

CAPÍTULO I. INVESTIGACIÓN EVALUATIVA Y EVALUACIÓN DE PROGRAMAS.

“Una de las actividades básicas de todo profesional de la educación es diseñar y aplicar programas para dar respuesta a la cada vez más numerosa demanda educativa por parte de la sociedad...

... Por tanto el diseño y evaluación de programas es una actividad fundamental, hoy día, del profesional de la educación.”

(Colás, 2000)

1.1. INTRODUCCIÓN.

En cualquier ámbito social, en cualquier Sistema Educativo, no puede entenderse una intervención sin el establecimiento de un programa, de unas líneas claras de actuación ordenadas temáticamente y temporalmente que ofrezcan coherencia y sentido a dicha intervención (Forner y Latorre, 1996). Sobre todo encontrándonos en un periodo de reordenación educativa en el que hay que justificar política, económica y socialmente las actuaciones de intervención.

En el mismo sentido, puede entenderse lo necesario de la evaluación de las intervenciones pedagógicas para promover su eficacia, mejora y justificar su rendimiento. La evaluación que desde la L.G.E.¹ de 1970 ha recibido una atención creciente, no solo como medio a través del cual mejorar el rendimiento de los estudiantes sino también de todo el proceso educativo, como criterio de su calidad (Pérez Carbonell *et al.*, 1999).

En 1987, con el proyecto de Reforma de la enseñanza, la evaluación sigue considerándose en términos semejantes, aunque se amplía su ámbito introduciéndose la evaluación de todas las personas involucradas en la acción educativa. Es un periodo en el que se producen grandes cambios sociales que generan innumerables demandas del proceso evaluativo. Siendo en estos momentos la LOGSE² el mejor marco para fomentar y favorecer la Evaluación Educativa y de forma más concreta la Evaluación de Programas.

En el mismo sentido que se plantea la necesidad del diseño e implementación de planes y programas como medio de intervención educativa, lo hace la evaluación de éstos, por lo que podría decirse que la evaluación de programas es, en la actualidad, inherente al hecho educativo.

Los contenidos de este primer capítulo aluden a la génesis de los estudios evaluativos y a la delimitación de evaluación de programas como campo disciplinar en la investigación educativa.

En su primer epígrafe se realiza una introducción a la caracterización de la Investigación Educativa (IE, en adelante) en general, y particularmente, a aquella orientada hacia un sentido evaluativo de las realidades analizadas. Haciendo mención específica a los estudios de síntesis, como mecanismos de obtención de visiones globales acerca de un campo de investigación.

¹ L.G.E.: Ley General de Educación.

² L.O.G.S.E.: Ley de ordenación general del sistema Educativo.

A continuación, se describe como los procesos de investigación evaluativa aplicados al campo de la evaluación de programas han alcanzado unas características específicas en cuanto a su objeto de estudio y los medios para estudiarlo, de forma que se ha constituido en un campo disciplinar diferenciado de la I.E.

Este campo de investigación posee ciertas peculiaridades derivadas de la convergencia en él de numerosas disciplinas pedagógicas, que han determinado su desarrollo. Las relaciones que mantiene con ellas son analizadas en un epígrafe específico de este capítulo.

Capítulo que finaliza, con una visión en sentido diacrónico y sincrónico, del desarrollo histórico de la disciplina de la evaluación de programas en el ámbito internacional y en el contexto específico español.

1.2. LA INVESTIGACIÓN EDUCATIVA, LA INVESTIGACIÓN EVALUATIVA Y LOS ESTUDIOS DE SÍNTESIS.

El origen de la investigación educativa se halla estrechamente vinculado al nacimiento de la pedagogía como disciplina científica (Colás y Buendía, 1998), que por reivindicar tal carácter pretende generar conocimientos científicamente aceptables de los hechos educativos.

Pero este establecimiento de una ciencia de la educación como disciplina social estuvo asociado a numerosas dificultades, como el abandono de una perspectiva moral, para adoptar una adecuación positivista en la que se hiciera posible la objetivación y por tanto un carácter empírico consustancial al método científico. Y también, el abandono de los planteamientos, metafóricos y simbolistas, basados en la descripción de la acción educativa; por la posibilidad de actuación y transformación con vistas al desarrollo y mejora.

Esta incorporación del método científico-experimental al estudio de la educación, lo relaciona García Hoz (1978) con la transferencia desde disciplinas afines como la medicina y la psicología, al intentar desvelar cuestiones relacionadas con temas educativos.

Desde este particular origen, el enraizamiento y desarrollo de la investigación educativa no se produce simultáneamente en los distintos contextos, sino que ha dependido en gran cuantía de las condiciones socio-políticas particulares.

Así, en Estados Unidos la investigación pedagógica ha alcanzado un alto nivel de desarrollo, siendo precisamente allí dónde en 1882 apareció el primer seminario con el propósito de crear una pedagogía de carácter científico.

Bartolomé (1984) sintetiza los logros generales de la investigación educativa norteamericana en la introducción de instrumentos de medida y evaluación, desde los tests de rendimiento a los más recientes de carácter ecológico. Desde la incorporación del método experimental a los estudios pedagógicos (Campbell y Stanley, 1982), la introducción de nuevos diseños mejor adaptados a las peculiares condiciones de los contextos educativos, cuasiexperimentales de Cook y Campbell (1979) y de caso único de Kratochwill (1978).

Y también, desde una perspectiva funcional, la mejora del tratamiento de datos mediante la incorporación de métodos multivariados apoyados por los soportes informáticos; o la potenciación de la difusión de la información a través de diversos soportes como la realizada por la A.E.R.A.³, que sintetiza, recopila y publica las principales aportaciones empíricas en el ámbito educativo. Todo ello en un fuerte ámbito de apoyo económico y político.

³ A.E.R.A.: *American Educational Research Association.*

En Europa, sin embargo, la investigación educativa se ha realizado en general de una forma mucho más aislada, con escaso reconocimiento e insuficiente colaboración de las instituciones oficiales, de una manera dispar.

La tradición investigadora sobre educación se remonta, en Gran Bretaña, a la creación del Instituto de Pedagogía de la Universidad de Londres en 1931. Destacando entre sus contribuciones más relevantes, por medio de la creación de centros de prestigio y la relevancia de ciertas figuras, o la publicación de revistas de prestigio mundial como *British Journal of Educational Technology*, *Research into Higher Education*, *Educational Research* o *British Journal of Educational Psychology*, la construcción de tests, la exploración de la inteligencia y delimitación de diferencias individuales, el diagnóstico de aptitudes escolares o la predicción del rendimiento, tomando como factor las implicaciones sociales.

En Alemania, la investigación educativa tiene sus orígenes en los institutos psicológicos que han tratado aspectos diferenciales y psicológicos del desarrollo, siendo de gran importancia las aportaciones metodológicas cualitativas y la filosofía humanista, bajo el liderazgo de la escuela de Frankfurt y autores como Habermas y Dilthey.

Cabe citar, en el ámbito francófono, el prolífico pasado en la primera veintena del siglo veinte, de las instituciones educativas belgas, como la Universidad de Bruselas. Y de las francesas a partir de los años cuarenta.

Desde la revolución bolchevique de 1923, en Rusia, se hace patente la preocupación por la pedagogía, que lleva a fundar en 1943 la Academia de las Ciencias Pedagógicas. Enmarcándose todos sus estudios en planes quinquenales con planes de evaluación, centrados en el estudio de los contenidos curriculares, de métodos de enseñanza y formación del profesorado.

En el caso de España, desde el primer tercio del siglo veinte se aprecia una sensibilización hacia la investigación educativa, plasmada en los años treinta

con la creación de las secciones de pedagogía en las Universidades, o con la edición de la Revista Española de Pedagogía una década más tarde.

Si embargo no es hasta los movimientos políticos, económicos y sociales de los años sesenta, que culminan con el establecimiento de una política educativa regida por la Ley General de Educación de 1970, cuando se comienza a hacer hincapié en ella coordinándola, asignándole fondos y creando instituciones como los ICEs⁴ encargados de llevarla a cabo.

La investigación en estos centros se caracteriza por su conexión con la formación del profesorado y la búsqueda de nuevas soluciones a los problemas de la sociedad y de la educación, el predominio del enfoque descriptivo, apoyado especialmente en la encuesta, observación y elaboración de pruebas y materiales, con una introducción creciente de métodos explicativos, predictivos y experimentales. Aunque de importancia, Tejedor, Castro y Mínguez (1988) caracteriza estas aportaciones de irregulares y declinantes.

Con posterioridad, el perfeccionamiento del profesorado, de los niveles básicos y secundarios, pasa a los CEPs⁵. Asignándole también a éstos funciones de investigación aplicada, que Escolano (1982) describe con una clara orientación práctica y corte didáctico, con la adopción de una metodología eminentemente empirista, realizada por equipos de investigación y con una proyección de planes en el ámbito nacional.

Este tipo de actuaciones para el desarrollo de la investigación pedagógica, en los niveles primarios y secundarios del sistema educativo, según Tejedor, Castro y Mínguez (1988) apenas ha tenido incidencia efectiva en la realidad educativa, ni influencia en las estructuras administrativas, ni generado innovaciones en las instituciones.

⁴ I.C.E.: Instituto de Ciencias de la Educación, asociados a las Universidades.

⁵ C.E.P.: Centros de Profesores.

La labor investigativa en educación ha quedado sobre todo relegada al ámbito universitario, por medio de trabajos que Colás y Buendía (1998) describen como trabajos personales, no de equipo; con temáticas que pudieran tener gran importancia científica y el uso de avanzadas y sofisticadas metodologías, pero con escaso interés práctico; y con una escasa continuidad, que dificulta la generalización y profundización de líneas de investigación.

Todos estos graves problemas de la investigación educativa, han tratado de solucionarse en las últimas décadas del siglo veinte, por medio de su reconsideración conceptual y metodológica, y el desarrollo de nuevos planteamientos sobre la auténtica finalidad de la investigación y la relación que ésta debe guardar con la práctica educativa.

La dinamización de todas estas propuestas debe comenzar por la formación y desarrollo profesional, en el que las técnicas y métodos de investigación ocupan un papel relevante. Formación que debe comenzar por los estudios universitarios sobre educación, alcanzando también a la práctica profesional.

De forma concreta deben plantearse concepciones y aproximaciones metodológicas más adecuadas a la realidad educativa actual, que permitan un mayor impacto de los hallazgos.

En este sentido, Colás y Buendía (1998), describen la evolución metodológica en los últimos años hacia la ampliación de las formas de proceder, una investigación orientada a la toma de decisiones que va ganando terreno a la investigación orientada a la obtención de decisiones. Y la apertura de nuevos campos de investigación tales como la investigación evaluativa, y la potenciación de la investigación-acción como forma de vincular la innovación y la práctica escolar.

