000): *Proyecto Docente presentado en el concurso ordinario para la plaza de Titular del área de Métodos de Investigación y Diagnóstico en Educación*, Universidad de Valencia.
- Bericat, E. (1998): *La integración de los métodos cuantitativo y cualitativo en la investigación social*, Barcelona, Ariel Sociología.
- Berk, R. A. (ed.) (1981): *Educational evaluation methodology. The state of the art*, Londres, The Johns Hopkins University Press.
- Best, J. W. (1972): *Cómo investigar en educación*, Madrid, Morata.
- Bickman, L. y Rog, D. J. (1998): «Introduction: Why a handbook of applied social research methods?», en L. Bickman y D. J. Rog (eds.) *Handbook of applied social research methods*, Thousand Oaks (CA), Sage, pp. IX-XIX.
- Bisquerra, R. (1989): *Métodos de investigación educativa*, Barcelona, Ceac.
- Blanco Prieto, F. (1994): *La evaluación en la educación secundaria*, Salamanca, Amarú.
- Bloom, B. S. et al. (1975): *Evaluación del aprendizaje*, Buenos Aires, Troquel.

- Bordas Alsina, I. (1995): «La evaluación de programas para el cambio», en A. Medina Rivilla y L. M. Villar Angulo (coor.): *Evaluación de programas educativos, centros y profesores*, Madrid, Universitas, pp. 176-205.
- (1999): «La evaluación de programas para la innovación», en B. Jiménez Jiménez (ed.): *Evaluación de programas, centros y profesores*, pp. 237-264.
- Botella, J. y Barriopedro, M. I. (1995): «Análisis de datos», en R. Fernández-Ballesteros (ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*, pp. 173-203.
- Bramley, P. (1996): *Evaluating training effectiveness*, Londres, McGraw Hill.
- Brannen, J. (2005): «Mixed methods research: A discusión paper». *International Journal of Social Research Methodology: Theory and Practice*. v. 8, n.º 3. pp. 173-185.
- Brinkerhoff, R. O.; Brethower, D. M.; Hluckyj, T. y Nowakosky, J. R. (1983): *Program evaluation: A practitioners guide for trainers and educators*, Boston, KluwerNijhoff Publishing.
- Burgess, R. G. (1993): *Research methods*, Walton-On-Thames, Thomas Nelson.
- Cabrera, F. (1986): *Proyecto Docente presentado en el concurso ordinario para la plaza de Profesor Titular del área de Métodos de Investigación y Diagnóstico en Educación (Técnicas de medición y evaluación educativas)*, Universidad de Barcelona.
- (1987): «La investigación evaluativa en la educación», en E. Gelpi *et al.* (comp.), *Técnicas de evaluación y seguimiento de programas de formación profesional*, Barcelona, Fundación Largo Caballero.
- (2000a): *Evaluación de la formación*, Madrid, Síntesis.
- (2000b): «Iruidi soziometrikoak eta haien interpretazioa soziograma baten azterketan» *Tantak*, n.º 23, pp. 83-102.
- Campbell, D. T. y Stanley, J. C. (1966): *Experimental and quasi-experimental designs for research*, Chicago, Rand McNally.
- Carballo Santaolalla, R. (1990a): «Evolución del concepto de evaluación: desarrollo de los modelos de evaluación de programas», *Bordón*, v. 42, n.º 4, pp. 423-431.
- (1990b): *Evaluación de un programa de formación ocupacional*, Madrid, Editorial de la Universidad Complutense de Madrid.
- Cardona Moltó, M. C. (2002): *Introducción a los métodos de investigación en educación*, Madrid, EOS.
- Caride Gómez, J. A. (2002): «Evaluación de programas sociales. Presentación», en F. Trillo Alonso (coor.), *Evaluación*, Barcelona, Praxis, pp. 145-147.
- Carr, W. (1989): «¿Puede ser científica la investigación educativa?», *Investigación en la Escuela*, n.º 7, pp. 37-47.
- y Kemmis, S. (1988): *Teoría crítica de la enseñanza*, Barcelona: Martínez Roca.
- Carreño, H. F. (1985): *Enfoques y principios teóricos de la evaluación*, México, Trillas.
- Carrión Carranza, C. (2001): *Valores y principios para evaluar la educación*, México, Paidós.
- Casanova Rodríguez, M. A. (1995a): *Manual de evaluación educativa*, Madrid, La Muralla.
- Castillo Arredondo, S. (1994): «Diagnóstico-evaluación: cooperación y complementariedad», *Revista de Investigación Educativa*, n.º 23, pp. 640-642.
- Cea, M. A. (1996): *Metodología cuantitativa. Estrategias y técnicas de investigación social*, Madrid, Síntesis.

- Cohen, L. y Manion, L. (1990): *Métodos de investigación educativa*, Madrid, La Muralla.
- Colás Bravo, P. (1997): «Enfoques en la metodología cualitativa: sus prácticas de investigación», en L. Buendía Eisman, M. P. Colás Bravo y F. Hernández Pina (eds.), *Métodos de investigación en psicopedagogía*, Madrid, McGraw-Hill, pp. 225-249.
- y Buendía Eisman, L. (1998): *Investigación educativa*, Sevilla, Alfar, 3.ª edición.
- y Rebollo, M. A. (1993): *Evaluación de programas. Una guía práctica*, Sevilla, Kronos.
- Coleman, J. S. et al. (1966): *Equality of Educational Opportunity*, Washington, US Government Printing Office.
- Cook, T. D. y Reichardt, Ch. S. (eds.) (1986): *Métodos cualitativos y cuantitativos en investigación evaluativa*, Madrid, Morata.
- Crocker, L. y Algina, J. (1986): *Introduction to classical and modern test theory*, Nueva York, Holt, Rinehart and Winston.
- Cronbach, L. J. (1963): «Course improvement through evaluation», *Teachers College Record*, n.º 64, pp. 672-683.
- (1980): *Toward reform of program evaluation*, San Francisco, Jossey-Bass Publishers.
- (1982): *Designing evaluations of educational and social programs*, San Francisco, Jossey-Bass.
- Chadwick, C. B. y Rivera, N. (1991): *Evaluación formativa para el docente*, Buenos Aires, Paidós.
- Choppin, B. H. (1990): «Evaluation, assessment and measurement», en H. J. Walberg y G. D. Haertel (eds.), *The international encyclopedia of educational evaluation*, Oxford, Pergamon Press.
