

EL DISEÑO Y LOS DISEÑOS EN LA EVALUACIÓN DE PROGRAMAS

Fco. Javier Tejedor

Universidad de Salamanca

I. LA EVALUACIÓN DE PROGRAMAS COMO INVESTIGACIÓN EVALUATIVA

La Evaluación de Programas es un proceso para generar formas útiles de comprensión sobre una innovación (o una intervención). Este sentido de utilidad hay que entenderlo como la posibilidad de utilizar la intervención como un recurso crítico para mejorar los procesos de acción educativa, razón por la que, a nivel metodológico, puede entenderse como un proceso de investigación evaluativa (Tejedor y otros, 1994; Pérez Juste, 1995; Amezcua y Jiménez, 1996; Martínez Mediano, 1997).

La investigación evaluativa ha ido perfilándose con una metodología concreta en los últimos quince años. Esta metodología queda delimitada por: una terminología propia, un conjunto de herramientas conceptuales y analíticas específicas, y unos procesos y fases también específicos. El conjunto define un campo floreciente en el que realmente lo específico es la mezcla y no las partes constitutivas. Para Alvira (1991), algunas circunstancias que han contribuido a conformar este campo de actuación serían:

- a) Mayor interrelación entre las actividades de evaluación y programación. Si en una perspectiva clásica la evaluación se realiza después de las actividades de planificación, programación y desarrollo de la intervención, en la actualidad ambas actividades caminan de forma paralela.
- b) La investigación evaluativa es investigación (con todos los considerandos epistemológicos, metodológicos y técnicos que ello entraña), pero es más que investigación, ya que se entiende como un proceso cuya finalidad es aplicar procedimientos científicos para acumular evidencia válida y fiable sobre la manera y grado en que un conjunto de actividades producen resultados o efec-

tos concretos. Intenta determinar el valor de la acción educativa, buscar soluciones a problemas concretos o mejorar las estrategias de actuación.

- c) Se produce una mayor implicación de las partes interesadas (observación participante, investigación acción,...).
- d) Pueden integrarse estrategias de investigación muy diversas, tanto cuantitativas como cualitativas. Es un campo de trabajo que posibilita la utilización simultánea y sucesiva de diferentes metodologías para el tratamiento de los temas objeto de estudio e intervención. Será la naturaleza específica del programa y la pericia del evaluador quienes a la postre determinarán la utilización de una o varias opciones.

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. Vemos que aparecen importantes puntos de contacto respecto a lo que entendemos por «investigación»: proceso sistemático, riguroso, controlado; utilización de técnicas de medición y análisis de datos. La evaluación se distingue de la investigación por los objetivos y no por los métodos...

Algunos autores, sin embargo, creen que conviene diferenciar ambas tareas, señalándose como características peculiares de la investigación evaluativa las siguientes (De la Orden, 1985):

- a) Los juicios de valor acompañan al desarrollo de todo el programa: desde la selección del problema a estudio hasta en la aplicación de metodologías concretas.
- b) Resulta difícil (a veces inadecuado) formular hipótesis precisas.
- c) La replicación es prácticamente imposible dadas las características casi irrepetibles de la ejecución de un programa de intervención.
- d) La recogida de datos está condicionada por la viabilidad del proceso.
- e) Multitud de variables relevantes sólo son susceptibles de un control superficial limitándose por tanto la utilización de diseños de carácter mas experimental.
- f) Las decisiones importantes en relación con la paralización, sustitución o repetición del programa no son competencia del evaluador.
- g) El informe debe adaptarse a las exigencias de quien lo financia.

Una característica básica de la investigación evaluativa es que se lleva a cabo en un proceso de intervención, en la misma acción, que es lo realmente importante. La investigación, señalaba Weis (1982), es una prioridad secundaria, lo que debe llevarnos a pensar que la evaluación/investigación tiene que adaptarse al contexto del programa y perturbar lo menos posible su desarrollo.

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 apoyan mutuamente. La intervención sacará partido de la investigación en la medida en que ésta se extienda a la solución de problemas prácti-

cos. La investigación activa se asimila, enriqueciéndola, a la noción de investigación aplicada.

Toda situación educativa es susceptible de este doble tratamiento. Por una parte, existe la necesidad de «conocerla» y de «explicarla». Por otra, es preciso «comprenderla» y «mejorarla». Se requiere que la investigación alcance un nivel de integración suficiente, en este caso, entre el pensamiento y la decisión: la preocupación por el saber se une de alguna manera a la preocupación por el actuar. Este campo de estudio, que algunos denominan «investigación orientada», está a la base de toda investigación evaluativa.

La investigación orientada/evaluativa nace como respuesta a necesidades sociales y se elabora en función de tales necesidades; se realiza con objeto de llegar a la solución de los problemas que exige una acción informada científicamente. Habrá de tener necesariamente un carácter interdisciplinar y viene determinada por los objetivos, los resultados y las motivaciones.

Al referirnos al tema del diseño en la Evaluación de Programas nos vamos a encontrar, como ocurre en el más amplio campo de la investigación, con una doble acepción de dicho término:

- a) Diseño como planificación general del proceso de intervención (el diseño o diseño de evaluación)
- b) Diseño como planificación puntual del análisis de los efectos de la intervención realizada (los diseños o diseños de investigación).

Esta diferenciación es la que se ha tenido en cuenta a la hora de titular esta aportación y que pasamos a comentar con detenimiento.

2. EL DISEÑO EN LA EVALUACIÓN DE PROGRAMAS

Un diseño es un plan que fija qué medidas y cuándo serán tomadas en el curso de un proceso de evaluación. Plantearse la especificación de un diseño supone haber realizado reflexiones previas en torno a cuestiones relacionadas con el concepto general de evaluación y con las finalidades de las diferentes opciones metodológicas; posteriormente delimitaremos, como parte de la planificación prevista, los fundamentos de la misma, los objetivos, los elementos personales implicados en el programa, los ámbitos de intervención, los instrumentos de recogida de datos, las técnicas de análisis y los momentos de la evaluación; todo ello, de acuerdo al modelo que se haya elegido para realizar la actividad evaluativa y teniendo en cuenta que en todos estos elementos existen diferencias según la perspectiva o enfoque (paradigma o modelo) por se siga.