Estas nuevas prácticas profesionales, generadas desde los intereses y necesidades de los propios profesionales, conciben la investigación como una actividad próxima a la práctica educativa e imprescindible en el que hacer profesional, desde nuevos planteamientos metodológicos, como es el caso de las perspectivas cualitativas o de la investigación-acción.

Perspectiva desde la que a los profesionales de la educación, como queda patente en los nuevos diseños curriculares de las enseñanzas, se les asigna un nuevo rol en la participación de construcción del conocimiento científico. Conocimiento para solucionar problemas y promover el desarrollo de la calidad educativa.

Es desde esta perspectiva, desde dónde puede entenderse la auténtica finalidad de la investigación en el campo educativo, en los términos que especifica Hernández Pina (1998a), como el estudio de los métodos, los procedimientos y las técnicas utilizadas par obtener un conocimiento, una explicación y una comprensión científica de los fenómenos educativos, así como para solucionar los problemas educativos y sociales. Orientándose más hacia planteamientos prácticos, que a la generación y resolución de problemas teóricos.

También en la organización y financiación de la investigación se observa también una mayor diversificación de centros y programas y la ampliación de los canales de financiación y difusión, con el incremento de los seminarios y congresos y la generación de fuentes de publicación de gran importancia, como es el caso de la *Revista de Investigación Educativa*.

Sin embargo, en general, la investigación educativa sigue adoleciendo de las limitaciones de las disciplinas incipientes o en proceso de reestructuración. Sólo una enorme seriedad en las fundamentaciones teóricas, mayor rigor en los procesos metodológicos, aumento de los canales y estilos en la comunicación harán generar la suficiente consistencia científica como para ofrecer mayor

credibilidad y mejorar las posibilidades de difusión (Slavin, 1997; Klein, 1997 y Pogrow, 1998).

La caracterización de la investigación evaluativa, entendiendo ésta como aquella que posee una finalidad valorativa, en cualquiera de los objetos de estudio a los que pueda referirse, sigue como investigación educativa un proceso similar de implantación y desarrollo.

Arnal *et al.* (1992), muestran estas coincidencias entre la investigación educativa general, la investigación evaluativa y de forma específica la referida al campo de la evaluación de programas, en la que se puede apreciar también un periodo de profundos cambios.

La investigación evaluativa debe entenderse como un proceso riguroso, controlado y sistemático de recogida y análisis de información válida y fiable para tomar decisiones acerca del programa educativo; y así aumentar la racionalidad de las decisiones acerca de la puesta en marcha, sobre su desarrollo y evaluación evitando la excesiva dependencia de las decisiones políticas (Tejedor, García-Varcárcel y Rodríguez, 1994).

En su aspecto general es similar a cualquier proceso investigativo, distinguiéndose más por los objetivos que por los métodos empleados; por lo que podríamos entenderla como una mera extensión, con ciertas caracterizaciones propias, hacia un campo práctico concreto. Sin embargo, siendo algunas sus diferencias, como señala De la Orden en 1985, conviene resaltar ciertas características propias de la investigación evaluativa:

- En todo el proceso de investigación evaluativo, desde la selección del problema hasta la aplicación de una metodología concreta, viene acompañado de juicios de valor acerca del programa.

- La complejidad de su objeto de conocimiento y los contextos en que se desempeñan hacen enormemente complejo tanto la formulación de hipótesis precisas, como un completo control de las variables intervinientes, limitándose el uso de los diseños con carácter más experimental.
- La recogida de datos queda condicionada a la viabilidad del proceso y posibilidades de los sujetos intervinientes.
- La replicación, dadas las peculiares e irrepitibles características de la ejecución, es prácticamente imposible.
- La toma de decisiones sobre la paralización, sustitución o repetición del programa no queda en manos del evaluador; adaptándose al igual que el informe a las exigencias de quien lo financia.

También, Tejedor, García-Varcárcel y Rodríguez (1994), caracterizan la investigación evaluativa, en una comparación con la investigación educativa general, en algunos aspectos fundamentales:

- Mientras el propósito de la investigación es el incremento del conocimiento para sacar conclusiones generalizables a otros ámbitos, el proceso de evaluación de programas es de intervención en la toma de decisiones de problemas prácticos limitados a un grupo de referencia.
- En la investigación el interés personal o científico delimita el problema sobre el que se lanzan unas hipótesis explícitas. Por medio de un proceso replicable, basado en un diseño preestablecido y una metodología usualmente cuantitativa, el investigador controla y manipula las variables para obtener datos de una muestra aleatorizada, e intentar alcanzar de manera autónoma un conocimiento objetivo.

Sin embargo, en la investigación evaluativa, son numerosos los aspectos que pueden venir asignados por condicionantes externos, como son el tema y/o el grupo de interés, impidiendo un muestreo aleatorizado. Es un proceso difícilmente replicable el que sin hipótesis explícitas, a través de diseños flexibles en los que difícilmente se van a poder controlar las variables, y por medio de metodologías diversas, el investigador intenta asignar criterios de valor a un determinado programa.

- En otros aspectos, mientras la investigación encuentra su validez internamente, externamente y mediante constructos teóricos, se interpretan los datos de acuerdo a reglas definidas, adaptando los informes a la comunidad científica; la evaluación de programas, basa su credibilidad en la propia utilidad que permita valorar el programa, adaptando sus interpretaciones a los gestores y usuarios.

Precisamente en periodos de transición como éstos, en los que han variado las cosas en evaluación educativa, su uso se ha profesionalizado e institucionalizado (Madaus, 1983), han aumentado sus objetos, métodos y audiencias (Stufflebean, 1987), y se proponen la elaboración de estudios de evaluación o metaevaluación (Scriven, 1974; Stufflebean, 1974,1978; Cook y Gruder, 1978).

1.3. LA EVALUACIÓN EDUCATIVA Y LA EVALUACIÓN DE PROGRAMAS COMO CAMPO DISCIPLINAR.

El origen de la palabra evaluación se encuentra en *valore*, del latín *tasar*, valorar o atribuir valor a una cosa.

En la lengua española muestra, por tanto, una clara descendencia etimológica al considerar sinónimos los términos valorar y evaluación. Aunque pese a tal sinonimia, existen matices y diferencias en su aplicación, puesto que incluso atendiendo a cuestiones de valoración social, se estima más el término

valorar. En otras lenguas, como la inglesa, se realizan distinciones entre términos, como *evaluation* o *assessment*, si se evalúan objetos o personas.

Como en su origen, no existe una definición única de evaluación de programas, pues la aceptación de una, presupone la aceptación de algunas consideraciones básicas que determinan a ésta. Así puede ser entendida, como el proceso de determinar en qué medida los objetivos de un programa se han cumplido (Tyler, 1950), suponer la estimación del mérito de un objeto (Scriven, 1967) o poner a examen los resultados o efectos de un programa (Levin, 1975).

Funcionalmente, la evaluación de programas comienza a ser una preocupación dentro del ámbito educativo en los años setenta, como medio para valorar la eficacia de los programas de acción social que se estaban llevando a cabo y comprobar la consecución de sus objetivos previstos.

Esta acción de estimación o juicio de valor de un determinado programa exige un proceso contextualizado de análisis normativo científico, en el que se recoja, analice e interprete informaciones validas y fiables que conduzca a la evaluación de dichos procesos y a la toma de decisiones (Mateo, 1991). Es decir, una evaluación debería poseer los rasgos característicos de una investigación científica.

Así, la evaluación de programas hay que entenderla como un proceso sistemático de obtención y análisis de informaciones significativas en las que basar juicios de valor sobre estos programas. De forma más específica (Fernández Ballesteros, 1996) como un conjunto de conocimientos teóricos y metodológicos, con ámbito de aplicación en las Ciencias Sociales, para dar respuesta a la necesidad de enjuiciamiento de ciertas decisiones políticas cuyo fin es el beneficio de la intervención social. Pero dicho mecanismo debe suponer una secuencia dinámica de diseño, aplicación, evaluación, re-diseño (Alvarez Rojo y Hernández Fernández, 1997).

Este proceso puede llegar a ser enormemente complejo, como señala Colás (1993) ya que en la medida que un programa de actuación pública o carácter social sea más ambicioso, más complejas son las realidades a analizar que además se desarrollan dinámicamente en sistemas interdependientes.

La gran diversidad de propuestas y planteamientos hace de éste un campo difícil de concretar y conceptualizar, resaltando cada autor los aspectos que le resultan más relevantes. Aunque buscando patrones de similitud podemos encontrar dos tipos básicos de apreciaciones: los que ponen hincapié en la dimensión técnica y los que ponen el énfasis en los aspectos ideológicos, aunque ni siquiera estas caracterizaciones extremas pueden considerarse tan distintas puesto que realmente no son opuestas ni contradictorias, resultando en el fondo complementarias y de importancia más sustancial, dependiendo del momento del proceso evaluativo: selección del contenido del programa, toma de decisiones, implicación de los participantes, recogida y análisis de la información, etc.

En esta diversidad de propuestas sobre evaluación de programas existen algunos aspectos claves que se han pretendido enmarcar en unos parámetros o estándares claros que la delimitasen (Sanders, 1985; Stufflebean, 1987; *Joint Committe*, 1998). Aunque este esfuerzo de definición ha recibido numerosas críticas (Cronbach *et. al.*, 1980; Stake, 1981; Talmage, 1982 y Guba, 1988), por considerarlo como una restricción conceptual del campo disciplinar.

Aunque se parta de una gran diversidad epistemológica y/o metodológica, de forma común, la evaluación se entiende un proceso dinámico, en el que aplicando procedimientos científicos, se recogen informaciones válidas y fiables, se realizan juicios de valor para tomar decisiones contextualizadas sobre un determinado programa.

Su finalidad puede entenderse desde la medición de la implementación y resultados de un programa (Rutman, 1980) a la toma de decisiones para una intervención (Cronbach, 1982), que puede servir a una intervención política

(Talmage, 1982). Así podremos definir, la evaluación de programas como un conjunto de conocimientos teóricos y metodológicos con ámbito de aplicación en las ciencias sociales, pretende dar respuesta a la necesidad de enjuiciamiento de ciertas decisiones políticas cuyo fin es el beneficio de la intervención social (Fernández-Ballesteros, 1996). Pero no sólo describe los resultados de la evaluación, en otro sentido también pretende explicar los hallazgos encontrados (relación entre variables, incidencia de unas en otras,...) construyendo su propio campo de conocimiento (Tejedor, García-Varcárcel y Rodríguez, 1994).

Para Alvira (1991), algunas de las circunstancias que han contribuido a conformar este campo de actuación serían:

- Mayor interrelación entre las actividades de evaluación y programación. Si en una perspectiva clásica, las actividades evaluativas se posponían a la planificación y desarrollo del programa, en la actualidad son procesos paralelos.
- La evaluación de programas desde una perspectiva metodológica es investigación evaluativa. Es investigación, con todos los condicionamientos epistemológicos, metodológicos y técnicos; centrada en el carácter evaluativo.
- En la evaluación de programas se produce una mayor implicación de las partes interesadas, tanto de quienes patrocinan el trabajo, como de los potenciales beneficiarios.
- En la evaluación de programas pueden integrarse estrategias de investigación muy diversas y complementarias, siendo la naturaleza del programa y la habilidad del evaluador, la que determine una o varias opciones.