- Darder, P. y Mestres, J. (1994): «Control, evaluación e innovación», en J. Garín y P. Darder (coor.), *Organización y gestión de los centros educativos*, Barcelona, Praxis.
- De la Orden Hoz, A. (1977): «Prólogo a la edición española» en P. D. Lafourcade, *Evaluación de los aprendizajes*, Madrid, Cincel, pp. 13-17.
- (1982): *La evaluación educativa*, Buenos Aires, Docencia-Proyecto CINAIE.
- (1985a): «Investigación evaluativa», en A. De la Orden Hoz (ed.), *Investigación educativa. Diccionario de Ciencias de la Educación*, Madrid, Anaya.
- (1985b): «Hacia una conceptualización del producto educativo», *Revista de Investigación Educativa*, v. 3, n.º 6, pp. 271-283.
- (1989): «Investigación cuantitativa y medida en educación», *Bordón*, v. 41, n.º 2, pp. 217-235.
- (1991): «La evaluación de programas en el contexto educativo», en *Seminario sobre evaluación de programas educativos*, Madrid.
- (2000): «La función optimizante de la evaluación de programas evaluativos», *Revista de Investigación Educativa*, v. 18, n.º 2, pp. 381-389.
- De Miguel Díaz, M. (1988): «Paradigmas de la investigación educativa española», en I. Dendaluze (coor.): *Aspectos metodológicos de la investigación educativa*, Madrid, Narcea, pp. 60-77.
- (1995): «Indicadores de rendimiento y evaluación de programas», en R. Pérez Juste, J. L. García Llamas y C. Martínez Mediano (coor.): *Evaluación de programas y centros educativos*, Madrid, UNED, pp. 169-192.
- (1997): «La evaluación de los centros educativos», en H. Salmerón (ed.) *Evaluación educativa. Teoría, metodología y aplicaciones en áreas de conocimiento*, Granada, Grupo Editorial Universitario, pp. 151-174.

- (2000a): «La evaluación de programas sociales: fundamentos y enfoques teóricos», *Revista de Investigación Educativa*, v. 18, n.º 2, pp. 289-317.
- (2000b): «Modelos y diseños en la evaluación de programas», en L. Sobrado (ed.), *Orientación profesional: Diagnóstico e inserción*, Barcelona, Estel, pp. 193-215.
- Del Rincón, D. (1996): «Ebaluazioan informazioa bildu eta analizatzeko estrategia kualitatiboak», *Tantak*, n.º 16, pp. 7-35.
- ; Arnal, J.; Latorre, A. y Sans, A. (1995): *Técnicas de investigación en ciencias sociales*, Madrid, Dykinson.
- Delors, J. et al. (1996): *La educación encierra un tesoro*. Informe a la Unesco de la Comisión Internacional sobre la Educación para el Siglo XXI, París, Ediciones Unesco.
- Dendaluze, I. (1988): «Una reflexión metodológica sobre la investigación educativa», en L. Dendaluze (coord.): *Aspectos metodológicos de la investigación educativa*, Madrid, Narcea, pp. 13-41.
- (1995): «Avances en los métodos de investigación educativa en la intervención psicopedagógica», *Revista de Investigación Educativa*, n.º 26, p. 932.
- (2001): «Ciencias Sociales e investigación en Euskal Herria», *Actas del XV Congreso de Estudios Vascos: Ciencia y cultura vasca, y redes telemáticas*, Donostia, pp. 1139-1158.
- Denzin, N. K. y Lincoln, Y. S. (eds.) (1994): *Handbook of Qualitative Research*, Thousand Oaks, Sage.
- D'Hainaut, L. (1988): *Sistemas educativos. Análisis y regulación*, Madrid, Narcea.
- Doval, L. (1995): «Aportaciones al diagnóstico en educación. De la historiografía al concepto», *Revista de Investigación Educativa*, n.º 26, pp. 201-216.
- Eisner, E. W. (1976): «Educational connoisseurship and criticism: Their form and functions in educational evaluation», *Journal of Aesthetic Education*, v. 10, n.º 3/4, pp. 135-150.
- (1985): *The art of educational evaluation. A personal view*, Londres, The Falmer Press.
- Elliott, J. (1990): *La investigación-acción en educación*, Madrid, Morata.
- Escudero Escorza, T. (2003): «Desde los tests hasta la investigación evaluativa actual. Un siglo, el xx, de intenso desarrollo de la evaluación en educación», *Revista Electrónica de Investigación y Evaluación Educativa (RELIEVE)*, v. 9, n.º 1. http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm.
- Escudero Muñoz, J. M. (1990): «Tendencias actuales en la investigación educativa», *Currículo* v. 1, n.º 2, pp. 2-25.
- Fernández Cano, A. (1996): *Métodos para evaluar la investigación en psicología y pedagogía*, Madrid, Síntesis.
- Fernández Pérez, M. (1986): *Evaluación y cambio educativo: El fracaso escolar*, Madrid, Morata.
- Fernández-Ballesteros, R. (1985): «Evaluación psicológica y evaluación valorativa», *Evaluación psicológica*, v. 1, n.º 1/2, pp. 7-31.
- (1995): «El proceso de evaluación de programas», en R. Fernández-Ballesteros (ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*, Madrid, Editorial Síntesis, pp. 75-113.
- Ferrández Arenaz, A. (1993): «Diseño y proceso de la evaluación de adultos», en A. Ferrández et al. (coord.), *La evaluación de la educación de personas adultas*, Madrid, Diagrama.

- Ferrer, F. (1999): «L'avaluació del sistema educatiu. Aproximació teòrica des de i per Catalunya», en *Avaluació i Educació*, Barcelona, Generalitat de Catalunya. Consell Escolar de Catalunya.
- Fetterman, D. M. (1989): *Ethnography. Step by Step*, Newbury Park, California, Sage.
- Fornier, A. y Latorre, A. (1996): *Diccionario terminológico de investigación educativa y psicopedagógica*, Barcelona, EUB.
- Forns Santacana, M. (1980): «La evaluación del aprendizaje», en C. Coll y M. Forns Santacana (eds.), *Áreas de intervención de la psicología*, Barcelona, Horsori, pp. 107-141.