El diseño de Evaluación de Programas es tanto un proceso como un conjunto de escritos o planes que incluyen unos elementos que presentan los propios contenidos y que generan numerosas preguntas. No hay un único camino para realizar la evaluación de un programa por lo que habrá necesariamente que prestar atención a numerosos elementos a la hora de concretar un diseño o plan de análisis: condiciones

organizacionales, necesidades, intereses y valores de los implicados en el proceso, objetivos del programa, recursos necesarios y disponibles,...

En cualquier caso, independientemente del modelo que se haya elegido, cualquier mínima planificación tiene que atender las siguientes tareas:

- establecer las actividades que se van a evaluar
- fijar los criterios de evaluación
- elegir las estrategias para la obtención de información
- analizar la información
- tomar decisiones a partir de la información analizada

Donde podemos identificar los dos elementos básicos del diseño de una evaluación, que expresados en términos sintéticos serían:

- La determinación de los datos que se requieren para valorar la eficacia del programa
- La determinación del plan para obtener la información necesaria que nos permita establecer relaciones entre el programa y los resultados obtenidos

En el cuadro 1 presentamos una reformulación simplificada de la propuesta de Pérez Carbonell (1998, pp. 188-192) respecto a los aspectos a considerar en la delimitación del diseño en el marco de la Evaluación de Programas.

Algunas cuestiones importantes a tener en cuenta serían (Carballo, 1990, págs. 166):

- ¿Se van a establecer las pautas que se seguirán en el proceso evaluativo? (diseños prefijados vs emergentes)
- Los resultados de la evaluación, ¿se van a utilizar para mejorar el programa, para emitir un informe sobre su funcionamiento o para ambas cosas a la vez? (diseños para una evaluación formativa vs normativa?)
- ¿Se van a incluir en el estudio procesos de intervención o sólo se va a describir el funcionamiento del programa? Si se incluye un proceso específico de intervención será necesario analizar sus efectos, lo que requerirá la explicitación de un «diseño de investigación»

a) Diseños prefijados vs emergentes

Los diseños prefijados siguen la pauta del modelo positivista, requiriéndose que los objetivos de la evaluación estén claramente definidos y fijados con anterioridad a la ejecución del programa, lo que supone que todos los componentes del modelo (objetivos variables, tratamientos, fuentes de información, datos a recoger, análisis de datos,...) estén determinados a partir de presupuestos teóricos o de experiencias en evaluación de otros programas similares.

Estos diseños tratan de informarnos sobre la consecución de objetivos en la ejecución del programa, intentando el evaluador esclarecer los factores (variables y relacio-

CUADRO 1
ASPECTOS A CONSIDERAR EN LA DELIMITACIÓN DEL DISEÑO

Elementos	Contenidos	Dificultades	Consecuencias
Desarrollo del programa	Entorno Objeto de la evaluación Destinatarios Criterios de evaluación Actividades	No se considera a los implicados	No se implican los interesados y se tiende a descalificar la evaluación
Recogida de datos	Variables de interés Fuentes de información Instrumentos Técnicas de recogida	Procedimientos que violan derechos	Pleitos
		Procedimientos que utilizan tiempo valioso de usuarios	Falta de cooperación Datos no reales
		Procedimientos que implican cambios de hábitos	
Análisis de datos	Tipo de datos recogidos Criterios estadísticos Análisis adecuados Interpretación correcta	Datos con incidencia individual	Información criticada y no valorada
		Análisis demasiado técnicos	
Realización de informes	Destinatarios Número de informes Difusión	Los implicados no participan en el informe	La información no se utiliza por los grupos afectados
		La audiencia no tiene oportunidad de reaccionar	
		El informe afecta a puestos de trabajo	
		Se pone en aprieto a personas o instituciones	
Manejo de la información	Decisiones a tomar Personas afectadas Presupuesto de las decisiones a adoptar	No se especifican bien responsabilidades	Agravio colectivo
		Presupuestos deficientes Información desaprovechada	

nes entre ellas) que han mediado en dicha consecución. En síntesis, estos diseños «fijos» se caracterizan porque:

- son planeados con anterioridad
- están estructurados, aunque admite reajustes
- se especifican problemas y objetivos, fuentes de información, instrumentos, temporalización,...
- se fijan los destinatarios de los informes
- los datos suelen ser de carácter cuantitativo y se valora todo aquello que se aproxima al paradigma racionalista

La práctica evaluativa ha ido poniendo de manifiesto que la evaluación de la consecución de los objetivos, planteada en estos términos, resultaba muy difícil de realizar

ya que las cosas no suelen desarrollarse en los términos inicialmente previstos, lo que llevó a los evaluadores a flexibilizar sus planteamientos evaluativos iniciales, abriéndose a nuevas perspectivas valorativas, no sólo de los resultados sino también del propio proceso seguido (evaluación del proceso) intentando relacionar la valoración de los resultados con las condiciones y circunstancias en las que se desarrolló el programa, lo que nos conduce a criterios de eficiencia y no sólo de eficacia.

En seguida, este tipo de alternativa tuvo seguidores fervorosos (Stake, Owens, Parlett y Hamilton, Guba...) y, especialmente, Scriven que incorpora como novedad la idea de valorar los efectos no previstos, dejándose llevar por el desarrollo del programa y cuestionándose incluso la utilidad de la formulación inicial de objetivos (Scriven, 1973).