En los últimos años la evaluación de programas, aunque de forma rápida y desordenada, se ha desarrollado como campo disciplinar con gran potencial de crecimiento en el futuro (Municio, 1992). Es un campo disciplinar porque en él

precipitan múltiples disciplinas (Psicología, pedagogía, sociología,...). Y ha ido perfilándose como un campo concreto, delimitado por una terminología propia, un conjunto de herramientas conceptuales y analíticas específicas, y unos procesos y fases también específicos.

En su desarrollo se ha tendido a consolidar una teoría de síntesis (Shadish, Cook y Levinton, 1991), la realización de evaluaciones sofisticadas y la publicación de gran cantidad de libros, artículos e informes de evaluación.

Algunos autores como Colás (1993), Gairín (1991) o Pérez Juste (2000) señalan aspectos básicos de caracterización general la investigación y evaluación de programas y que facilitan su delimitación como campo disciplinar específico:

- La evaluación de programas no debe entenderse como un proceso uniforme con un modelo único de evaluación o con una única metodología.
- La evaluación de programas se ajustará inicialmente a la idea que se tenga de la evaluación. La concepción más amplia es aquella que entiende que evaluación es el conocimiento empírico del mérito o valor de las actividades y objetivos de un programa.
- Irá referida a un programa o conjunto sistemático de actuaciones que se ponen en marcha para alcanzar unos objetivos.
- Junto a la función educativa, el programa debe estar cumpliendo otras funciones de carácter ideológico o político.
- Planificación y evaluación deben pensarse simultáneamente. Considerando incluso la evaluación de necesidades como un tarea previa a la planificación de un programa.

- Los modelos de evaluación deben ser al mismo tiempo modelo de análisis y de toma de decisiones. Detallando la planificación–evaluación de un programa para facilitar las tareas de un control posterior.
- El evaluador debe estar familiarizado con el objeto de la evaluación procurando, mediante el contacto y la negociación previa, la participación de los interesados, ya que aportan validez al proceso y garantía de utilidad y aplicabilidad a los resultados. Aunque en algunos diseños no se incluye de forma explícita o se duda de su utilidad, el problema actual no se plantea en términos de admitir o rechazar la participación de los interesados (gestores y destinatarios) en el proceso evaluativo, sino en delimitar la forma y el grado en que debe realizarse.
- Los criterios o indicadores son básicos en la evaluación así como el establecimiento de estándares o grados aceptables de éxito.
- La evaluación se entiende como proceso, confiriéndole un carácter dinámico. El concepto concreto de evaluación que se adopte marcará la elección de posibles modelos teóricos y enfoques metodológicos a seguir, así como otros aspectos a considerar (tema, actividades, rol del evaluador, audiencia o informes).
- Deben aplicarse procedimientos científicos, que incluyan estrategias de diseño, recogida y análisis de información rigurosos y sistematizados, lo que no excluye en modo alguno la diversidad de enfoques metodológicos (técnicas de recogida y análisis de datos o estrategias de toma de decisiones). Tampoco el carácter científico del proceso evaluador debe hacernos concebir dicho proceso como una actividad exclusivamente teórica, mera aplicación de los procedimientos científicos.
- Actualmente el énfasis de la tarea evaluativa se pone en la emisión del juicio de valor, a partir de una adecuada recogida y análisis de la información pertinente.

- Deben superarse las opciones excluyentes de metodologías específicas, la evaluación de programas permite utilizar de forma conjunta y coordinada metodologías distintas.
- No todos los diseños son válidos, debiendo responder a criterios de validez y utilidad. Además de a diversas tipologías dependiendo del momento o fase del programa que quiera analizarse: de necesidades, de proceso, de producto...
- Debe tener como proyección la toma de decisiones, lo que implica otorgarle una utilidad a la evaluación.
- Se remarca el carácter contextualizado, en el sentido que ha de estar condicionada por las circunstancias y ha de ajustarse a las condiciones reales en las que se aplica el programa objeto de la evaluación.
- Hay que prestar especial atención a los efectos no previstos, ya que pueden deberse tanto a la potencialidad del programa como a su incorrecta aplicación. El éxito del programa no se mide únicamente por la consecución de los objetivos ya que hay programas que no los formulan o que se plantean desde otras perspectivas menos funcionales.

Todos estos aspectos dejan patente la delimitación de la investigación sobre evaluación de programas como un campo de la investigación educativa con características específicas diferenciales. No sin dificultades, como señalan Colás y Rebollo (1993), por las escasas realizaciones sistemáticas y el predominio de determinados enfoques. Siendo una consecuencia de este panorama la dificultad de su conceptualización, encontrándose múltiples definiciones en la que cada autor destaca lo que más le interesa.

1.4. RELACIÓN ENTRE LA EVALUACIÓN DE PROGRAMAS Y LAS DISCIPLINAS PEDAGÓGICAS.

La Evaluación de Programas podría ser uno de los campos disciplinares del ámbito educativo que reciben un mayor número de influencias de diversas disciplinas pedagógicas. Esta multiplicidad produce variadas interacciones entre ellas.

Así en un sentido, éstas ofrecen su corpus axiológico como principios de ordenación de las realizaciones teórico prácticas delimitando consistentes diseños de investigación y facilitan procesos metodológicos, aportando instrumentos de recogida y análisis de informaciones válidos y fiables. En definitiva, vierten toda una serie de principios que orientan la investigación en E.P.

Pero a su vez todas éstas disciplinas requieren de la E.P. que ésta centre su objeto de conocimiento en contenidos específicos del interés particular, mediante aproximaciones y realizaciones investigativas acordes con los principios teórico prácticos específicos de cada una de ellas.

Por ello es común que en la investigación sobre E.P. se recurran a teorías propias de la Orientación o de la Didáctica, se empleen procesos examinados en la Metodología de la Evaluación o se apliquen instrumentos característicos de multitud de disciplinas pedagógicas. Centrando a su vez su objeto de conocimiento en aspectos del interés de cada una de estas disciplinas.

Esta actuación de diversas áreas en un mismo ámbito puede entenderse como un rasgo de riqueza disciplinar o plantear contradicciones si en lugar del empleo de aproximaciones complementaristas se acude a absurdas posturas solapadas y/o excluyentes.

El origen de la evaluación de programas se encuentra muy vinculado a la Orientación Educativa, puesto que los modelos de intervención por programas

comienzan a utilizarse en EEUU en los años sesenta, introduciéndose su implementación en España en los años ochenta con la implementación de los servicios de orientación educativa (Álvarez Rojo y Hernández Fernández, 1997). Es por ello que es ésta un área de conocimiento habitual en las temáticas evaluativas.

La E.P. como campo disciplinar eminentemente empírico y aplicado, en el que vierten sus conocimientos numerosas disciplinas, ha sido un caldo de cultivo excelente en el que plantear diseños y métodos de la más diversa índole desde perspectivas complementarias y eclécticas, por lo que también ha sido foco de atención para la Metodología de la Investigación. Prueba palpable de ello es el papel crucial que juegan los Departamentos de Métodos de Investigación como filiación institucional de los autores más importantes en este campo.

Las intervenciones por programas limitados y las consiguientes evaluaciones de éstos, ofrecen también una oportunidad excelente a la Didáctica de considerar y experimentar diversos aspectos. Es por ello que son comunes las evaluaciones de programas que centran gran parte de su interés en cuestiones específicas, como la Didáctica de Idiomas, la Didáctica para la atención a N.E.E. o las Didácticas para la intervención en contextos sociales desfavorecidos.

Pero quizás una de las relaciones más paradójicas de la E.P. es la que establece con la Teoría de la Pedagogía. Puesto que pese a ser la E.P. un campo de experimentación práctica no se ha desarrollado especialmente como centro de ensayo de las teorías pedagógicas, sino que más bien se ha limitado a emplear un puñado relativamente pequeño de ellas como justificación epistemológica de sus realizaciones. Es precisamente ésta una de las debilidades de este campo, denotando las en ocasiones difíciles relaciones entre la teórica y la práctica educativa real y aplicada, suponiendo una pérdida de oportunidades por parte de los investigadores educativos.

Para completar esta demarcación, desde una perspectiva diacrónica, se considera su génesis y desarrollo histórico como disciplina en el siguiente epígrafe.

1.5. LA EVALUACIÓN DE PROGRAMAS EN EL ÁMBITO INTERNACIONAL. DESARROLLO HISTÓRICO.

El empleo de programas como estrategia de intervención educativa tiene una historia relativamente reciente, en la que es posible constatar la evolución de la propia concepción de los que es un programa y cómo debe ser su evaluación, como los avatares de su coexistencia con el resto de la investigación educativa.

Como en muchos otros campos de la investigación educativa, la evaluación de programas encuentra a sus pioneros en EEUU, donde desde finales del siglo XIX el sentido de la evaluación era entendido como sinónimo de aplicación de test para marcar diferencias individuales dentro de un grupo normativo, con poca relación con la evaluación de programas (Guba y Lincoln, 1985).

En los años 30 del siglo XX, se comienzan a promover modelos centrados en la experimentación para comprobar la relación entre objetivos/resultados. Focalizando la evaluación su atención en la detección de diferencias individuales, mediante puntuaciones específicas y comprobar el nivel de cumplimiento de parámetros previamente establecidos. Consolidándose esta etapa entorno a los años cincuenta, en los trabajos de Tyler, que hacen referencia a los objetivos del programa y a la descripción de la medida.

Para algunos autores es aquí dónde comienza y se consolida la Evaluación de Programas, tal y como hoy se entiende, en el entorno de las actividades de evaluación de objetivos desarrolladas por Tyler. Modelo que ha tenido una gran influencia posterior debido a su lógica básica, y a que era científicamente aceptable y fácil de aplicar (Carballo, 1990).

En los planteamientos de este autor, el principal papel de la evaluación consistía en verificar la consecución de los objetivos propuestos de forma previa a la realización de la actividad. Y cuanto más claramente estaban especificados estos objetivos, más sencilla resultaría su evaluación, mediante las medidas *pre* y *post* programa.

La sustitución de estos modelos basados en las concepciones taylerianas se produce precisamente por su reduccionismo, ya que la consideración exclusiva de la consecución de objetivos no parece un factor para medir el éxito global de un programa.

En los años sesenta con la promulgación de la Ley E.S.E.A. en la que se concede gran importancia a los mecanismos evaluativos dentro del propio proceso educativo, se comienza a hablar de programas, sobre todo en el ámbito de la orientación psicoeducativa, en su doble faceta profesional de *guidance* y *counseling*. Siendo en esta segunda mitad del s. XX, cuando se produce un gran auge de la disciplina, evolucionando de la etapa de inocencia al realismo.