- Fox, D. J. (1981): *El proceso de la investigación en Educación*, Pamplona, Eunsa.
- Fraenkel, J. R. y Wallen, N. E. (2000): *How to design and evaluate research in education*, Nueva York, McGraw-Hill, tercera edición.
- Franco, R. (1971): *Algunas reflexiones sobre la evaluación del desarrollo*, Santiago de Chile, Ilpes.
- Fuentes, A. (1986): «Los modelos causales en la investigación del rendimiento académico», *Revista de Investigación Educativa*, v. 4, n.º 7, pp. 35-48.
- García García, M. (2000): «Evaluación y calidad de los sistemas educativos», en T. González Ramírez (coor.), *Evaluación y gestión de la calidad educativa. Un enfoque metodológico*, Málaga, Aljibe, pp. 227-267.
- García Garrido, J. L. (1996): *Diccionario europeo de la educación*, Madrid, Dykinson.
- García Hoz, V. (1975): *Principios de pedagogía sistemática*, Madrid, Rialp.
- García Jiménez, E. (1994a): «¿Diagnóstico versus evaluación?», *Revista de Investigación Educativa*, n.º 23, pp. 585-587.
- (1994b): «Investigación Etnográfica», en V. García Hoz (dir.), *Problemas y métodos de investigación en educación personalizada*, Madrid, Rialp, pp. 343-375.
- ; Gil Flores, J. y Rodríguez Gómez, G. (1999a): «Elaboración de un TAI sobre contenidos escolares», en J. Olea, V. Ponsoda y G. Prieto (eds.), *Tests informatizados. Fundamentos y aplicaciones*, Madrid, Pirámide.
- ; — y — (1999b): «La evaluación de tests adaptativos informatizados», *Revista Electrónica de Investigación y Evaluación Educativa*, v. 4, n.º 2.
- ; López Górriz, I.; Moreno Sánchez, E. y Padilla Carmona, M. T. (2002): «Otras técnicas y estrategias para el diagnóstico en educación» en M. T. Padilla Carmona, *Técnicas e instrumentos para el diagnóstico y la evaluación educativa*, Madrid, CCS, pp. 259-316.
- García Llamas, J. L. (1995): «Evaluación de programas educativos», en R. Pérez Juste, J. L. García Llamas y C. Martínez Mediano (coor.), *Evaluación de programas y centros educativos*, Madrid, UNED, pp. 45-60.
- García Ramos, J. M. (1989): *Bases pedagógicas de la evaluación. Guía práctica para educadores*, Madrid, Síntesis.
- Gil Flores, J. (1994): *Análisis de datos cualitativos*, Barcelona, PPU.
- ; García Jiménez, E. y Rodríguez Gómez, G. (1997): «Testu-datu en azterketarako metodo kualitatibo eta kuantitatiboak», *Tantak*, n.º 17, pp. 77-107.
- y Perera Rodríguez, V. H. (2001): *Análisis informatizado de datos cualitativos. Introducción al uso del programa NUD.IST-5*, Sevilla, Kronos.
- Gimeno Sacristán, J. (1985): *Teoría de la enseñanza y desarrollo del currículo*, Madrid, Anaya.
- Glaser, R. (1963): «Instructional technology and the measurement of learning outcomes: some questions», *American Psychologist*, n.º 18, pp. 519-521.

- Goetz, J. y LeCompte, M. D. (1988): *Etnografía y diseño cualitativo en investigación educativa*, Madrid, Morata.
- González Such, J. (2000): *Proyecto Docente presentado en el concurso ordinario para la plaza de Profesor Titular del área de Métodos de Investigación y Diagnóstico en Educación (Medición y evaluación educativas: Bases de medición educativa)*, Universidad de Valencia.
- Goyette, G. y Lessard-Hebert, M. (1988): *Investigación-acción: Funciones, fundamentos e instrumentación*, Barcelona, Laertes.
- Greene, J.C.; Caracelli, V.J. y Gram, W.F. (1989): «Toward a conceptual framework for mixed-method evaluation designs», *Educational Evaluation and Policy Analysis*, v.11, n° 3, pp. 255-274.
- Gronlund, N. E. (1985): *Measurement and evaluation in teaching*, Nueva York, Macmillan Publishing Company.
- Guba, E. G. (1989): «Criterios de credibilidad en la investigación naturalista», en J. Gimeno Sacristán y A. Pérez Gómez (eds.), *La enseñanza: Su teoría y su práctica*, Madrid, Ediciones Akal, pp. 148-165.
- y Lincoln, Y. S. (1982): *Effective evaluation*, San Francisco, Jossey-Bass.
- y — (1989): *Fourth generation evaluation*, Newbury Park, California, Sage.
- y — (1994): «Competing paradigms in qualitative research», en N. Y. Denzin y Y. S. Lincoln (eds.), *Handbook of qualitative research*, Thousand Oaks, California, Sage, pp. 105-118.
- Hambleton, R. K. (1990): «Criterion referenced testing. Methods and practices», en T. Gutkin y C. Reynolds (eds.): *The handbook of school psychology*, Nueva York, Wiley.
- ; Swaminathan, H. y Rogers, H. J. (1991): *Fundamentals of item response theory*, Beverly Hills, California, Sage.
- Hammond, D. L. (1973): «Evaluation at the local level», en B. R. Worthen y J. R. Sanders, *Educational evaluation: Theory and practice*, Worthington, Ohio, Jones.
- Hernández López, J. M. (1995): «Procedimiento de recogida de información en evaluación de programas», en R. Fernández-Ballesteros (eds.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*, Madrid, Síntesis, pp. 117-147.
- Hernández Pina, F. (1995): *Bases metodológicas de la investigación educativa. I. Fundamentos*, Murcia, DM.
- ; Fuentes, P.; Iglesias, E. y Serrano, F. J. (1995): *Introducción al proceso de investigación en educación*, Murcia, DM.
- Hontangas, P. (1999): «Software para la construcción y administración de tests informatizados», en J. Olea, V. Ponsoda y G. Prieto (eds.), *Tests informatizados. Fundamentos y aplicaciones*, Madrid, Pirámide, pp. 251-286.
- House, E. R. (ed.) (1986): *New directions in educational evaluation*, Londres, The Falmer Press.
- Howe, K. R. (1985): «Two dogmas of Educational Research», *Educational Researcher*, v. 14, n.º 8, pp. 10-18.