Estos diseños «emergentes» pueden, globalmente, caracterizarse en los siguientes términos:

- se adaptan mejor a las necesidades e intereses de los participantes
- están abiertos a posibles modificaciones que durante el desarrollo puedan incorporarse
- los problemas y objetivos no tienen por qué ser especificados inicialmente; se produce una reformulación constante de los objetivos del programa
- no se especifican elementos tales como instrumentos de recogida de datos, momentos de obtención de la misma,..
- se exige que el evaluador siga el desarrollo del programa estando en contacto permanente con los participantes; este distinto grado de implicación del evaluador va a condicionar las técnicas e instrumentos de obtención de datos
- los datos suelen ser de carácter cualitativo y se valora todo aquello que se aproxima a los paradigmas interpretativo y crítico
- debe cuidarse la adecuada selección de información (demasiado abundante e irrelevante en muchos casos) y una implicación excesiva del evaluador en la valoración de logros

Quizás puede pensarse que esta diferenciación dicotómica establecida en los diseños («fijo/emergente») es más teórica que práctica. En realidad parece posible pensar que no exista tal dicotomía; pensamos que el diseño, en el marco de la Evaluación de Programas, debe ser establecido a priori, respondiendo a la concepción teórica que sobre la evaluación tienen los responsables del programa y el propio evaluador, debiendo incluir referencias a los objetivos del programa, sus componentes, los recursos necesarios, las estrategias de recogida y análisis de datos, las posibles decisiones que de su desarrollo puedan derivarse y la presentación y difusión de los resultados.

Pero abogar por la previsión no es estar a favor de la inmutabilidad. Nuestra postura en modo alguno excluye la posibilidad de ir incorporando durante el desarrollo de la ejecución del programa todos aquellos elementos que se consideren más favorables a la optimización del proceso planificado: desde la reformulación de objetivos a la difusión de resultados pasando por la incorporación de nuevos recursos o la modificación de las estrategias de recogida y análisis de datos.

b) Diseños de evaluaciones formativas vs sumativas

Puede afirmarse que fue Cronbach quién, a partir de 1980, extendió la aplicación de los términos «formativa» y «sumativa», inicialmente propuestos por Scriven para su aplicación a contextos instructivos, al ámbito de la evaluación de programas. En este contexto, la evaluación formativa es una parte importante del proceso que sigue el desarrollo del programa, proporcionándole retroalimentación de su funcionamiento para favorecer el mismo e incorporar los cambios que puedan contribuir a mejorar su realización, lo que implica la vinculación del evaluador con el discurrir del programa: evaluación de carácter interno con efectos potenciales, sobre todo en la mejora del desarrollo del programa. Sin duda, responde al papel más noble que puede desempeñar la evaluación: su incardinación en el desarrollo del propio programa, de la propia intervención, del propio proceso educativo.

La evaluación formativa permite establecer una interacción continua entre la evaluación y la acción educativa, destacando su repercusión en la mejora del proceso y en la orientación del programa y, por consecuencia, en la optimización de los productos (logros) que se obtengan tras la aplicación de dicho programa, sobre todo si la evaluación formativa se plantea la búsqueda de las causas que están explicando los datos que nos vamos encontrando.

En síntesis, el diseño formativo puede caracterizarse por los siguientes considerandos:

- la evaluación se plantea desde una perspectiva interna
- su objetivo es intentar la mejora del programa
- los datos son utilizados por el propio equipo técnico del programa o por los implicados (el evaluador debe ser parte integrante del equipo)
- se persigue sobre todo la utilidad para los usuarios
- el diseño puede ser tanto fijo como emergente (si es fijo deberán aceptarse cambios)

A la hora de valorar un diseño de corte formativo deberá considerarse (Bartolomé, 1990):

- la coherencia lógica del diseño
- su capacidad para responder a las necesidades del grupo y/o a su proyecto social y educativo
- su flexibilidad de adaptación y mejora
- su posibilidad de realización, a través de un acercamiento progresivo y de las percibidas por el grupo y de un acercamiento de las actividades previstas y realizadas
- su eficacia progresiva, valorando su incidencia real en la mejora educativa

La evaluación sumativa, por su parte, se relaciona con la terminación del proceso de intervención, intentando globalizar la evaluación formativa que se haya podido ir

haciendo a lo largo del proceso. La evaluación sumativa nos informa sobre los resultados obtenidos y, por tanto, sobre las necesidades que han podido satisfacerse con la aplicación del programa. Parece importante destacar el hecho de relacionar la consecución de logros con la satisfacción de necesidades en mayor medida que con la consecución de objetivos, lo que puede ayudarnos a reconocer la especificidad de los procesos de intervención en contextos de carácter no estrictamente instructivo.

El diseño sumativo respondería a las pautas siguientes:

- la evaluación puede plantearse desde una perspectiva externa
- su objetivo es intentar ofrecer un juicio sobre la satisfacción de necesidades y/o sobre la consecución de los objetivos
- interesa en gran medida la credibilidad externa
- el evaluador debe ser ajeno al equipo técnico del programa para favorecer la objetividad en la valoración de los resultados
- el diseño suele ser mas bien de tipo fijo que emergente

c) *Distintos tipos de evaluaciones*

Desde una perspectiva metodológica, la puesta en marcha de un programa supone una actuación que debe atender ordenadamente a las distintas fases del proceso. La supervisión del desarrollo de cada una de esas fases en términos evaluativos nos lleva al establecimiento de diferentes tipos de evaluación que pueden llegar a tener interés en si mismo, incluso con su metodología propia (necesidades, objetivos, proceso, resultados,...); pero no podemos olvidar que todos ellos forman parte de un esquema único de investigación que nos informará sobre la puesta en marcha del programa y la calidad de sus logros. Este esquema global del proceso evaluativo de un programa de intervención sería similar al que presentamos a continuación.

1) **Evaluación de necesidades**

Debe ser la primera actividad a llevar a cabo en la puesta en marcha de un programa de intervención (Tejedor, 1990). Aunque algunos autores plantean la evaluación de necesidades con el mismo énfasis que la evaluación de programas, considero que es mas pertinente concebirla como una fase de dicho proceso, limitando entonces sus objetivos, lo que no implica desvirtuar su esquema operativo que entiendo está suficientemente conformado. Podemos pensar que la evaluación de necesidades, siendo una fase del proceso de investigación para la puesta en marcha de un programa de intervención, tiene entidad propia y es susceptible de ser analizada con un esquema completo de investigación evaluativa (Witkin, 1984).