Las nuevas pretensiones no alcanzan tan sólo a aspectos como la medida, sino que además se pretende describir e interpretar; aunque todavía bajo el predominio de la evaluación como investigación positivista, mediante el empleo de procedimientos estandarizados para la recogida y el análisis de datos orientados a la toma de decisiones.

Durante los años 70, la evaluación centra su objeto y atención en los programas formativos, existiendo una gran demanda y voluntad de rendir cuentas sobre el mérito y efectividad de los programas educativos *accountability*; y realizándose un gran esfuerzo por garantizar la exactitud de las medidas que se toman, mediante el establecimiento de unas claras pautas metodológicas, que como las ofrecidas por Campbell y Stanley (1982), siguen orientando en cierto sentido, la investigación evaluativa actual.

Incluso durante este periodo, se produce una profesionalización de la evaluación, y se elaboran normas para la evaluación de programas y materiales (*Joint Committee on Standards for Educational Evaluation*, 1981/1998).

Este desarrollo culmina en EEUU, Canadá y algunos países europeos durante los años ochenta, cuando comienza a producirse un asentamiento científico y la aceptación social de la evaluación de programas educativos. Y aunque se cuestiona la viabilidad de los estudios experimentales en el ámbito de las Ciencias Sociales, se abren nuevas perspectivas que suponen la incorporación de nuevos modelos evaluativos y de nuevas estrategias metodológicas, tanto cuantitativas (diseños correlacionales, series, técnicas de clasificación,...) como cualitativas.

Algunas de estas múltiples concepciones, propuestas desde una pluralidad de perspectivas, han sido en ocasiones consideradas exageradamente como modelos (Alvira, 1991), sin tener el rigor y validez propio de éstos. Siendo éste uno de los principales *handicaps* en el desarrollo de la disciplina como campo científico.

Hoy en día, nadie duda de lo complejo del objeto de estudio de las Ciencias Sociales, de las diversidades metodológicas de su investigación y los problemas que plantea su evaluación. Y tampoco, que esta situación encuentra algunas dificultades al no existir una concepción paradigmática lo suficientemente fuerte y unitaria que oriente la teoría y sustente la práctica investigativa y evaluativa de forma diferencial a otras. Las posibilidades de la evaluación de la investigación como proceso de mejora, se ven empañadas por una escasa mentalidad y concienciación evaluativas.

A escala internacional, la situación actual se define por la utilización de múltiples modelos que se aplican a distintos contenidos (necesidades, diagnóstico, productos o impacto), la multidisciplinaridad (análisis etnográfico y estadístico), y

la multidimensionalidad en objetivos, en perspectivas y niveles de valoración (validez, credibilidad, costes, rentabilidad, aplicabilidad). En síntesis, múltiples metodologías, múltiples impactos, múltiples formatos de informa,... Es una multiplicidad de lo múltiple (Colás y Rebollo, 1993, citando a Corner).

Una consecuencia de este panorama es sin duda la dificultad de conceptualizar la evaluación de programa como campo, encontrándose múltiples definiciones en la que cada autor destaca lo que más le interesa. Sin embargo, prueba de la importancia que se le comienza a ofrecer a esta tarea evaluativa es que en algunos países la evaluación de programas está institucionalizada, no sólo como ajuste de las intervenciones públicas a unos estándares o criterios y a un gasto público, sino como proceso de mejora y perfeccionamiento.

Algunos de estos problemas de la evaluación educativa en general y de la evaluación de programas en concreto, lo expresan Worthen *et. al.* (1997), en los siguientes términos:

- Los modelos de evaluación fracasan como modelos científicos. Para Cook y Reichardt (1986) la relación entre paradigmas y métodos es instrumental, siendo la metodología una simple colección de técnicas y procedimientos, no siendo axiomáticos ni deductivos y con limitaciones en su generalización.
- Es necesario imprimir una dinámica evaluativa como proceso constructivo formativo/sumativo, utilizando metodologías complementarias, diversidad de instrumentos y técnicas, que permitan una evaluación válida y eficaz con alto grado de responsabilidad social. Su valor reside en ayudarnos a pensar, ser fuente de nuevas ideas o técnicas y servir de esquema mental de cuestiones a considerar o que no debemos olvidar.
- Se hacen necesarias fuentes de triangulación en diversos aspectos del proceso: datos, investigador, teoría y metodología (Denzin, 1989). Existiendo dos tendencias actuales, una hacia la complementariedad metodológica, pero

primando las metodologías cualitativas, y otra el "amateurismo" al considerar que cualquier docente puede investigar, aún sin preparación específica.

Y desde la perspectiva de la práctica evaluativa, los problemas que habrá que cuidar en la evaluación de programas se refieren sobre todo a la tendencia del programa o centro a cambiar mientras se está siendo evaluado, las relaciones entre evaluadores/evaluados y el hecho de que el sistema socio político en el que se inserta el programa determine éste incluso en cuestiones que no debería.

Por tanto, en general, la investigación evaluativa sigue adoleciendo de las limitaciones de las disciplinas incipientes o en ciernes de reestructuración. Sólo una enorme seriedad en las fundamentaciones teóricas, mayor rigor en los procesos metodológicos, aumento de los canales y estilos en la comunicación harán generar la suficiente consistencia científica como para ofrecer mayor credibilidad y mejorar las posibilidades de difusión (Slavin, 1997; Klein, 1997 y Pogrow, 1998).

1.6. LA INVESTIGACIÓN SOBRE EVALUACIÓN DE PROGRAMAS EN ESPAÑA.

1.6.1. El contexto diacrónico. Una perspectiva histórica.

La evolución de la investigación educativa en nuestro país guarda semejanzas con los desarrollos de ésta en el ámbito internacional, pero con ciertas peculiaridades específicas.

Prueba de la importancia ofrecida en ésta a la evaluación y por extensión a la evaluación de programas, puede encontrarse en la relevancia del marco legal desarrollado y dedicado en ciertos aspectos a su orientación.

A partir de 1970 con la Ley General de Educación, la evaluación recibe una atención creciente, no sólo como medio a través del cual mejorar el rendimiento de los estudiantes sino también todo el proceso educativo, y reflejar su calidad.

Aunque en ella se habla explícitamente de evaluación no sólo para valorar el rendimiento de los alumnos sino también como proceso educativo de mejorar, y concretamente en la O.M. 16/11/70⁶ algunos procesos de esta evaluación se refieren a evaluación de programas, son muy escasos los estudios durante este periodo.

Con esta Orden Ministerial de 16/11/1970 se regulan los principios que deben regir los procesos de evaluación e indirectamente se refiere a la Evaluación de Programas. En su apartado de principios y normas generales ya se define la evaluación como una actividad sistemática, integrada en un proceso educativo, cuya finalidad es el mejoramiento del mismo mediante un conocimiento lo más exacto posible del alumno en todos sus aspectos de personalidad y una información ajustada sobre el proceso educativo y los factores personales y ambientales que en éste inciden.

Algunas de las finalidades de la evaluación allí previstas son: Valorar los métodos y procedimientos empleados, determinar el contenido de los programas seleccionándolos por su valor formativo, determinar en que medida se alcanzan los objetivos previstos en la programación educativa y contrastar su validez.

Mientras en EEUU parece perderse cierto interés por la evaluación de programas durante los años 80, sobre todo por una restricción de los fondos dedicada a esta tarea investigativa, en otros países comienza a surgir los modelos de intervención y la preocupación de evaluación de programas.

⁶ O.M.: Orden Ministerial.

En España comienza a cobrar auge, a existir preocupación por el tema y a cobrar significación esta línea de investigación educativa (Garanto, 1989), con un énfasis creciente en que las evaluaciones sean útiles. Produciéndose un rápido pero descompensado crecimiento.

Los Reales Decretos 69/1981 y 710/1982 ordenan la E.G.B. en tres ciclos a efectos de programación y evaluación de los alumnos y se insiste de nuevo en la necesidad de una evaluación no limitada exclusivamente a la comprobación de los resultados obtenidos al término de unos periodos más o menos amplios, sino que centrada en el grado de cumplimiento de los objetivos propuestos, durante el desarrollo del proceso educativo, con el fin de hacer posible una reordenación del mismo.

Durante esta década se producen grandes cambios sociales que requieren cambios educativos en el propio sistema, desde los procedimientos metodológicos hasta nuevas demandas al proceso evaluativo.

En 1987, con el proyecto de Reforma de la enseñanza y la implementación de la LOGSE, los poderes públicos comienzan a prestar una atención mayor a la evaluación del sistema educativo, como uno de los factores favorecedores de la calidad y mejora de la enseñanza. La evaluación sigue considerándose en los mismos términos, aunque se introduce la evaluación de todas las personas involucradas en la acción educativa.

Este marco legal es en esos momentos el mejor marco para fomentar y favorecer la Evaluación Educativa, al quedar especificado en algunos de sus artículos, como el 55, la importancia y funcionalidad que se le concede:

“Los poderes públicos prestarán una atención prioritaria al conjunto de factores que favorecen la calidad y mejora de la enseñanza, entre ellos la evaluación del sistema educativo”. (Art. 55)

De manera mucho más explícita y ampliando el artículo anterior en cuanto a los ámbitos evaluativos:

"La evaluación del sistema educativo se orientará a la permanente adecuación del mismo a las demandas sociales y a las necesidades educativas y se aplicará sobre los alumnos, el profesorado, los centros, los procesos educativos y la propia administración".

(Art. 62)

Incluso en el art. 28 se plantea la creación de un centro específico para la evaluación general del sistema educativo, el I.N.C.E.⁷.

De esta forma, el ámbito de la evaluación se desarrolla desde los procesos de enseñanza / aprendizaje a todo lo que se refiera al sistema educativo en su totalidad y a los agentes que intervienen en él. Se produce un gran auge de la evaluación, de proyectos experimentales, se convierte en un instrumento imprescindible para implementar la reforma educativa experimental, 1981/1989 empiezan a aparecer trabajos en el campo de la salud y de los servicios sociales.

Pese a este apoyo de la Administración, patente desde la normativa legal establecida, no fue la comunidad científica el principal canal de comunicación y debate, siendo en los 90 cuando se produce un periodo de expansión del modelo de intervención por programas en el ámbito educativo - orientador y de servicios sociales, aumentando el número de reuniones y *congresos* dedicados a la evaluación (AEDES⁸, AIDIPE⁹ ...).

⁷ I.N.C.E.: Instituto Nacional de Calidad y Evaluación.

⁸ A.D.E.S.: Asociación Española de Desarrollo de la Educación Especial.

⁹ A.I.D.I.P.E.: Asociación Interuniversitaria De Investigación Pedagógica Experimental.