- (1988): «Against the quantitative-qualitative incompatibility thesis or dogmas die hard», *Educational Researcher*, 17, 8, pp. 10-16.
- (1992): «Getting over the quantitative-qualitative debate», *American Journal of Education*, v. 100, n.º 2, pp. 236-255.
- Husén, T. y Postlethwaite, T. N. (eds.) (1985): *The International Encyclopedia of Education*, Oxford, Pergamon Press, [ed. cast.: *Enciclopedia Internacional de la Educación*, 10 volúmenes, Barcelona, VicensVives/MEC].

- Hutmacher, W. (1999): «L'avaluació en la transformació de les modalitats de govern dels sistemes educatius. Tendències europees», *Avaluació i Educació*, Barcelona, Generalitat de Catalunya. Cosell Escolar de Catalunya, pp. 15-34.
- ISEI/IVEI (2002): *Evaluación de 6.º curso de Educación Primaria 1999*, Vitoria Gasteiz, Servicio Central de Publicaciones del Gobierno Vasco.
- Jiménez Jiménez, B. (ed.) (1999a): *Evaluación de programas, centros y profesores*, Madrid, Síntesis.
- (1999b): «Presentación», en B. Jiménez Jiménez (ed.), *Evaluación de programas, centros y profesores*, Madrid, Síntesis, pp. 9-23.
- (1999c): «Evaluación de la docencia», en B. Jiménez Jiménez (ed.), *Evaluación de programas, centros y profesores*, Madrid, Síntesis, pp. 173-206.
- Jiménez Lara, A. (1996): «Diseño de la evaluación», en C. Amezcua Viedma y A. Jiménez Lara, *Evaluación de programas sociales*, Madrid, Díaz de Santos, pp. 91-103.
- Joint Committee on Standards for Educational Evaluation (1988): *Normas de evaluación para programas, proyectos y material educativo*, México, Trillas.
- (1994): *The program evaluation standards: How to assess evaluations of educational programs*. Beverly Hills, California, Sage.
- (1998): *Estándares para la evaluación de programas*, Bilbao, Mensajero.
- (2001): *Estándares de evaluación de personal. Cómo evaluar sistemas de evaluación de educadores*, Bilbao, Mensajero
- Jornet, J. M. (2001): *Proyecto Docente presentado en el concurso ordinario para la plaza de Catedrático del área de Métodos de Investigación y Diagnóstico en Educación (Diagnóstico en educación)*, Universidad de Valencia.
- Kaufman, R. y Engilsh, F. W. (1979): *Needs assessment. Concept and application*, Englewood Cliffs, Educational Technology.
- Kemmis, S. (1986): «Seven principles for programme evaluation in curriculum development and innovation», en E. R. House (ed.), *New directions in educational evaluation*, Londres, The Falmer Press, pp. 117-138.
- (1988): *El currículum. Más allá de la teoría de la reproducción*, Madrid, Morata.
- y McTaggart, R. (1988): *Cómo planificar la investigación-acción*, Barcelona, Laertes.
- Kerlinger, F. N. (1987): *Investigación del comportamiento. Técnicas y metodología*, México, Interamericana, segunda edición.
- (2002): *Investigación del Comportamiento: Un enfoque conceptual*. México D.F., Interamericana, cuarta edición.
- Ketetele, J. M. de y Roegiers, X. (1995): *Metodología para la recogida de información*, Madrid, La Muralla-Hespérides.
- Koetting, J. R. (1984): *Foundations of naturalistic inquiry: developing a theory base for understanding individual interpretations of reality*, Oklahoma, University Press.
- Koretz, D. (1992): «State and national assessment», en M. C. Alkin (ed.), *Encyclopedia of Educational Research*, Nueva York, McMillan, pp. 1262-1267.
- Krueger, A. R. (1991): *El grupo de discusión. Guía práctica para la investigación aplicada*, Madrid, Pirámide.
- Kuhn, T. S. (1982): *La estructura de la revoluciones científicas*, México, Fondo de Cultura Económica.
- Lafourcade, P. D. (1977): *Evaluación de los aprendizajes*, Madrid, Cincel.
- Landsheere, G. de (1996): *El pilotaje de los sistemas educativos*, Madrid, La Muralla.

- Latorre, A.; Del Rincón, D. y Arnal, J. (1996): *Bases metodológicas de la investigación educativa*, Barcelona, GR92.
- Lawson, M. J. (1997): «Experimental Studies», en J. P. Keeves (ed.), *Educational Research. Methodology, and measurement: an international handbook*, Oxford, Pergamon Press, pp. 126-134.
- Lázaro Martínez, A. J. (1990): «Problemas y polémicas en torno al diagnóstico pedagógico», *Bordón*, v. 42, n.º 1, pp. 7-15.
- (1991): «La formalización de indicadores de evaluación», *Bordón*, v. 43, n.º 4, pp. 477-494.
- (1994): «¿Se evalúa y/o se diagnostica?», *Revista de Investigación Educativa*, n.º 23, pp. 592-594.
- Le Boterf, G. (1985): *Cómo invertir en formación*, Barcelona, EADA GESTIÓN.
- Levine, H. M. (1975): «Cost-effectiveness in evaluation research», en M. Guttentag y E. L. Struening (eds.), *Handbook of Evaluation Research*, Beverly Hills, California, Sage.
- Lincoln, Y. S. (1995): «Emerging criteria for quality in qualitative and interpretative research», *Qualitative Inquiry*, v. 1, n.º 3, pp. 275-289.
- Livas González, I. (1980): *Análisis e interpretación de los resultados de la evaluación educativa*, México, Trillas.
- Lizasoain Hernández, L. y Joaristi Olariaga, L. (2000): «El análisis de datos en la evaluación de programas educativos», *Revista de Investigación Educativa*, v. 18, n.º 2, pp. 357-379.
- López Górriz, I. (1998): *Proyecto Docente presentado en el concurso ordinario para la plaza de Profesor Titular del área de Métodos de Investigación y Diagnóstico en Educación (Métodos de Investigación en Educación: Metodologías de Investigación Cualitativa y de Investigación-Acción)*. Universidad de Sevilla.
- López Mojarro, M. (1999): *A la calidad por la evaluación. La evaluación de centros docentes*, Madrid, Escuela Española.