2) **Evaluación del diseño (evaluación de entrada)**

Supone la preocupación por la secuencia de los pasos a seguir, por la claridad de formulación de los objetivos planteados, por la calidad de los instrumentos o técni-

cas de obtención de información, por la suficiencia de la muestra fijada, por la adecuada localización de las unidades muestrales, por la procedencia de las actividades sugeridas, por la adecuación de las técnicas de análisis,... (Mayer, 1985; Bartolomé, 1990). Se pretende valorar el plan de acción (diseño) y los recursos del programa, a fin de seleccionar las estrategias que permitan satisfacer más adecuadamente las necesidades detectadas y que mejor se adapten al grupo de personas al que el programa va dirigido.

Incluye igualmente la evaluación de las posibilidades de evaluación del propio diseño (evaluación de la evaluabilidad). La experiencia determina que pocos de los programas llevados a la práctica son evaluables al tener problemas de especificación de objetivos, de implementación, de adecuación de las actividades a la consecución de objetivos, de carencia de medios para su desarrollo,...Por ello estamos de acuerdo con quienes sugieren la conveniencia de llevar a cabo una revisión previa de las posibilidades evaluativas del programa (Anguera, 1990).

3) Evaluación de la implementación

Este tipo de evaluación pretende analizar el funcionamiento del programa y determinar si existen diferencias entre el funcionamiento real y el previsto inicialmente. Si hubiera diferencias notorias entre ambos esquemas de funcionamiento trataríamos de hallar las causas y, si fuera posible, subsanar las deficiencias (Finnegan y otros, 1988). Puede pensarse que incluye:

- Evaluación de la cobertura: Trataremos de analizar si el programa cubre todos los objetivos previstos y si llega a todos los sujetos para los que se había pensado, determinando, si procede, los sesgos de cobertura existentes y las causas que los han motivado.
- Evaluación del proceso (seguimiento): En algunos casos, sobre todo si el programa es de larga duración, conviene realizar un análisis de su funcionamiento a fin de verificar si su desarrollo sigue las pautas previstas. Implica la revisión de todos los elementos constitutivos del programa (personales, funcionales y materiales), en su dimensión interactiva: calidad de las actividades, distribución del tiempo y del trabajo, utilización de recursos,...

La evaluación adecuada del proceso supone la consideración de todos sus elementos por lo que se sugiere un esfuerzo previo de reconsideración respecto a las fases anteriores de ejecución del programa (necesidades, objetivos, diseño,...). Los elementos a considerar serán:

- utilización adecuada de recursos
- temporalización de actividades
- implicación de los destinatarios en las actividades
- obtención de información

4) Evaluación de resultados

Hasta hace poco tiempo era el único tipo de evaluación que se llevaba a cabo tras la finalización de las actividades que implicaba la ejecución del programa. En términos generales, se trata de analizar el grado en que el programa consigue los resultados previstos (también pueden evaluarse, siempre que se hayan previamente identificado, resultados no previstos inicialmente). Se intenta determinar hasta que punto pueden vincularse los logros obtenidos con los efectos del programa, lo que tiene especial interés en los diseños de intervención dada la dificultad para establecer nexos de causalidad y covariación entre las variables consideradas en el programa.

La evaluación de resultados admite la modalidad de «evaluación de verificación», que pretende efectuar comprobaciones a base de patrones claramente identificados, tanto en base a datos cuantitativos como cualitativos, y la modalidad de «evaluación comprensiva», que pretende ampliar o profundizar la comprensión de situaciones y que utiliza básicamente datos cualitativos.

La temática de la evaluación de resultados es amplia y compleja, apareciendo tratada en la bibliografía con diferentes terminologías: evaluación de productos, de la efectividad, de la eficacia, de la eficiencia,... (De la Orden, 1985, 1990). Los elementos a considerar en la evaluación de productos serán:

- calidad de la información obtenida
- posibilidad de valorar la consecución de objetivos a partir de la información disponible
- calidad de los análisis y adecuación a los datos disponibles
- propuestas para la toma de decisiones
- pertinencia de las decisiones tomadas

5) Metaevaluación

Los elementos a considerar serán las pautas de cambio estratégico en la intervención y en la planificación evaluativa

3. LOS DISEÑOS DE INVESTIGACIÓN EN LA EVALUACIÓN DE PROGRAMAS

La evaluación de efectos supone la elección de un diseño adecuado de investigación que permita obtener, analizar e interpretar la información relevante para formular un juicio de valor y toma de decisiones apropiadas. El diseño determina la posibilidad de detectar y comprender los procesos y el impacto de un programa educativo, dada la multiplicidad de factores que influyen su operación y resultados (De la Orden, 1990).

Los diseños de evaluación de programas pueden variar desde el más riguroso experimento de campo a la descripción naturalista, pasando por la diversidad de diseños cuasiexperimentales o correlacionales. Existen muchas propuestas de clasificación

de diseños de investigación utilizables en el ámbito específico de la Evaluación de Programas (Fitz Gibbon y Morris, 1987; Colás y Rebollo, 1993). Nuestra propuesta es la siguiente:

A) Diseños para la investigación cuantitativa

- 1) Diseños descriptivos-comparativos (estudios tipo survey)
- 2) Diseños longitudinales
 - secuenciales (diacrónicos)
 - transversales (sincrónicos)
- 3) Diseños Correlacionales:
 - correlación y regresión
 - técnicas de clasificación
- 4) Diseños causales
- 5) Diseños experimentales:
 - preexperimentales
 - cuasiexperimentales
 - experimentales de campo (intergrupo e intragrupo)

B) Diseños para la investigación cualitativa

- 1) Observación naturalista:
 - Estudios de casos
 - Estudios del grupo
 - Estudios de comunidades (etnográficos)
 - Estudios de situaciones o contextos
- 2) Diseños participativo-críticos
 - Investigación-acción

A) El entorno cuantitativo

El diseño experimental

Presenta las características propias del método científico (repetibilidad, estabilidad y comunicabilidad), siendo sus rasgos distintivos la manipulación de variables por parte del investigador y el control de variables intervinientes en el fenómeno estudiado.

Pretende llegar a establecer relaciones causales entre una o más variables independientes y una o más variables dependientes, lo que únicamente será posible a partir de un riguroso control de las circunstancias en que podemos llegar a establecer con claridad dicha relación causal.