A finales del 95, con la aprobación de la L.O.P.E.G.¹⁰ además de ampliar los ámbitos de la evaluación se plantea como un proceso con carácter continuo, implantando procesos amplios para la evaluación de Centros Públicos (Orden 29 de febrero de 1996). Encontrándose en un momento trascendental de clara preocupación por la calidad educativa de las instituciones y del sistema educativo, reflejado en el propio marco legal, en el que la evaluación de programas es una exigencia ineludible para la búsqueda progresiva y permanente de una mejor educación.

La situación actual de la evaluación, que ya denotaron Tejedor, García-Varcárcel y Rodríguez (1994) y confirma este mismo autor (Tejedor, 2000), podría calificarse de plural y compleja, en la que las nuevas perspectivas no sustituyen, complementan a las anteriores. Se ha pasado de la postura tradicional de la evaluación como el empleo de métodos científicos para medir los logros educativos e iluminar la toma de decisiones; a la situación actual, en la que la evaluación es entendida como el conocimiento del mérito o valor de un programa para tomar decisiones ajustadas a las necesidades de los usuarios y/o responsables.

1.6.2. El contexto sincrónico. La situación actual.

En España la fase de evaluación de programas ha recibido paradójicamente mayor atención que las de diseño e implementación, por la necesidad de demostrar, por parte de los gestores de los programas, la efectividad de sus propuestas.

Velaz de Medrano *et. al.* (1996) señalan que hasta esa fecha el 73% de las investigaciones financiadas por el C.I.D.E. se referían a evaluación de programas y de forma específica: 17% Reformas y programas experimentales, 60% Programas curriculares de aplicación restringida y 22.8% Programas de formación del profesorado.

¹⁰ L.O.P.E.G.: Ley de Participación, Evaluación y Gobierno de los Centros Educativos.

Es por tanto un campo consolidado y con entidad propia, que produce evaluaciones externas, formativas, globales, normativas y criteriosales al mismo tiempo, fundamentalmente cuantitativas experimentales, con métodos mixtos en algunos casos. Bartolomé (1990) señala un notable aumento de los estudios evaluativos, sobre todo al amparo de la LOGSE, aunque en los últimos años se ha realizado una sola metaevaluación.

También se ha producido en España importante literatura sobre el tópico (Gelpi, 1987; Repetto, 1987; Alvira, 1991; Colás y Rebollo, 1993; Medina y Angulo, 1994; Rodríguez, 1995; Sanz Oro, 1996; Cruz, 1997; Sobrado y Ocampo, 1998), utilizando unos pocos modelos (CIPP de Stufflebean, evaluación democrática, iluminativa...). Sin embargo, se requiere ampliar el ámbito de la evaluación de programas al propio mundo académico y a la difusión científica y técnica, para que los informes de evaluación no queden en las bibliotecas y puedan ser discutidos por los técnicos y los propios interesados.

Como afirmaba Garanto (1989) estamos asistiendo a un gran desarrollo de la evaluación de programas educativos, similar al proceso ocurrido en EEUU hace algunos años, en el que concurren factores similares a los señalados por Talmage (1982): “preocupación por la mejora de la enseñanza, necesidad de controlar el gasto público y la participación de los ciudadanos en la gestión y dinámica de las estructuras formativas”.

En la actualidad y en nuestro contexto estos factores sobre la necesidad de evaluación de las intervenciones educativas quedan justificados (Colás, 2000) en aspectos concretos:

- Por la necesidad de justificar la rentabilidad política y económica de los programas educativos.

- Por los cambios sociales de valores, en el proceso de construcción europeo y autonómico, que requieren la puesta en marcha / evaluación de nuevas y múltiples intervenciones.

- Por los nuevos retos generados por el gran avance tecnológico en el último cuarto de siglo, entre los que aparecen nuevos tipos de intervenciones, metodologías y materiales educativos.

Es decir, la evaluación de programas educativos, como principio de calidad, más que en una necesidad se ha convertido en una exigencia económica, política y social. Por tanto la actividad científica en este campo necesita del establecimiento de pautas y criterios de orientación claros que permitan un firme desarrollo de la disciplina, aspectos que intentan establecerse en el siguiente capítulo de este trabajo.

CAPÍTULO II. LA EVALUACIÓN DE PROGRAMAS COMO ACTIVIDAD CIENTÍFICA.

La evaluación de programas, como técnica de investigación sobre los procesos de intervención socioeducativos, es una estrategia que contempla en su realización un potencial riquísimo para contribuir a la mejora de la calidad educativa pues tiene todos los requerimientos exigibles a la investigación educativa útil.

(Tejedor, García – Varcárcel y Rodríguez, 1994)

2.1. INTRODUCCIÓN.

La evaluación de programas o *educational evaluation*, en términos anglosajones, desde los años sesenta que emergió con fuerza, se ha convertido en un núcleo teórico-práctico de la investigación educativa, que al igual que toda ella se rige por unos principios propios del conocimiento científico.

Además de estos requerimientos generales, la investigación sobre evaluación de programas se fundamenta en una gran diversidad y complejidad de teorías y modelos, derivados de las ciencias sociales, que exigen unos adecuados conocimientos técnicos para llevarla a cabo con garantías.

El propósito en este capítulo es ofrecer el marco oportuno para realizar una delimitación técnica de la disciplina de la evaluación de programas, en aquellos rasgos que permiten caracterizar las investigaciones.

Se parte, en el primer epígrafe, del establecimiento de unos conceptos básicos y planteamientos generales, haciendo mención de los elementos diferenciales que distinguen a las investigaciones en este campo evaluativo.

Estas investigaciones, por las propias características de la disciplina y de su objeto de estudio, se enfrentan a una serie de problemas, que se describen a continuación.

En el origen de estos procesos de evaluación se origina una cuestión clave entorno a qué evaluar. Esta cuestión requiere, en el cuarto epígrafe del capítulo, el desarrollo de los contenidos objeto de estudio en las evaluaciones de programas (necesidades, contextos, procesos, productos...), que pueden plantearse en distintas fases o momentos del programa y cuya finalidad en última instancia es la toma de decisiones acerca del desarrollo de éste.

Procesos que se fundamentan en unas determinadas posiciones epistemológicas y paradigmáticas, que ofrecen modelos prácticos de evaluación que delimitan y orientan el funcionamiento empírico investigativo. Además en este capítulo se describen las principales consideraciones metodológicas propias de la investigación en evaluación de programas, como son los diseños, las fuentes de datos y las técnicas más adecuadas para la recogida de información, y los análisis y presentaciones de datos más usuales.

La intención al ofrecer todas estas informaciones es delimitar la investigación sobre evaluación de programas como disciplina científica, ofreciendo para ello aproximaciones y categorizaciones generales a distintos aspectos de ella, remitiéndose a la consulta de literatura especializada y de manuales específicos para una mayor profundización.

Para finalizar esta caracterización de la investigación en este campo disciplinar ofrecida en este capítulo, se señalan los principales avances en evaluación de programas. Haciendo mención especial a los estudios metaevaluativos y de síntesis.

2.2. CONCEPTOS BÁSICOS, PLANTEAMIENTOS GENERALES Y ELEMENTOS DIFERENCIALES EN LA INVESTIGACIÓN SOBRE EVALUACIÓN DE PROGRAMAS.

La educación, como actividad humana sistemática, se orienta al perfeccionamiento y mejora de las personas, por medio de acciones intencionadas de los educadores generalmente concretadas en planes o programas.

Como en cualquier ámbito, el ser humano ha propuesto el conocimiento del ámbito educativo. Y aunque el saber pedagógico ha partido de planteamientos y conocimientos cercanos a procesos casi artísticos, pronto se observó la necesidad del establecimiento de métodos rigurosos, que permitieran respuestas válidas a los múltiples problemas y cuestiones que se plantean en este campo con rigor científico.

En la actualidad la educación puede entenderse como tecnología, en el sentido ofrecido por Bunge (1981), aplicada a la mejora de la vida de las personas; aunque en un sentido amplio como planteaba Álvarez Rojo (1994). Generando conocimientos a los que se le asignan finalidades para su comprensión profunda, explicación y transformación de la realidad natural o social.

Así, la investigación evaluativa ha sido uno de los campos más desarrollados dentro de las Ciencias Sociales. En concreto, la evaluación de programas reclama procesos de análisis y reflexión que nos ayuden a precisar su alcance y significado en un contexto general de investigación científica (De Miguel, 1999).

La delimitación de las intervenciones educativas mediante el empleo de programas hay que entenderla desde la lógica que considera a éstos planes sistemáticos diseñados por el educador como medio al servicio del desarrollo de las metas educativas. No debiendo faltar en ellos (Pérez Juste, 2000): Unas metas u objetivos, acomodados a las características de los destinatarios; un establecimiento claro y exhaustivo en cuanto a estrategias, medios y recursos suficientes y adecuados para llevarlo a cabo; y la consideración de un sistema que permita su valoración.

Este concepto de actuación mediante la elaboración de programas, queda justificada en cualquier acción de tipo educativa para autores como Barr y Cuyjet (1991) o Álvarez Rojo, *et al.* (1994). Y se puede especificar en definiciones como las ofrecidas por diversos autores para este término:

Actividad educativa continua diseñada para producir cambios esperados en la conducta de los sujetos que se exponen a ella

(Astin y Panos, 1971)

Conjunto especificado de acciones humanas y recursos materiales diseñados e implementados organizadamente en una determinada realidad social con el propósito de resolver algún problema.

(Fernández Ballesteros, 1992)

Sistemáticos esfuerzos realizados para lograr objetivos planificados con el fin de mejorar la salud, el conocimiento, las actitudes y la práctica.

(Fink, 1993)

Es decir los procesos educativos quedarían establecidos en acciones sistemáticas y planificadas, especificando acciones y recursos materiales, para cambiar su conducta y/o actitud, mejorar su salud o su conocimiento y en definitiva solucionar sus problemas.

Este mismo razonamiento aboca a la evaluación de tales intervenciones mediante procesos evaluativos, que de forma general se han venido justificando a partir de dos funciones: la rendición de cuentas (*accountability*) y la mejora del programa (*improvement*) y los procesos. Pero además, resaltando el carácter investigativo de esta evaluación, se puede generar una comprensión más profunda de este objeto de conocimiento a la que no se debe renunciar, ya que permite una mayor comprensión y explicación de los fenómenos sociales (Chelimsky, 1997).

Estos procesos de evaluación no deben entenderse como una medida puntual, sino como instrumento de valoración crítica y formativa que puede ayudar a mejorar el proceso de enseñanza (Carrera, 1987; Bartolomé, 1990).

Así, la valoración de los programas tiene como finalidad específica la facilitación de la toma de decisiones orientada a la consecución de la calidad para la mejora (Wroeijerstijn y Acherman, 1990). Entendiéndose en este sentido, la evaluación como un proceso de búsqueda de conocimientos respecto a las estrategias y factores que facilitan o inhiben los procesos de mejora, comprometiéndose como mecanismo fundamental del cambio social.