- Lord, F. M. y Novick, M. R. (1968): *Statistical theories of mental test scores*, Reading, Massachusetts, AddisonWesley Publishing Company.
- Lukas, J. F.; Santiago, C.; Oihartzabal, L.; Pikabea, I.; Forner, A. y Latorre, A. (2000a): *Hezkuntza ikerkuntzarako hiztegia*, Donostia, Erein.
- ; —; Munarriz, B.; Moyano, N. y Sedano, M. (2000b): «Euskal Autonomia Erki-degoko ikastetxeen kanpoebaluazioa», *Tantak*, n.º 24, pp. 39-63.
- ; —; Sedano, M.; Moyano, N.; Joaristi, L. y Lizasoain, L. (2001): «Usos de la evaluación en profesores/as de la ESO de la Comunidad Autónoma Vasca», en AIDIPE (comp.), *Actas del X Congreso Nacional de Modelos de Investigación Educativa*, A Coruña, AIDIPE, pp. 427-433.
- Mager, R. F. (1975): *Medición del intento educativo*, Buenos Aires, Guadalupe.
- Marcelo, C. y Villar Angulo, L. M. (1994): «Sistema y métodos de evaluación de profesores», en L. M. Villar Angulo (coor.), *Manual de entrenamiento: Evaluación de procesos y actividades educativas*, Barcelona, PPU.
- Martínez González, R. A. (1994): «¿Diagnóstico pedagógico o evaluación educativa? Hacia una clarificación conceptual y terminológica», *Revista de Investigación Educativa*, n.º 23, pp. 623-630.
- Martínez Mediano, C. (1996): *Evaluación de programas educativos. Investigación evaluativa. Modelos de evaluación de programas*, Madrid, UNED.

- Massons, M. T. (1997): *Teoría i pràctica de l'avaluació dels serveis socials*, Barcelona, Diputació de Barcelona.
- Matarazzo, J. D. (1965): «The interview», en B. B. Wolman (ed.), *Handbook of clinical psychology*, Nueva York, McGraw-Hill.
- Mateo, J. (1986): *Proyecto Docente presentado en el concurso ordinario para la plaza de Catedrático del área de Métodos de Investigación y Diagnóstico en Educación (Medición y evaluación educativas)*. Universidad de Barcelona.
- (1998): «La evaluación educativa», en J. Mateo (dir.), *Enciclopedia general de la educación*, Barcelona, Océano, pp. 532-586.
- (2000a): *La evaluación educativa, su práctica y otras metáforas*, Barcelona, ICE - Horsori.
- (2000b): «L'avaluació de les institucions educatives», *Ponencia presentada en las I Jornadas de Medición y Evaluación Educativas*, Valencia, pendiente de publicación.
- y Martínez, F. (2008): *Medición y evaluación educativa*. Madrid, La Muralla.
- y Rodríguez, S. (1984): «Precisiones y limitaciones explicativas en los métodos correlacionales: Alternativas metodológicas», *Revista de Investigación Educativa*, v. 2., n.º 4, pp. 103-132.
- Matud, P. (1993): «Cuestionarios e inventarios de autoinforme», en W. Peñate, P. Matud e I. Ibáñez, *Evaluación psicológica: Concepto y técnicas de análisis*, Valencia, Promolibro, pp. 187-210.
- Mayor Ruiz, C. (1998): *La evaluación como estrategia de mejora. Evaluación de programas, centros y profesores*, Sevilla, Kronos.
- McDonald, B. (1975): «Evaluation and the control of education», en D. Tawney (ed.), *Evaluation: The state of the art*, Londres, Schools Council.
- (1977): «A political classification of evaluation studies», en D. Hamilton, D. Jenkins y C. King, *Beyond the numbers game*, Londres, McMillan.
- Medina Rivilla, A. (1994): «Evaluación de proyectos curriculares de centros», en M. L. Villar (coord.), *Manual de entrenamiento: evaluación de procesos y actividades educativas*. Título propio de expertos en evaluación educativa, Universidad de Sevilla, GID.
- Mehrens, W. A. y Lehmann, I. J. (1982): *Medición y evaluación en la educación y en la psicología*, México, Compañía Editorial Continental.
- Merriam, S. B. (1990): *Case study research in education*, Oxford, University Press.
- Metfessel, N. S. y Michael, W. B. (1967): «A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs», *Educational and Psychological Measurement*, n.º 27, pp. 931-943.
- Monedero Moya, J. J. (1998): *Bases teóricas de la evaluación educativa*, Málaga, Aljibe.
- Morgan, D. (1997): *Focus groups as qualitative research*, Thousand Oaks, Sage.
- Muñiz, J. (1998): *Teoría clásica de los tests*, Madrid, Pirámide.
- y Hambleton, R. K. (1999): «Evaluación psicométrica de los tests informatizados», en J. Olea, V. Ponsoda y G. Prieto (ed.), *Tests informatizados. Fundamentos y aplicaciones*, Madrid, Pirámide.
- Muñoz-Repiso, M. y Murillo, F. J. (2000): «Mejora de la eficacia escolar. Investigación internacional para analizar la capacidad de las escuelas para el cambio», *XII Congreso Nacional y I Iberoamericano de Pedagogía*, Madrid.
- Nevo, D. (1986): «The conceptualization of educational evaluation: An analytical review of the literature», en E. R. House (ed.), *New directions in educational evaluation*, Londres, The Falmer Press, pp. 15-29.

- (1997): *Evaluación basada en el centro*, Bilbao, Mensajero.
- Nieto Gil, J. M. (1994): *La autoevaluación del profesor: Cómo puede el profesor evaluar su propia práctica docente*, Madrid, Escuela Española.
- Nieto Gil, J. M. (1996). *La autoevaluación del profesor: Cómo puede el profesor evaluar su propia práctica docente*, Madrid, Escuela Española, segunda edición.
- Nirenberg, O., Brawerman, J. y Ruiz, V. (2000): *Evaluar para la transformación. Innovaciones en la evaluación de programas y proyectos sociales*, Buenos Aires, Paidós.
- Nisbet, J. (1988): «Policy-oriented research», en J. P. Keeves (ed.), *Educational research, methodology, and measurement: An international handbook*, Oxford, Pergamon Press, pp. 211-217.
- Nunnally, J. C. y Bernstein, I. J. (1995): *Teoría psicométrica*, México, Trillas.