Se estudian las modificaciones que sufren las variables dependientes a partir de cambios intencionados (manipulación o intervención) en las variables independientes.

La exactitud de la relación establecida entre ambas dependerá de la calidad del control ejercido sobre otras variables intervinientes, bien a partir de medios físicos (similitud de situaciones), bien a partir de técnicas aleatorias.

Los diseños que se enmarcan en el método experimental se caracterizan por tener una gran validez interna, que no es más que la capacidad para atribuir ciertamente la variación observada en la variable dependiente a la modificación intencional llevada a cabo sobre la variable independiente en base al control ejercido sobre el resto de las variables (Tejedor, 1994).

El grado de seguridad de los resultados de una investigación no debe juzgarse exclusivamente en términos de validez interna. Conviene reconsiderar también sobre otros tipos de validez: externa, estadística, de constructo y ecológica.

Los presupuestos exigidos a toda investigación en términos de valor de verdad, aplicabilidad, consistencia y replicabilidad, se reinterpretan en el método experimental en términos de validez interna, validez externa, fiabilidad y objetividad (en las estrategias de investigación activa y participativa se reinterpretarán en términos de credibilidad, transferibilidad, dependencia y conformabilidad).

Un aspecto importante a tener en cuenta por el evaluador que decida seguir algunas de las variantes del método experimental es el problema que representa la «experimentación con sujetos humanos». Desde los distintos efectos perturbadores que pueden modificar las conclusiones obtenidas hasta las interacciones que pueden producirse en las características personales de los sujetos. Pese a todo, las críticas más fuertes al método experimental le llegan por sus características de artificialidad, de reduccionismo de la realidad que pretende explicar y por la operacionalidad implícita de las variables que se incluyen. Por ello no consideramos adecuada la aplicación de los diseños experimentales, en sentido estricto, en el marco de la Evaluación de Programas.

Experimentos de campo

Sin embargo, sí creemos posible, aunque con muchas limitaciones, pensar en la modalidad de «experimentos de campo», que presentando los mismos presupuestos formales, su viabilidad de aplicación reside en el hecho de llevarse a cabo en marcos naturales, lo que supone una evidente reducción de su validez interna. Kerlinger (1977) caracterizaba los experimentos de campo en los siguientes términos:

- se realizan en condiciones naturales, interviniendo todas las variables que influyen en el fenómeno
- los efectos producidos por las variables son generalmente más fuertes que en el laboratorio
- existe mayor validez externa, por su referencia a situaciones más reales
- existe la posibilidad de estudiar fenómenos complejos en situaciones similares a la vida misma; son aplicables a gran variedad de problemas
- son apropiados para la demostración de teorías, comprobación de hipótesis y la resolución de problemas prácticos.

Entre las dificultades que puede encontrarse el investigador social en la aplicación de estos diseños cabría citar los aspectos relacionados con el control de variables, con la aleatorización de las muestras y con la disponibilidad de experimentadores bien entrenados.

Aunque son muchos los diseños que pueden aplicarse en los experimentos de campo, en todos ellos deben seguirse en su planificación los siguientes pasos:

- determinación de las muestras a utilizar;
- especificación de los tratamientos de cada una de las variables independientes;
- asignación de las condiciones experimentales (tratamientos) a los diferentes grupos;
- selección de las técnicas adecuadas de control de variables;
- constitución de los «grupos de control», si procede.

Cuasi-experimentos

Cuándo no es posible establecer un control efectivo sobre las variables intervinientes, el investigador social puede recurrir a diseños de control parcial o cuasi-experimentos. Deberá prestarse en su aplicación especial interés a las variables que no se han podido controlar y considerar la posibilidad de que los resultados que obtenga (relaciones entre las variables independientes y dependientes) puedan deberse más a la influencia de las variables no controladas que a los efectos de las variables independientes.

La carencia de control se produce básicamente por la imposibilidad de manipular la variable independiente y por las dificultades de asignar aleatoriamente los sujetos a las condiciones experimentales. Las principales características de un diseño cuasi-experimental serían:

- el empleo de escenarios naturales, generalmente de tipo social;
- la carencia de un control experimental completo;
- el uso de procedimientos (producción de observaciones múltiples) como sustitutos del control experimental intentando minimizar y, en algunos casos eliminar, los efectos de tantas fuentes de invalidez interna como sea posible;
- su disponibilidad, ya que pueden realizarse cuando no es posible la aplicación de un diseño experimental (de laboratorio o de campo).

Pueden considerarse aportaciones ingeniosas de los diseños experimentales a investigaciones naturales, que utilizan para la evaluación del cambio producido por las condiciones estudiadas grupos de control no equivalentes.

Los numerosos diseños cuasi-experimentales se pueden encuadrar en una de las dos modalidades siguientes:

1. *Diseños simples de series temporales.* Realizan medidas de la variable dependiente antes y después de la aplicación de la variable independiente; el posible efecto

de ésta puede ser estimado en función de la discontinuidad que presentan las medidas tomadas después, que se puede manifestar bien por un «cambio de nivel», bien por un «cambio de tendencia o inclinación». El análisis de la serie temporal puede conducirnos a evaluar el efecto como continuo o permanente, pasajero o momentáneo, instantáneo o retardado.

2. *Diseños de grupo de control no equivalente*. Donde el investigador trabaja con grupos ya formados de una manera natural resultando muy difícil asegurar la equivalencia de los grupos de control y experimental. Se incluyen medidas pre-tratamiento y post-tratamiento, la que nos indica la influencia de la variable independiente; el hecho de que los grupos no sean equivalentes introduce una posible fuente de error en la interpretación. Aunque estos diseños pierden validez interna respecto a los experimentos de campo, su aplicabilidad con grupos naturales les convierte en los diseños más utilizados en la investigación educativa y social.

El método psicométrico

Cuando el objetivo del estudio no es el establecimiento de relaciones de causalidad entre las variables analizadas sino el análisis de la covariación o concomitancia entre las mismas, el investigador puede recurrir a los métodos psicométricos o correlacionales. Su aplicación es procedente cuando no se tiene información sobre los antecedentes del hecho que se estudia y no procede la asignación aleatoria de los sujetos a grupos diferenciados.