Por ello, debe orientar sus intervenciones a promover el desarrollo de las políticas más adecuadas o conseguir programas más eficaces, mediante la participación activa y democrática de los protagonistas en los propios procesos evaluativos. Siendo éste un compromiso del evaluador con la educación cívica, el desarrollo humano y la emancipación social (Smith, 1998).

Fernández Ballesteros (1996) realiza una conceptualización entorno a otras cuestiones básicas que rodean la evaluación de programas entorno a qué es, cómo se realiza, para qué se hace y qué se evalúa en los procesos.

La primera cuestión sobre a qué se refiere la evaluación de programas se plantea desde la consideración que se ha dado a la investigación evaluativa desde

su desarrollo histórico, que parte desde la obra de autores como Tyler (1957), Scriven (1967) o Levin (1975) que asociaban la evaluación a la estimación del mérito de un objeto y específicamente en la evaluación de los programas a la consecución y cumplimiento de las metas planteadas y al examen de sus resultados. Hasta las más relacionadas con la toma de decisiones para la mejora de las personas y contexto donde se aplican. Actuaciones que el *Joint Committe* trata de sistematizar en 1981.

Evaluar un programa consiste, por tanto, en recoger información sistemática sobre las operaciones implicadas en ese conjunto de acciones encaminadas a unos objetivos (Hernández, 1990). Dicha información debe servir para establecer un juicio acerca de la bondad, mérito o valor de dicho programa (Fernández Ballesteros, 1983), que permita a los encargados tomar decisiones acerca de lo que dio resultado y lo que no sirvió para nada, qué es lo que debería adaptarse en todo sistema o de lo contrario modificarse. Cuestiones a las que sólo puede responderse fundamentadamente y presentando pruebas mediante la evaluación (Weiss, 1975).

En este mismo sentido ha evolucionado la funcionalidad de la evaluación, desde esta justificación de las decisiones evaluativas ofrecida por Weiss, a planteamientos para medir la implementación y los resultados (Rutman, 1980), para la toma de decisiones sobre intervención (Cronbach, 1982) o para la justificación de las actuaciones y gastos públicos (Fernández Ballesteros, 1984).

Este mismo autor, delimita la funcionalidad de estas evaluaciones en diversos campos además de la contabilidad pública o base para decisiones presupuestarias y el de la justificación de decisiones, en la realización de ajustes del propio programa y la constatación de teorías que subyacen en los propios modelos evaluativos.

Los modelos de ingeniería en la utilización de la evaluación (Alback, 1988 y Fischer, 1990) utilizan la evaluación de programas como procesos en los

que los usuarios identifican problemas sociales, establecen objetivos y formula preguntas, para juzgar los resultados y emplearlos en la toma de decisiones.

Los modelos de iluminación o uso conceptual de la evaluación incorporan los resultados gradualmente al marco del usuario influyendo más en su pensamiento que en la acción directa (Weiss, 1975).

Incluso podrían delimitarse empleos de la evaluación como mera legitimación del poder y medio de racionalización de la toma de decisiones, que en un sentido poco lícito, permitiría su uso táctico en la eliminación de responsabilidades.

Esta necesidad de la evaluación de los planes y programas, queda incluso especificada en la normativa legal, ya que en el artículo 2.3 de la LOGSE habla de la evaluación de nuestro sistema educativo, la valoración de los procesos educativos y de la propia enseñanza y aprendizaje.

Es éste un tipo de investigación, al que Briones (1991) delimita su objeto, en el análisis de la estructura, del funcionamiento y de los resultados de un programa con el fin de proporcionar información de la cual se deriven criterios útiles para la toma de decisiones en relación con su administración y desarrollo. Para Cook y Shadish (1986), la evaluación de programas es un modelo racional de solución de problemas de carácter psicológico, social o educativo, que requiere que:

- Los problemas se definan con claridad y a partir de los que....
- Se genere una amplia variedad de soluciones potenciales...
- Algunas de las cuales se implementan y...
- Finalmente se evalúan.
- El conocimiento de las soluciones positivas se presenta a los responsables políticos quienes...

- Lo utilizarán para tomar las decisiones políticas que correspondan.

Pero como se puede constatar en la descripción de estos procesos, suponen situaciones complejas en las que se relacionan elementos de diversas naturalezas, lo que implica una serie de dificultades, descritas por Pérez Juste (2000).

La primera de ellas se relaciona precisamente con la complejidad de las acciones evaluativas que requiere la existencia de profesionales expertos en el campo, que no siempre se encuentran en número necesario, ni con la formación adecuada. En muchas ocasiones no son reconocidos en competencia y experiencia por los evaluados, lo que unido al recelo que levanta en personas e instituciones a ser evaluado, hace que estas acciones pierdan una enorme efectividad.

En cierto sentido, las relaciones necesarias entre los distintos implicados en los procesos evaluativos son las que determinan la mayoría de sus dificultades. Así la necesaria participación de todos ellos o la complementariedad metodológica, conducen a un problema sin resolver sobre quienes deben tomar las decisiones definitivas que rigen estos procesos, en el que no siempre es posible llegar a consensos en sesiones evaluativas y que éstos sean aceptados por las instituciones.

Estos problemas también se refieren a la relación entre los expertos contratados para la realización de evaluaciones y las instituciones que las encargan, pues aunque a éstos se le suponga honradez o probidad legal y ética en sus acciones, la desigualdad de condiciones con las grandes instituciones y la falta de fuertes asociaciones profesionales, hacen que las evaluaciones asuman las condiciones de quienes las contratan. Incluso se pueden constatar influencias de carácter eminentemente político.

Otra dificultad se relaciona con la consecución de evaluaciones factibles, en el sentido ofrecido a este estándar por el *Joint Committe*, que relaciona la adquisición de informaciones relevantes y los recursos empleados para conseguirlas, que no siempre son suficientes y adecuados.

Y por último, el problema típico de la investigación social referido a la comparabilidad de los datos, en momentos en los que se opta por la descentralización curricular y la autonomía de centros y profesores, debe solucionarse no por la comparación en estándares nacionales o externos, sino por la comparación diacrónica con valores propios y específicos, cuya finalidad sea siempre las opciones de mejora y calidad.

Este aspecto, es sólo parte de las dificultades derivadas de la diversidad de teorías, modelos y metodologías, que ha embarcado a muchos autores en polémicas sobre su uso (Guba, 1978), solo superable por un multiplismo necesario en acciones investigativas en situaciones complejas como son las que se encuentran en los entornos sociales.

Todas estas peculiaridades, del objeto de estudio y los procesos evaluativos en E.P., hacen que la investigación evaluativa en este campo posea unas peculiaridades específicas y elementos diferenciales con respecto al resto de la investigación educativa.

Una de estas características que distingue la investigación evaluativa sobre programas es la de su realización en un marco de acción. Es decir, además de la investigación, su objeto el programa, está ocurriendo simultáneamente. Para Weiss (1975) hay que tener claro que la investigación es secundaria, no supeditarla al desarrollo del propio programa. Es por tanto un campo en el que domina la investigación aplicada.

Este objeto de conocimiento, aunque se plantee en una única forma como programa, posee naturalezas diversas, puesto que el contenido de éstos puede

abarcar conceptos propios de cualquiera de las disciplinas pedagógicas. Esta diversidad plantea situaciones en las que aún respetando las condiciones generales evaluativas, se plantee la necesidad de adaptación a cada disciplina concreta en aspectos como sus conceptos básicos, temáticas, métodos e incluso vocabulario específico.

Las investigaciones sobre E.P. quedan delimitadas en modelos teóricos, que fundamentados en los paradigmas usuales de la investigación educativa, establecen unas condiciones o criterios básicos de acción que orientan los procesos evaluativos de forma global ofreciendo coherencia de las acciones evaluativas.

Desde algunos de estos modelos básicos, en los que es usual la eponimidad, como son el modelo de Tyler o el de Scriven entre otros, se realizan adaptaciones casi tan diversas y específicas a situaciones concretas como autores transitorios caracterizan las realizaciones en este campo. Aplicaciones empíricas que por la dificultad que contemplan y la carencia de experiencia no siempre poseen las garantías suficientes, lo que en definitiva supone un detrimento de la credibilidad del propio campo disciplinar.

Con cierta lógica, los contenidos de las evaluaciones guardan coherencia con el área o disciplina en la que se desarrolla el programa evaluado y los contenidos desarrollados en la implementación de éste. Es por ello que siendo éste un campo disciplinar en el que convergen numerosas áreas de conocimiento, los contenidos evaluativos puedan ser muy diversos en distintas evaluaciones e incluso posean cierta variedad en la evaluación de un mismo programa por ser abordado éste desde perspectivas dispares pero complementarias.

Además esos contenidos quedan relacionados con los momentos del proceso de implementación del programa a los que se refieran, pudiendo referirse al contexto a necesidades de implementación, diseño o planificación, evaluación del proceso y del producto.

En cuanto a los procesos empíricos concretos se suelen emplear una extensa variedad de diseños y metodologías de investigación usuales en cualquier otra área de la investigación educativa. En la que, dado lo complejo de los contextos en que se desarrollan, las dificultades de aleatorización y manipulación/control de variables, abundan los diseños cusiexperimentales con connotaciones descriptivas orientadas a la descripción de procesos.

Estas características metodológicas determinan el uso de técnicas de recogida de información y análisis e interpretación de datos típicos de los estudios descriptivos. Siendo el cuestionario un instrumento casi omnipresente en este tipo de trabajos, hasta el punto que puede considerarse de cierto abuso (Fernández Cano, 2002a) por el reduccionismo empírico que supone.

Las peculiaridades diferenciales de las investigaciones sobre E.P. llegan hasta la elaboración y uso de los informes de investigación. Ya que si la finalidad de la evaluación del programa es aportar información válida para valorar una intervención educativa, el uso y explotación de estas informaciones debe quedar al alcance de las audiencias a las que se dirige (Worthen y Sanders, 1987; Santos, 1990 o Alkin y Ellett, 1990). Siendo por tanto determinante quién evalúa e programa, para qué, cómo lo hace y sobre todo para quién.

En la mayoría de ocasiones, la finalidad última de las investigaciones educativas es revertir los conocimientos adquiridos en el corpus teórico de conocimiento de la disciplina tratada, para a partir de ahí promover acciones prácticas para el cambio y mejora en la educación. Sin embargo, en la investigación sobre E.P. la aplicación es más directa es sobre la propia práctica educativa, pues en primer lugar se plantea la mejora del programa desarrollado, para con posterioridad realizar generalizaciones e inferencias que pueden alcanzar los principios teóricos desde los que se derivan las diversas acciones implementadas en el programa o desarrolladas en las acciones evaluativas.