- Oakes, J. (1986): *Education indicators. A guide for policy makers*, New Brunswick, New Jersey, Center for Policy Research in Education.
- Olea, J. y Ponsoda, P. (1996): «Tests adaptativos informatizados», en J. Muñoz (coor.), *Psicometría*, Madrid, Editorial Universitas, pp. 729-783.
- OMS (1981): *Evaluación de programas de salud*, Ginebra.
- Owens, T. R. (1973): «Education evaluation by adversary proceeding», en E. R. House (ed.), *School evaluation: The politics and process*, Berkeley, California, McCutchan.
- Padilla Carmona, M. T. (2002): *Técnicas e instrumentos para el diagnóstico y la evaluación educativa*, Madrid, CCS.
- Pain, A. (1992): *Evaluar las acciones de formación*, París, Editions d'Organisation.
- (1993): *Cómo evaluar las acciones de capacitación*, Argentina, Granica.
- Parlett, M. y Hamilton, D. (1972): «Evaluation as illumination: A new approach to the study of innovatory programs», *Occasional Paper*, n.º 9, Edimburgo, Centro de Investigación sobre Ciencias Educativas.
- y — (1983): «La evaluación como iluminación», en J. Gimeno Sacristán y A. Pérez Gómez (ed.), *La enseñanza: Su teoría y su práctica*, pp. 450-466.
- Parra, J. (1996): *Diagnóstico en educación*, Barcelona, PPU.
- Patton, M. Q. (1990): *Qualitative evaluation research methods*, Londres, Sage.
- Pawson, R. y Tilley, N. (1997): *Realistic evaluation*, Londres, Sage.
- Perales Montolío, M. J. (2002a): «Estudio de validación de un modelo de evaluación de la formación ocupacional y continua», *Bordón*, v. 54, n.º 1, pp. 95-113.
- (2002b): *Proyecto Docente presentado en el concurso ordinario para la plaza de Profesor Titular del área de Métodos de Investigación y Diagnóstico en Educación (Medición y evaluación en educación)*. Universidad de Valencia.
- Pérez Carbonell, A. (2000): *Proyecto Docente presentado en el concurso ordinario para la plaza de Profesor Titular del área de Métodos de Investigación y Diagnóstico en Educación (Metodología de la investigación evaluativa)*. Universidad de Valencia.
- (2006): *Evaluación de programas educativos*, Madrid, La Muralla.
- Pérez Gómez, A. (1983): «Modelos contemporáneos de evaluación», en J. Gimeno Sacristán y A. Pérez Gómez (ed.), *La enseñanza: Su teoría y su práctica*, pp. 426-449.
- Pérez Juste, R. (1985): «Estudios correlacionales», en A. De la Orden, *Diccionario de Investigación Educativa*, Madrid, Anaya.
- (1986): *Evaluación de los logros escolares*, Madrid, UNED.
- (1989): *Pedagogía experimental. La medida en educación*, Madrid, UNED.
- (1992): *El modelo evaluativo de la reforma educativa*, Madrid, UNED.

- (1994a): «Investigación evaluativa», en V. García Hoz (dir.), *Problemas y métodos de investigación en educación personalizada*, Madrid, Rialp, pp. 405-418.
- (1994b): «¿Por qué diferenciar los términos evaluación y diagnóstico?», *Revista de Investigación Educativa*, n.º 23, pp. 588-591.
- (1995a): «Evaluación de programas educativos», en A. Medina Rivilla y L. M. Villar Angulo (coord.), *Evaluación de programas educativos, centros y profesores*, Madrid, Universitas, pp. 73-106.
- (1995b): «La evaluación en el proceso educativo. Etapas y elementos», en R. Pérez Juste y J. M. García Ramos, *Diagnóstico, evaluación y toma de decisiones*, Madrid, Rialp, pp. 17-34.
- (1999): «Nuevas perspectivas de gestión, control y evaluación», *Bordón*, v. 51, n.º 4, pp. 449-477.
- (2000): «La evaluación de programas educativos: conceptos básicos, planteamientos generales y problemática», *Revista de Investigación Educativa*, v. 18, n.º 2, pp. 261-287.
- (2002): «Hezkuntzaprogramen ebaluaketa: Oinarrizko kontzeptuak, ikuspuntu nagusiak eta arazoak», *Tantak*, n.º 27, pp. 53-81.
- Pérez Serrano, G. (1994): *Investigación cualitativa. Retos e interrogantes. I. Métodos*, Madrid, La Muralla.
- Perrenoud, P. (1981): «De las diferencias culturales a las desigualdades culturales: La evaluación y la norma en una enseñanza indiferenciada», *Infancia y Aprendizaje*, n.º 14, pp. 19-50.
- Pineault, R. y Daveluy, C. (1987): *La planificación sanitaria. Conceptos, métodos, estrategias*, Barcelona, Masson.
- Popham, W. J. (1969): «Objective and instruction», en R. Stake (ed.), *Instructional objectives*, Chicago, Rand McNally.
- (1980): *Problemas y técnicas de la evaluación educativa*, Madrid, Anaya.
- (1983): *Evaluación basada en criterios*, Madrid, Magisterio Español.
- (1988): *Educational evaluation*, Englewood Cliffs, New Jersey, Prentice Hall, segunda edición.
- Popkewitz, T. S. (1988): *Paradigma e ideología en investigación educativa*, Madrid, Mondadori.
- Provus, M. (1971): *Discrepancy evaluation. For educational program improvement and assessment*, Berkeley, California, McCutchan.
- Pujadas, J. J. (1992): *El método biográfico: el uso de las historias de vida en ciencias sociales*, Madrid, CIS (Cuadernos metodológicos, 5).
- Real Academia Española (1992): *Diccionario de la lengua española*, Madrid, Espasa Calpe.
- Rebollo Catalán, M. Á. (1993): «El informe de evaluación», en M. P. Colás Bravo y M. Á. Rebollo Catalán, *Evaluación de programas. Una guía práctica*, Sevilla, Kronos, pp. 171-180.
- Rippey, R. M. (comp.) (1973): *Studies in transactional evaluation*, Berkeley, McCutchan.
- Rodríguez Conde, M. J. (1999): *Proyecto Docente presentado en el concurso ordinario para la plaza de Profesor Titular del área de Métodos de Investigación y Diagnóstico en Educación (Métodos de Investigación Educativa)*. Universidad de Salamanca.