El estudio de las relaciones entre variables va a llevarse a cabo en un único momento temporal. El interés del investigador reside más en la búsqueda de regularidades o irregularidades en el conjunto de datos analizados que en el contraste de hipótesis específicas. Su principal contribución consiste en determinar la importancia relativa de las variables para la descripción y explicación de un hecho o fenómeno.

La finalidad, en la mayor parte de los casos, es descriptiva. Se estudia un conjunto de variables (una selección rigurosa de las mismas contribuye a clarificar la interpretación de los resultados) en una muestra con el fin de obtener información sobre el entramado de variables e individuos analizados. Los pasos a seguir en la aplicación de estos métodos serían:

- especificación de la población de interés;
- fijación de la forma de obtención de los datos;
- delimitación de las variables a estudio;
- selección o construcción de los instrumentos para la medida de las variables;
- forma o recogida de los datos (a través de encuestas y cuestionarios);
- establecimiento de las categorías de clasificación de los datos;
- tipo de análisis a realizar, en función del objetivo del estudio.

Este método puede concretarse en diseños muy variados, todos ellos encuadrados en una de estas dos categorías: diseños descriptivos o diseños correlacionales.

1) *Diseños descriptivos*. Su objetivo fundamental es posibilitar la comparación de comportamientos de grupos en diferentes variables, que no están bajo control del investigador sino que han ocurrido de forma natural y sin intervención experimental. Su finalidad, por tanto, es básicamente descriptiva. Sus limitaciones más importantes son: la falta de control sobre la variable criterio, la imposibilidad de establecimiento de relaciones causales y la distinta familiaridad que los grupos puedan tener respecto a los instrumentos de medida utilizados.

Como modalidad básica de este tipo de diseños destacamos los estudios de encuesta (tipo survey), con un esquema tipo de investigación que incluye las siguientes fases:

- Definición del problema,
- Establecimiento de hipótesis, considerandos,...
- Selección de variables
- Selección/elaboración de instrumentos
- Medida de variables
- Análisis de datos, conclusiones y toma de decisiones

2) *Diseños correlacionales*. Nos permiten comprender la complejidad del fenómeno o hecho social estudiado, determinando las variables relacionadas con el mismo o las dimensiones básicas en las que se puede descomponer. El esquema tipo de trabajo es similar al presentado para los diseños descriptivos. Pueden distinguirse a su vez tres modalidades:

- a) En la modalidad de «diseños de covariación» su finalidad es descriptiva resultan especialmente adecuados para los casos siguientes: realizar análisis exploratorios en campos poco estudiados o con estructuras complejas; establecer tipologías y/o dimensiones básicas en un conjunto de variables; conocer relaciones entre variables, que posteriormente pueden estudiarse con metodologías experimentales.
- b) La modalidad de «diseños correlacionales predictivos» se aplican cuando se tiene un mayor conocimiento de la relación entre las variables, de forma que pueden categorizarse como variables criterio o como variables predictoras (este *status* puede cambiar de una investigación a otra). Su objetivo fundamental es predecir, pronosticar, el comportamiento de un grupo de sujetos en una variable criterio determinada, a partir del conocimiento de su relación con un conjunto de variables predictoras.
- c) Una tercera modalidad de diseños correlacionales son los denominados «diseños correlacionales causales». Se sitúan en un punto intermedio entre la metodología correlacional y la experimental. Sobrepasan el nivel descriptivo y exploratorio aproximándose al explicativo, aunque para conseguir éste sea necesario que el modelo explicitado resulte de la deducción hipotética de una teoría. No permiten estos diseños generar nuevas teorías, sino contrastar empíricamente las deducciones obtenidas en un marco conceptual. Las relaciones causales se estudiarán a partir de las correlaciones empíricas entre las variables.

El argumento que lo posibilita es el hecho de considerar que la correlación no implica causalidad, mientras que la causalidad implica correlación (Batista y Coenders, 2000).

Todos los diseños correlacionales presentados habría que encuadrarlos en la categoría de «estáticos», en la medida en que se analizan las relaciones entre variables o los cambios ocurridos en momentos determinados de tiempo. La inclusión del «tiempo» como variable explicativa nos conduce a la consideración de «diseños dinámicos o evolutivos», de relativa complejidad y con un número importante de modalidades de aplicación: diseños evolutivos descriptivos, estrategias secuenciales, transversales y longitudinales, estrategias secuencial transversa, secuencial de cohorte, secuencial de tiempo..., todas ellas resultantes de la consideración aislada o conjunta de los componentes edad, momento de la medida, cohorte o generación a la que se pertenece.

B) La observación (científica y naturalista)

Esta estrategia metodológica básica de investigación en la Evaluación de Programas debe girar en torno a la observación realizada científicamente; es decir, utilizando técnicas que permitan lograr el máximo control y objetividad de la realidad o hecho educativo, siendo recomendada la utilización de diferentes modalidades metodológicas de observación: sistemática, participativa, externa,...

La observación desempeña un importante papel en la investigación al proporcionar a la ciencia los hechos, uno de sus elementos fundamentales. Mediante la observación, rigurosa y cuidada, el investigador va descubriendo pautas que le capacitan para ir dando forma a sus teorías. La observación científica supone la selección deliberada de algún aspecto significativo de los fenómenos, en cierta situación y en un momento determinado, un detenido examen que puede requerir el uso de procedimientos e instrumentos de precisión y la presentación de resultados.

La observación constituye pues uno de los aspectos más importantes del método científico (de hecho todas las ciencias comienzan por la observación), dependiendo, en gran parte, la consecución de los objetivos científicos del rigor y la precisión con que dicha observación se haya realizado. La observación requiere, para ser científica y en virtud de las condiciones que informan el quehacer científico, ser controlada, o sea, ser objetiva y comprobable; debe emplear una serie de procedimientos y técnicas que eliminen al máximo las posibles fuentes de distorsión y error y establecer adecuadamente las condiciones bajo las cuales se ha observado un determinado fenómeno.