2.3. PROBLEMAS CENTRALES EN LA EVALUACIÓN DE PROGRAMAS.

El campo de la evaluación de programas supone la puesta en marcha de situaciones complejas en las que se relacionan elementos de diversas naturalezas, adoleciendo estos procesos de una serie de problemas centrales paralelos al propio desarrollo de la investigación educativa en general y de la investigación sobre evaluación de programas en particular. Estas dificultades, descritas en algunos aspectos por Pérez Juste (2000), son de diversa índole y pueden agruparse en los siguientes aspectos:

El primero de ellos se deriva del propio carácter semántico de la evaluación, que como proceso de emisión de juicios y valores se basa en la asignación de significados previos: de su carácter, proceso y criterios valorativos. Así dependiendo de lo que se entienda por evaluación, cómo debe llevarse a cabo, con qué finalidad, qué criterios se emplean para ello, los procesos adoptaran y requerirán una dinámica determinada.

Muestra de las características semánticas asignadas al campo de la E.P. es la diversidad de definiciones adoptadas por cada uno de los autores, desde qué se entiende por programa o evaluación, hasta cómo debe realizarse ésta. Significados planteados de forma acorde a sus propuestas teóricas y adaptadas a los desarrollos evaluativos que plantean.

Este aspecto, es sólo parte de las dificultades derivadas de la diversidad de teorías, modelos y metodologías, que ha embarcado a muchos autores en polémicas sobre su uso (Guba, 1987), solo superable por un multiplismo necesario en acciones investigativas en situaciones complejas como son las que se encuentran en los entornos sociales.

Existe una clara tendencia a valorar la investigación aplicada frente a la básica o fundamental, lo que a la larga limita las posibilidades de impactar transformando la práctica y política educativa (Fernández Cano, 1997).

El propio objeto de conocimiento de la investigación sobre evaluación de programas puede ser, como ente social, además de enormemente complejo en continua evolución. Esto dificulta los procesos de investigación pues requiere combinar la fortaleza y credibilidad de la metodología científica con la flexibilidad de procesos con la suficiente capacidad adaptativa. Así, las nuevas necesidades sociales demandan nuevos modelos de intervención y por tanto nuevos modelos evaluativos (Bartolomé *et. al.*, 2000).

Otro de los principales problemas para Wholey (1983 y 1987) es que como en la mayoría de prácticas evaluativas, existe una resistencia por parte de los individuos e instituciones a participar en procesos evaluativos y a desarrollar procesos de cambio mientras se está siendo evaluado.

Las dificultades metodológicas en los procesos de evaluación de programas son numerosas, comenzando por la falta de claridad en el establecimiento de asunciones teóricas y prioridades de la evaluación, y de lógica entre las acciones y los resultados esperados. Las, en ocasiones, difíciles relaciones entre los evaluadores y el personal del programa o centro. Las dificultades de plantear procesos de revisión adecuados o la escasa utilización de los resultados de las evaluaciones.

Los procesos que con una naturaleza eminentemente empírica se asocian a hechos distintos al de la propia investigación, como es la existencia de un programa, deben establecerse de forma concurrente a éste. La evaluación no debe entenderse como un proceso paralelo o ajeno al desarrollo del programa, ni superarse a la implementación de éste, si no que forma parte de él en todas sus fases requiriendo un constante esfuerzo de crítica constructiva.

El hecho de que el programa se encuentre insertado en un sistema psicosocial más amplio y de que la naturaleza de dicho sistema determine de alguna manera la evaluación, hace que este proceso pueda encontrar un mayor número de dificultades.

Estos procesos evaluativos en contextos sociales se encuentran delimitados por la toma de numerosas decisiones evaluativas acerca de quién establece el valor de un programa educativo, con qué criterios de esfuerzo, efectividad o impacto se realiza, o cómo se emplea la información obtenida, lo que puede acarrear numerosos problemas de índole ético y deontológico (Suchman, 1967; Talmage, 1982).

Estas acciones complejas para el desarrollo de la evaluación también encuentran en su sentido más material dificultades de tipo logístico. Puesto que aunque en el mejor de los casos se cuenten con los suficientes recursos humanos y materiales para realizar un conveniente proceso evaluativo, en ocasiones éste queda tan sólo sujeto a la voluntad y decisión de la autoridad política o institucional.

Pese a lo complejo de los procesos evaluativos, la producción científica sobre evaluación de programas descansa en manos de demasiados autores ocasionales con escasa especialización en el campo. Así la existencia de profesionales expertos en el campo, no siempre se encuentran en número necesario, ni con la formación adecuada.

Esta falta de experiencia hace que en muchas ocasiones los profesionales de la evaluación no son reconocidos en competencia y experiencia por los evaluados, lo que unido al recelo que levanta en personas e instituciones a ser examinado, hace que estas acciones pierdan una enorme efectividad.

En cierto sentido, las relaciones necesarias entre los distintos implicados en los procesos evaluativos son las que determinan la mayoría de sus dificultades. Así la necesaria participación de todos ellos o la complementariedad metodológica, conducen a un problema sin resolver sobre quienes deben tomar las decisiones definitivas que rigen estos procesos, en el que no siempre es posible llegar a consensos en sesiones evaluativas y que éstos sean aceptados por las instituciones.

Estos problemas también se refieren a la relación entre los expertos contratados para la realización de evaluaciones y las instituciones que las encargan, pues aunque a éstos se le suponga honradez o probidad legal y ética en sus acciones, la desigualdad de condiciones con las grandes instituciones y la falta de fuertes asociaciones profesionales, hacen que las evaluaciones asuman las condiciones de quienes las contratan. Incluso se pueden constatar influencias de carácter eminentemente político.

Así, la investigación queda en ocasiones supeditada a agendas o líneas de investigación prioritarias establecidas centralmente por autoridades autonómicas, nacionales o internacionales. Siendo éste un problema típico de la investigación social referido a la comparabilidad de los datos, en momentos en los que se opta por la descentralización curricular y la autonomía de centros y profesores, debe solucionarse no por la comparación en estándares nacionales o externos, sino por la comparación diacrónica con valores propios y específicos, cuya finalidad sea siempre las opciones de mejora y calidad.

Otra dificultad se relaciona con la consecución de evaluaciones factibles, en el sentido ofrecido a este estándar por el *Joint Committe*, que relaciona la adquisición de informaciones relevantes y los recursos empleados para conseguirlas, que no siempre son suficientes y adecuados.

Además existen dificultades de difusión y utilización de la investigación que podrían calificarse casi de endémicas, limitando o incluso haciendo estéril su

uso. Sobre todo debido a la escasa difusión e impacto internacional de las fuentes de publicación españolas, que provocan por la dependencia de ciertas fuentes de gran relevancia un colonialismo investigativo.

2.4. FASES Y CONTENIDOS DEL PROCESO EVALUATIVO.

La evaluación de los programas puede plantearse en muy diversos momentos: durante su planificación, en el desarrollo del mismo o en sus logros y resultados. Dependiendo además del momento del contenido asignado a cada uno de ellos, una gran diversidad de propuestas.

Los procesos evaluativos deben partir para Pérez Juste (2000) de la consideración de aspectos como lo educativo como referente, el tamaño y complejidad del programa, el papel que adoptará el evaluador, cual serán las unidades de análisis y qué informaciones se recogerán. Y se concretan en diversos momentos:

1. El **primer momento** de evaluación abarca **al programa en su globalidad**, aportando opciones preventivas de mejora y optimización, incluso antes de su puesta en marcha.

Su finalidad es establecer formativamente la calidad técnica del programa, recogiendo datos sobre su fundamentación, formulación y relación con las demandas e intereses de los destinatarios.

Emplea metodologías basadas en el análisis de contenido, estudios prospectivos, *técnica Delphi*, juicios de expertos y el empleo de pruebas diversas de evaluación inicial. Y criterios de calidad, pertinencia y realismo de objetivos y su adecuación a los destinatarios. Suficiencia de apoyos y compromiso e implicación de los agentes relacionados.

Y sobre la calidad técnica, tanto del proceso de evaluación, como del propio programa, en aspectos relacionados con la coherencia y congruencia de sus componentes y sus facilidades para ser evaluado; para tomar decisiones eminentemente formativas.

Este momento evaluativo, previo a la puesta en marcha del mismo programa, está orientado para Colás y Rebollo (1993) a dos contenidos prioritarios y diferenciados: La evaluación del contexto y de necesidades del programa y la evaluación del diseño y planificación del programa.

La *evaluación del contexto y de necesidades* tiene por objeto servir de guía para la selección de problemas y asignación de prioridades. Abarca cuestiones sobre las características específicas del entorno donde se desarrollará el programa; quiénes son los participantes; qué condiciones exigen su realización y qué recursos materiales, temporales y humanos son necesarios; qué necesidades pretende cubrir o qué expectativas tienen los receptores.

Para Stufflebean y Shinkfield (1987) esta es una etapa de diagnóstico de necesidades del contexto institucional y poblacional.

La *evaluación del diseño y planificación del programa* se ciñe a identificar y valorarla planificación de estrategias y procedimientos que el programa plantea para su desarrollo, atendiendo a cuestiones como la correcta definición de objetivos, si la secuencia de actividades está expresada claramente para su puesta en marcha y su temporalidad está ajustada, o si la planificación general del programa se ajusta al contexto donde se desarrollará.

La omisión de algunos de estos aspectos (conceptos, organización, materiales, etc.) en la evaluación puede. Como planteaba Álvarez Rojo (1991), incluso falsear el resto de la evaluación.

Para la valoración y juicio de estos objetos Fernández Ballesteros (1996) propone juicios relativos a su pertinencia (si responden a necesidades concretas bien establecidas) y suficiencia (si los medios que se proponen son los suficientes).

2. Como **segundo momento** hace referencia a las medidas evaluativas de la **implementación del programa**, para la toma formativa de decisiones de mejora.

Se recogen informaciones sobre el desarrollo del programa, sus efectos intermedios y los no deseados; mediante instrumentos como la observación, diálogos, entrevistas o pruebas intermedias, que se analizan en sesiones de grupo y debates.

La pretensión es descubrir durante el proceso los defectos de planificación y describir y juzgar las actividades que se están realizando. Es decir, los criterios de evaluación giran entorno al progreso propuesto por King, Morris y Fitz-Gibbon (1987) y cumplimiento de los planes previstos, a su eficiencia y a la satisfacción de los usuarios. Para promover medidas de ajuste.

Los contenidos de la evaluación se refieren a cuestiones como cuál es la percepción que tienen los interesados de la ejecución del programa, si se ajustan las actividades al plan previsto, cuáles son las dificultades y logros más importantes, cómo funcionan los grupos y cómo es el nivel de participación; y en definitiva, cuál es el desarrollo real del programa.

3. El **tercer momento** de evaluación posee un carácter sumativo para comprobar la **eficacia del programa y los resultados de su aplicación**.