- Rodríguez Espinar, S. (2000): «La evaluación del aprendizaje de los estudiantes», *Ponencia presentada al I Congreso Internacional: Docencia Universitaria e Innovación*, Barcelona, Institut de Ciències de l'Educació (UAB, UB, UPC).

- Rodríguez Gómez, G.; Gil Flores, J. y García Jiménez, E. (1996): *Metodología de la investigación cualitativa*, Málaga, Aljibe.
- Rodríguez Lajo, M. (1987): *Medición y evaluación del rendimiento académico asistido por ordenador*, Barcelona, Ferré Moret.
- Rosales, C. (1990). *Evaluar es reflexionar sobre la enseñanza*, Madrid, Narcea.
- Rossi, P. H. (1981): «The professionalization of evaluation research in the United States», en Levine et al. (ed.), *Evaluationn. Research and practice. Comparative en a international perspectives*, Beverly Hills, California, Sage, pp. 220-236.
- y Freeman, H. E. (1993): *Evaluation. A systematic approach*, Newbury Park, California, Sage.
- ; — y Lipsey, M. (1999): *Evaluation. A systematic approach*, Thousand Oaks, Sage.
- Rotger Amengual, B. (1983): «Evaluación», en S. Sánchez Cerezo (dir.), *Diccionario de las Ciencias de la Educación*, Madrid, Diagonal/Santillana, pp. 603-604.
- Rul Gargallo, J. (1995): «La evaluación comunicativa, factor de desarrollo humano, organizativo y curricular. Valorar, medir y evaluar», *Aula de Innovación Educativa*, n.º 39, pp. 70-77.
- Ruttman, L. (1984): *Evaluation research methods: A basic guide*, Beverly Hills, California, Sage.
- Sáez Brezmes, M. J. y Carretero, A. J. (1995): «Un modelo posible: La evaluación en la España actual», en M. J. Sáez Brezmes (ed.), *Conceptualizando la evaluación en España*, Madrid, Servicio de Publicaciones de la Universidad de Alcalá, pp. 131-141.
- Salinas, D. (2002): *¡Mañana examen! La evaluación: entre la teoría y la realidad*, Barcelona, Graó.
- Sancho Gil, J. M. (1990): «De la evaluación a las evaluaciones», *Cuadernos de Pedagogía*, n.º 185, pp. 8-11.
- Santiago, C.; Lukas, J. F.; Pikabea, I. y Oihartzabal, L. (2001): *Hezkuntza ikerkuntzarako Lexikoa. Léxico de investigación educativa. Educational research lexicon. Lexique de recherche éducative. Lexic d'investigació educativa. Léxico de investigación educativa*, Donostia, Erein.
- Santisteban, C. (1990): *Psicometría. Teoría y práctica en la construcción de tests*, Madrid, Norma.
- Santos Guerra, M. Á. (1990): *Hacer visible lo cotidiano: Teoría práctica de la evaluación cualitativa de centros escolares*, Madrid, Akal Universitaria.
- (1993): *La evaluación: Un proceso de diálogo, comprensión y mejora*, Málaga, Aljibe.
- Scriven, M. S. (1967): «The methodology of evaluation», en R. W. Tyler, R. M. Gagné y M. S. Scriven (eds.), *Perspectives in curriculum evaluation*, Chicago, Rand McNally.
- (1973): «Goalfree evaluation», en E. R. House (ed.), *School evaluation: The politics and process*, Berkeley, McCutchan.
- (1991): *Evaluation thesaurus*, Newbury Park, California, Sage.
- (1994): «Duties of the teacher», *Journal of Personnel Evaluation in Education*, n.º 4, pp. 19-32.
- Secretaría de Educación Pública (1991): *Evaluación de la educación superior*, México, Secretaría de Educación Pública-Comisión Nacional de Evaluación.
- Serrano Angulo, J.; Tójar Hurtado, J. C.; Galante, R. y Elósegui, E. (1995): «Aportaciones recientes en medición y evaluación educativas», *Revista de Investigación Educativa*, n.º 26, pp. 129-157.

- Shadish, W. R.; Cook, T. D. y Leviton, L. C. (1991): *Foundations of program evaluation: Theories of practice*, Newbury Park, California, Sage.
- Shavelson, R. J.; McDonnell, L. M. y Oakes, J. (1991): «What are educational indicators and indicator systems?», *American Institutes for Research*. ERIC-EDO-TM-91-5.
- Shulman, L. S. (1989): «Paradigmas y programas de investigación en el estudio de la enseñanza: Una perspectiva contemporánea», en M. C. Wittrock (dir.), *La investigación de la enseñanza I*, Barcelona, Paidós, pp. 9-91.
- Sierra Bravo, R. (1983): *Técnicas de investigación social*, Madrid, Paraninfo.
- Silva, F. (1989): *Evaluación conductual y criterios psicométricos*, Madrid, Pirámide.
- Smith, J. K. y Heshusius, L. (1986): «Closing down the conversation: the end of the Quantitative-Qualitative debate among Educational Inquirers», *Educational Researcher*, V. 1, n.º 15, pp. 4-12.
- Spindler, G. y Spindler, L. (1992): «Cultural process and ethnography: An anthropological perspective», en M. LeCompte, W. L. Millroy y J. Preissle, *The handbook of qualitative research in Education*, Nueva York, Academic Press, pp. 53-92.
- Stake, R. E. (1967): «The countenance of educational evaluation», *Teachers College Record*, n.º 68, pp. 523-540.
- (1975): «Program evaluation: Particular responsive evaluation», *Occasional Paper*, n.º 5, Kalamazoo, Centro de Evaluación, University of Western Michigan.
- (1976): *Evaluating educational programs. The need and the response*, Nueva York, McGraw-Hill.
- (1981): «Setting standards for educational evaluators», *Evaluation News*, v. 2, n.º 2, pp. 148-152.
- Stanley, A. J.; Glock, M. D. y Wardeberg, H. L. (1972): *Evaluación de los alumnos en la Escuela Primaria*, Madrid, Aguilar.
- Stecher, B. y Davis, W. A. (1988): *How to focus an evaluation*, Londres, Sage.
- Stenhouse, L. (1987): *La investigación como base de la enseñanza*, Madrid, Morata.