La observación tiene como finalidad obtener información sobre algo, por lo que hay que empezar por delimitar el fenómeno a analizar, situándolo en el marco de una teoría, de la que parte y a la que pretende conformar. La observación debe ser realizada en condiciones específicamente establecidas, las cuales actúan a modo de reglas que establecen la naturaleza de la situación en que se realizarán las observaciones y los procedimientos o instrumentos que deberán ser utilizados.

En el contexto de la Evaluación de Programas, la observación científica se concreta como método que podemos denominar «observación naturalista». Este método se caracteriza porque el análisis de la realidad se basa en la observación de los comportamientos naturales del grupo social de referencia: clase, barrio, comunidad, distrito... Su objetivo fundamental es el estudio intensivo de los fenómenos complejos que conforman la vida de un grupo de personas, con finalidad comprensiva que debe conducirnos a propuestas para subsanar las carencias educativas y culturales percibidas.

Existen numerosas técnicas de observación, que se generan en función del ambiente en que se realiza (ambiente natural y ambiente artificial), en función de la utilización o no de registros observacionales (observación estructurada o no estructurada), y en función de la actitud que adopte el investigador respecto al proceso social observado (investigación participativa o no participativa).

El diseño determina la posibilidad de detectar y comprender los procesos y el impacto de un programa educativo, dada la multiplicidad de factores que influyen en su operación y resultados (De la Orden, 1990). Podemos considerar las siguientes pautas como esquema general de trabajo en la aplicación del método observacional (Martínez Arias, 1981; Garaigordobil, 1999):

- definir las metas del estudio
- establecer las muestras que serán observadas
- decidir si la observación será participante o no
- modo de entrar en contacto con los sujetos
- muestrear los ambientes y situaciones en que se desarrollará
- muestrear las unidades temporales
- establecer el procedimiento de recogida de datos
- fijar las categorías de clasificación de las conductas a observar
- considerar las posibles técnicas de organización y análisis de datos

Las ventajas fundamentales residen en la posibilidad de estudiar conductas no verbales y estudiar comportamientos en toda su complejidad. Sus principales inconvenientes para una aplicación optimizada serían la carencia de información sobre variables antecedentes que puedan ayudar a describir los hechos observados; los riesgos de carga subjetiva que pueden asociarse a las conclusiones obtenidas por el observador; los riesgos de reactividad de la presencia del observador (pueden evitarse con una presencia larga o encubierta); la dificultad de establecer categorías claras que orienten los registros; la utilización de muestras reducidas; el entrenamiento de los distintos observadores; y las dificultades del investigador para introducirse en el ambiente que se desea observar (Rodríguez, Gil y García Jiménez, 1996; Goetz y LeCompte, 1988; LeCompte, 1995).

La observación como técnica básica en la investigación cualitativa puede tener aplicaciones en diferentes formas y contextos, lo que determina las diferentes estrategias de investigación cualitativa, con notas peculiares y diversidad de concreciones:

estudio de casos, de grupos, estudios etnográficos (Rodríguez, Gil y García Jiménez, 1999).

C) El entorno interpretativo-crítico

La elaboración de un diseño desde la perspectiva crítica implica aceptar las siguientes premisas básicas (Colás y Rebollo, 1993, p. 112):

- El diseño está estrechamente vinculado, conceptual y prácticamente, con la acción-intervención
- Los receptores del programa se convierten en agentes de la evaluación
- Los procesos evaluativos los realizan y sirven a los implicados
- La secuencia evaluativa es cíclica: comienza con el diagnóstico y termina en la valoración global del contexto social y personal. No se pueden aislar aspectos o dimensiones del programa para su evaluación.
- El análisis e interpretación de los datos se hace de forma colectiva a la luz de un contexto social y político más amplio. En el análisis se utiliza información sobre cómo, quién, cuando y dónde se obtienen los datos

La investigación en la acción deben ser entendida como la síntesis de dos dimensiones indisolubles de un mismo proyecto; sólo así se puede comprender la práctica; la finalidad de la investigación es a la vez explicativa y normativa: se investiga para perfeccionar nuestro conocimiento, fundamentar racionalmente la práctica y transformar la realidad. La actividad desarrollada es a la vez un acto de investigación y un acto sustantivo.

La investigación-acción podría definirse como el estudio de una situación social (escolar) con el fin de mejorar la calidad de acción dentro de la misma. En la investigación-acción las teorías no se hacen válidas de formas aislada, para ser aplicadas después a la práctica. Se hacen válidas por medio de la práctica (Rodríguez Rojo, 1991).

La investigación-acción propuesta por Kurt Lewin (1951) posee las siguientes características:

- Es una actividad desarrollada por grupos o comunidades con el propósito de cambiar las circunstancias de acuerdo con una idea compartida.
- Es una práctica social reflexiva, donde no cabe hacer ninguna distinción entre la práctica que se investiga y el proceso de investigación de esa práctica. Las prácticas (sociales, escolares,...) son consideradas como «actos de investigación», como «teorías en acción» o «explicaciones hipotéticas que han de ser examinadas y valoradas en términos de su potencialidad para realizar cambios valiosos. Si una práctica social como la enseñanza es considerada como una actividad reflexiva ya no existe la división del trabajo entre los que la ejercen y los que la investigan (Elliot y otros, 1986, p. 12; Elliot, 1990; Elliot, 1993).

Lewin establecía un proceso disciplinado de investigación en la acción, con un claro paralelismo con el método científico, y articulado en una espiral de actividades realizadas en el siguiente orden:

- clarificar y diagnosticar una situación problemática para la práctica
- formular estrategias de acción para resolver el problema
- poner en práctica y evaluar las estrategias de acción
- nueva aclaración y diagnosis de la situación

Los puntos esenciales que definen y caracterizan la investigación-acción y el papel a desempeñar por esta estrategia metodológica de investigación en los procesos de innovación son analizados con detalle por Pérez Serrano (1990, pp. 53-54).