- Stevens, S. S. (1951): «Mathematics, measurement and psychophysics», en S. S. Stevens (ed.), *Handbook of experimental psychology*, Nueva York, John Wiley.
- Stoll, L. y Wikeley, F. (1998): «Issues on linking school effectiveness and school improvement», en W. Hoeben (ed.), *Effective school improvement: State of the art. Contribution to a discussion*, Groningen, Institute for Educational Research. University of Groningen, pp. 29-58.
- Stronge, J. H. (1997): *Evaluating teaching. A guide to current thinking and best practice*, Thousand Oaks, California, Corwin Press.
- Stufflebeam, D. L. (1973): *A new conceptualization of evaluation competence*, Nueva Orleans, AERA paper.
- et al. (1971): *Educational evaluation and decision-making*, Itasca, Illinois, Peacock.
- y Shinkfield, A. J. (1987): *Evaluación sistemática: guía teórica y práctica*, Barcelona, Paidós.
- Suchman, E. A. (1967): *Evaluative research*, Nueva York, Russell Sage Foundation.
- Suen, H. K. y Ary, D. (1989): *Analyzing quantitative behavioral observation data*, Hillsdale, New Jersey, Erlbaum.
- Taba, H. (1983): «La evaluación de los resultados del currículo», en H. Taba, *Elaboración del currículo*, Buenos Aires, Troquel.
- Talmage, H. (1985): «Evaluating the curriculum, what, why and how», en *National Association for Secondary Schools Principals*, pp. 1-8.

- Taylor, S. J. y Bogdan, R. (1987): *Introducción a los métodos cualitativos de investigación*, Barcelona, Paidós.
- Tejada Fernández, J. (1999): «La evaluación: Su conceptualización», en B. Jiménez Jiménez (ed.), *Evaluación de programas, centros y profesores*, Madrid, Síntesis, pp. 25-56.
- Tejedor, F. J. (1994): «La experimentación como método de investigación educativa», en V. García Hoz (dir.), *Problemas y métodos de investigación en enseñanza personalizada*, Madrid, Rialp, pp. 256-285.
- (2000): «El diseño y los diseños en la evaluación de programas», *Revista de Investigación Educativa*, v. 18, n.º 2, pp. 319-339.
- ; GarcíaValcárcel, A. y Rodríguez Conde, M. J. (1994): «Perspectivas metodológicas actuales de la evaluación de programas en el ámbito educativo», *Revista de Investigación Educativa*, n.º 23, pp. 93-127.
- Tenbrink, T. D. (1984): *Evaluación. Guía práctica para el docente*, Madrid, Narcea.
- Tiana Ferrer, A. (1995): «Perspectivas españolas actuales para la evaluación del sistema educativo: En Instituto Nacional de Calidad y Educación», en M. J. Sáez Brezmes (ed.), *Conceptualizando la evaluación en España*, Madrid, Servicio de Publicaciones de la Universidad de Alcalá, pp. 143-155.
- (1997a): *Avaluació de programes, centres i professorat*, Barcelona, UOC.
- (1999): «Ikastetxeen kanpo-ebaluazioa Aukerak eta mugak», *Tantak*, n.º 21, pp. 7-32.
- Tójar Hurtado, J. C. (2001): *Planificar la investigación educativa: Una propuesta integrada*, Buenos Aires, Fundec.
- Tuckman, B. W. (1975): *Measuring educational outcomes. Fundamentals of testing*, Nueva York, Harcourt Brace Jovanovich.
- Tyler, R. W. (1950): *Basic principles of curriculum and instruction*, Chicago, University of Chicago Press.
- UNESCO (1979): *Evaluation terms: Unesco Glossary*, París, Unesco.
- Verdugo, M. Á. (dir.) (1994): *Evaluación curricular. Una guía para la intervención psicopedagógica*, Madrid, Siglo XXI.
- Villa Sánchez, A. y Morales Vallejo, P. (1993): *La evaluación del profesor: Una visión de los principales problemas y enfoques en diversos contextos*, Vitoria Gasteiz, Gobierno Vasco.
- Walberg, H. J. y Haertel, G. D. (ed.) (1990): *The International Encyclopedia of Educational Evaluation*, Oxford, Pergamon Press.
- Walker, J. C. y Evers, C. W. (1997): «Research in education: Epistemological issues», en J. P. Keeves (ed.), *Educational research, methodology, and measurement: An international handbook*, Oxford, Pergamon Press, pp. 22-31.
- Weiss, C. H. (1970): *Investigación evaluativa*, México, Trillas.
- Weiss, J. (dir.) (1991): *L'évaluation: problème de communication*, Cousset, DelVal.
- Wolf, D. P. y Reardon, S. F. (1996): «Acces to excellence through new forms of student assessment», en J. B. Baron y D. P. Wolf (eds.), *Performance based student assessment: Challenges and possibilities*, Ninety-fifth Yearbook of the National Society of Education, Part 1, Chicago, University of Chicago Press.
- Wolf, R. M. (1974): «The citizen as jurist: A new mode in educational evaluation», *Citizen Action in Education*, n.º 4.
- (1990a): *Evaluation in education*, Nueva York, Praeger.

- (1990b): «The nature of educational evaluation», en H. J. Walberg y G. D. Haertel (eds.), *The international encyclopedia of educational evaluation*, Oxford, Pergamon Press.
- (1990c): «Judicial evaluation. Foundation of competency assessment and program review», en H. J. Walberg y G. D. Haertel (eds.), *The international encyclopedia of educational evaluation*, Oxford, Pergamon Press.
- Woods, P. (1987): *La escuela por dentro. La etnografía en la investigación educativa*, Barcelona, Paidós.
- Worthen, B.R. y Sanders, J. R. (1973): *Educational Evaluation: Theory and practice*, Woethington (Ohio), C.A. Jones.
- y Sanders, J. R. (1987): *Educational evaluation*, Nueva York, Longman.
- ; — y Fitzpatrick. (1997): *Program evaluation*, Nueva York, Longman.
- Yin, R. K. (1994): *Case Study Research. Design and Methods. (Applied Social Research Methods Series. Vol. 5)*, Thousand Oaks, California, SAGE.
- Zabalza, M. Á. (1991): *Los diarios de clase. Documento para estudiar cualitativamente los dilemas prácticos de los profesores*, Barcelona, PPU.