La investigación participativa representa una propuesta metodológica para el cambio social. Es una perspectiva científica o ideológica para estimular, apoyar y facilitar los procesos de transformación. Sus principios epistemológicos significan una manera de acercamiento a la realidad social: se trata de conocer transformando (Gabarrón y Hernández, 1994, p. 23).

BIBLIOGRAFÍA

- Alvira, F. (1986): «Diseños de investigación social: Criterios operativos» en F. Alvira, M. García Ferrando y J. Ibáñez *Análisis de la realidad social*. Madrid, Alianza Editorial.
- Alvira, F. (1991): *Metodología de la evaluación de programas*. Cuadernos Metodológicos nº 2, Madrid, CIS.
- Amezcuca, C. y Jiménez, A. (1996): *Evaluación de Programas Sociales*. Madrid, Díaz de Santos.
- Anguera, M.T. (1990): «Programas de intervención. ¿Hasta que punto es factible la evaluación?». *Revista de Investigación Educativa*, nº 16, pp. 77-93.
- Bartolomé, M. (1990): «Evaluación y optimización de los diseños de intervención». *Revista de Investigación Educativa*, nº 16, 39-60
- Batista, J.M. y Coenders, G. (2000): *Modelos de ecuaciones estructurales*. Madrid, La Muralla-Hespérides.
- Carballo, R. (1990): *Evaluación de un Programas de Formación Ocupacional*. Tesis Doctoral. Madrid, Universidad Complutense.
- Colás, P. y Rebollo, M^a A. (1993): *Evaluación de programas. Una guía práctica*. Sevilla, Kronos.
- Cronbach, L.J. (1982): *Designing Evaluations of Educational and Social Programs*. San Francisco, Jossey-Bass.
- De la Orden, A. (1990): «Evaluación de los efectos de los programas de intervención». *Revista de Investigación Educativa*, nº 16, 61-76.
- Elliot, J. y otros (1986): *Investigación en el aula*. Generalitat Valenciana, Valencia.
- Elliot, J. (1990): *La investigación-acción en educación*. Madrid, Morata.

- Elliot, J. (1993): *El cambio educativo desde la investigación-acción*. Madrid, Morata.
- Finnegan, J.R. y otros (1988): «Measuring and tracking education program implementation: The Minnesota Heart Health Program». *Health Educational Quarterly*, 16 (1), 77-90.
- Fitz-Gibbon, C.T. y Morris, L.L. (1978): *How to design a Program Evaluation*.
- Gabarrón, L.R. y Hernández Landa, L. (1994): *Investigación participativa*. Cuadernos Metodológicos. Madrid, CIS.
- Garaigordobil, M. (1999): «Técnicas observacionales en la construcción de instrumentos de evaluación psicológica y en la evaluación continua de programas». En M.T. Anguera *Observación en el aula: Aplicaciones*. Barcelona, Ed. Universitat de Barcelona, Monografía de psicología.
- Goetz, J.D. y Lecompte, M.D. (1988): *Etnografía y diseño cualitativo en investigación educativa*. Madrid, Morata.
- Guba, E.G. (1985): «Criterios de credibilidad en la investigación naturalista» en J. Jimeno y A. Pérez (eds.) *La enseñanza: su teoría y su práctica*. Madrid, Akal.
- Kerlinger, F.N. (1977): «The influence of Reserach on Education Practice». *Educational Researcher*, Vol. 6, nº 8, pp. 5-12.
- Lecompte, M.D. (1995): «Un matrimonio conveniente: diseño de investigación cualitativa y estándares para la evaluación de programas». *Revista Electrónica de Investigación y Evaluación Educativa*, I, 1.
- Lewin, K. (1951): *Field theory in Social Science*. New York, Harper.
- Martínez Arias, R. (1981): «Diseños de investigación en psicología ambiental». En A. Jiménez Burillo *Psicología y medio ambiente*. Madrid, Ceotma, MOPU.
- Martínez Mediano, C. (1997): «La teoría de la evaluación de programas». En AIDIPE *Actas del VIII Congreso de Modelos de Investigación Educativa*.
- Mayer, R.R. (1985): *Policy and program planning*. Englewood Cliffs, N.J., Prentice Hall.
- Pérez Carbonell, A. (1998): *Metodología de la Evaluación de Programas: Evaluación de Programas de Postgrado en la Universidad de Valencia*. Tesis Doctoral. Valencia, Universidad de Valencia.
- Pérez Juste, R. (1995): «Evaluación de programas educativos». En L.M. Villar y A. Medina *La evaluación de programas educativos, centros y profesores*. Madrid, Universitat.
- Pérez Serrano, G. (1990): *Investigación-acción. Aplicaciones al campo social y educativo*. Madrid, Dykinson.
- Rodríguez, G., Gil, J. y García Jiménez, E. (1999): *Metodología de investigación cualitativa*. Málaga, Aljibe.
- Rodríguez Rojo, M. (1991): «Cómo aprender qué es una investigación acción». *Investigación en el Aula*, 13, 59-66.
- Sanders, J.R. (1998): *Estándares para la evaluación de programas (Comité Conjunto de Estándares para la Evaluación Educativa)*. Bilbao, Mensajero.
- Scriven, M.S. (1973): «Goal-free evaluation» en E.R. House *School Evaluations: The politics and process*. Berkeley, California, McCutchan.
- Tejedor, F.J. (1990): «Perspectiva metodológica del diagnóstico y evaluación de necesidades en el ámbito educativo». *Revista de Investigación Educativa*, nº 16, 15-38.

- Tejedor, F.J. (1994): «La experimentación como método de investigación educativa», en V. García Hoz *Problemas y métodos de investigación en educación personalizada*, pp. 256-285. Madrid, Rialp.
- Tejedor, F.J., García-Valcárcel, A. y Rodríguez, M^a J. (1994): «Perspectivas metodológicas de la evaluación de programas en el ámbito educativo». *Revista de Investigación Educativa*, nº 24, pp. 93-128.
- Weis, C.H. (1982): *Investigación evaluativa. Métodos para determinar la eficiencia de los programas de acción social*. México, Trillas.
- Witkin, B.N. (1984): *Assessing needs in Educational and Social Programs*. London, Jossey-Bass Publ.