Se basa en el análisis de la consecución de objetivos, mediante criterios de eficacia (logro de objetivos propuestos), eficiencia (logros en relación con los

medios disponibles) y efectividad (efectos beneficiosos no previstos). Satisfacción de los implicados e impacto del programa sobre el contexto en el que se aplica.

4. Suponiendo un **cuarto momento** para la **institucionalización de la evaluación** del programa, basada en procesos cíclicos de evaluación y mejora, en el que los individuos e instituciones asimilan los cambios que el programa pretendía provocar.

La finalidad de todo proceso evaluativo es valorar, interpretar y juzgar los logros de un programa. Esta valoración no sólo hay que considerarla sumativamente tras su implementación, sino que como se ha señalado con un carácter formativo, parte incluso antes de la propia elaboración del proyecto y afecta a cada una de sus fases.

Como efecto patente de esta toma de decisiones, las acciones a realizar se pueden enmarcar en tres posibles ámbitos: Sobre la valoración del programa en cada momento evaluativo puede decidirse que en líneas generales es positiva, neutra o negativa; esta valoración determina a su vez acciones relativas a la modificación de las actuaciones diseñadas o que incluso se están implementando, en cuanto a mantenerlas o modificarlas si no se consideran adecuadas o al menos se consideran mejorables; el tercer ámbito de decisión implica la continuidad o abandono del proceso evaluativo y del propio diseño, desarrollo e implementación del programa, por los juicios considerados y generados desde la consideración de su inadecuación total o parcial en aspectos humanos o materiales.

ESTOS ÁMBITOS DE DECISIÓN Y LAS POSIBLES ACCIONES DERIVADAS EN ELLOS AFECTAN AL PROCESO EVALUATIVO EN CUALQUIERA DE SUS FASES, ESTABLECIDAS ECLÉCTICAMENTE POR LAS CONSIDERACIONES DE COLÁS Y REBOLLO (1993) Y PÉREZ JUSTE (2000).

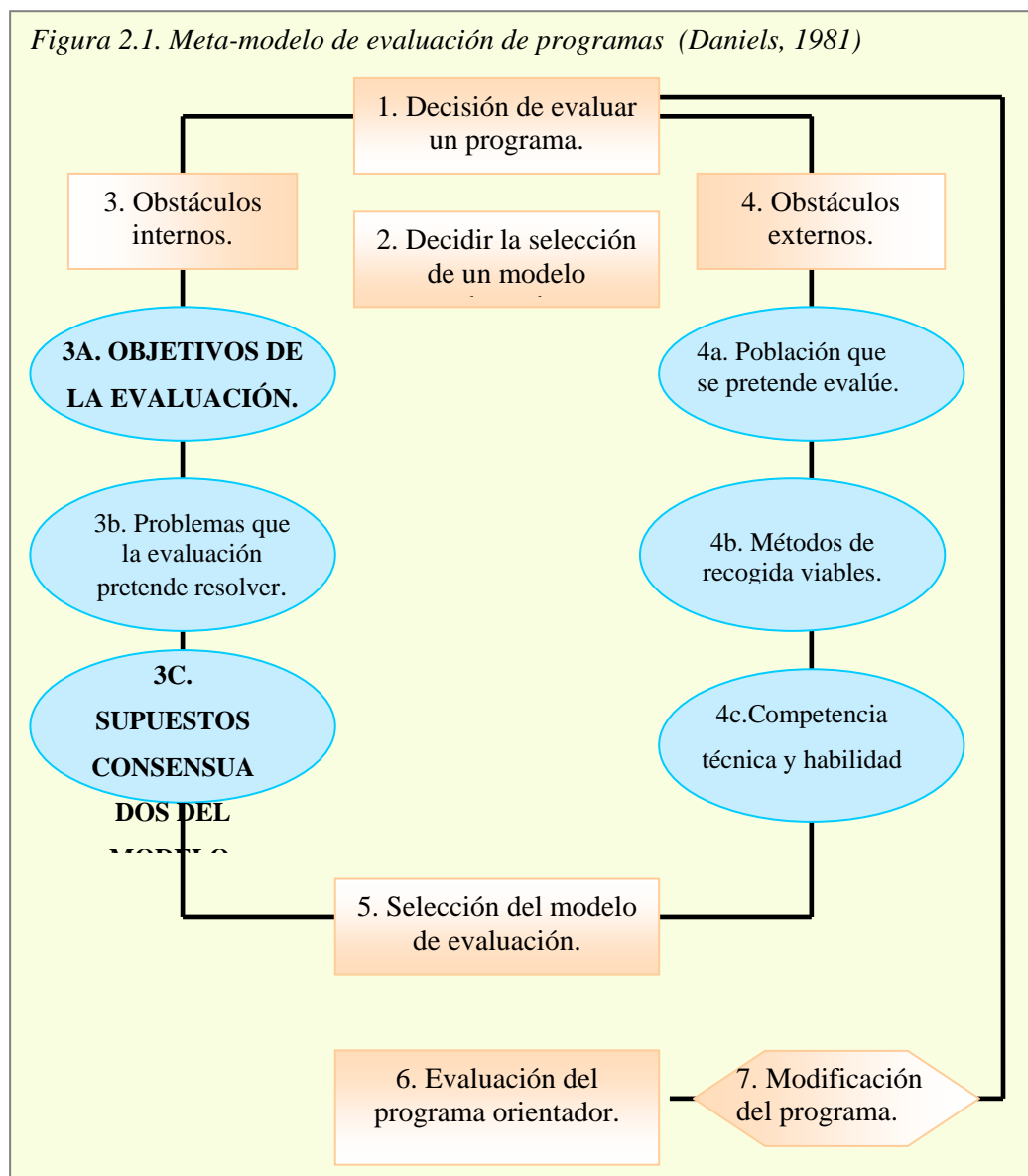
2.5. ENFOQUES TEÓRICOS Y MODELOS EN EVALUACIÓN DE PROGRAMAS.

Existe un gran intento de formalización de modelos sobre la evaluación de programas en función de múltiples factores como el objetivo de la evaluación, los tópicos y metodologías, las funciones e incluso el propio concepto de evaluación. Esta proliferación ha supuesto un cierto confusionismo para el usuario. Incluso para algunos autores (Alvira, 1991), en la mayoría de los casos no se puede hablar de modelos en sentido estricto, sino más bien de diferentes aproximaciones a unas pocas variantes básicas de la metodología investigativa.

En evaluación de programas, el paradigma evaluativo supone una conceptualización general del modo más apropiado para llevar a cabo. Es decir, las resoluciones que se tomen en el ámbito teórico determinarán la práctica evaluativa desde aspectos generales como el diseño de investigación, el modelo, contenidos y objetivos evaluativos y metodología; hasta aspectos tan concretos como el plan, secuencia, proceso recogida, análisis e interpretación de datos o el tipo de informe de evaluación.

Aunque esta selección del modelo se realice siguiendo las características específicas del programa ha desarrollar, existen pautas de carácter general que pueden ayudar a realizar tal selección, como la propuesta de Daniels, *et al.* (1981), que se muestra en la figura 2.1.

De la Orden (1985) hace un repaso de las principales tipologías de modelos de evaluación mostrando la multiplicidad de modelos existentes. De igual modo Talmage (1982) ensaya una clasificación haciendo referencia a clasificaciones previas de Ross y Cronbach (1976), Guba (1978), Stufflebeam y Webster (1980), House (1980) y Popham (1980). Utilizándose múltiples consideraciones, tanto en el ámbito teórico como práctico (finalidad, contenidos, efectos de la evaluación o enfoque evaluativo), para clasificar los diversos modelos de evaluación.



Ateniéndose tal diversidad a la lógica de la propia evolución del concepto de *evaluación* y del desarrollo histórico de la *investigación educativa*, como señala Carballo (1990).

Popham (1980) clasifica los modelos de evaluación de programas por los siguientes aspectos:

- Modelos de concreción de metas, como el *Modelo de Tyler*, caracterizado por el establecimiento de metas y objetivos definidos en términos operativos, la ordenación de los objetos en amplias clasificaciones, establecer en qué condiciones puede demostrarse la consecución de objetivos, explicar los propósitos de la evaluación a los destinatarios, recopilar datos y realizar comparaciones de los datos con los objetivos del programa.

- Modelos de enjuiciamiento que acentúan criterios intrínsecos, en los que el evaluador tiene un papel muy importante, al realizar un planteamiento técnico-profesional, pero sin conexión con el contexto.

- Modelos de enjuiciamiento que acentúan criterios extrínsecos. Como el *Modelo de Scriven*, el que se realiza una distinción entre la evaluación formativa y sumativa, demandando especialmente el interés de los objetivos. O el *modelo de evaluación respondente de Stake*.

- Modelos orientados a la toma de decisiones como el *Modelo de Cronbach*, el *Modelo CIPP de Stufflebeam* o el *Modelo CSE de Alkin* (variante CIPP).

- Modelos que se basan en la negociación como el *Modelo respondente constructivista de Guba y Lincoln*.

En ocasiones resulta difícil delimitar los diferentes paradigmas y encontrar modelos que puedan incluirse en una sola de estas categorías. Para Colás y Rebollo, (1993) los principales modelos en evaluación pueden agruparse atendiendo a criterios tales como su finalidad, su contenido, a quién se evalúa, quién toma las decisiones o qué papel tiene el evaluador en objetivistas o

positivistas, subjetivistas o interpretativos y críticos; incluyendo en cada categoría los modelos que más se ajusten. (Tabla 2.1.)

Los *modelos de evaluación objetivistas*, ya sea la basada en objetivos de Tyler, la planificación evaluativa de Cronbach, el modelo C.I.P.P. de Stufflebean o la evaluación sin referencia a objetivos de Scriven; poseen un determinismo hacia la eficacia. Propugnan evaluaciones promovidas externamente, en la que analizando con neutralidad y rigor los datos intentan explicar, predecir y controlar.

Mientras, la finalidad de los *modelos subjetivistas*, como el de evaluación respondiente de Stake, la evaluación iluminativa de Parlett y Hamilton o la evaluación democrática propuesta por Mc Donald; se diferencian sobre todo en la concepción de la realidad y en la forma de entender el conocimiento. De forma que, es la colaboración con los sujetos centros o instituciones la que permite la comprensión de las relaciones internas profundas. Mediante una concepción voluntarista y creativa se buscan visiones globales en las que el contexto en el que se desarrolla el programa juega un papel importante.

Si embargo, es en los *modelos críticos* (Parley, McKinney o Brown) donde más implicación se requiere con el objetivo principal de transformar la realidad mediante la evaluación. La evaluación se centra en el análisis de las circunstancias personales, sociales y políticas y posee una finalidad emancipativa de los destinatarios del programa.

Para algunos autores no hay modelos explícitamente vinculados a esta modalidad, aunque si hay experiencias que siguen las pautas marcadas (Martínez *et al.*, 1990; Pozo *et al.*, 1990).

Cada modelo presenta también características diferenciales en cuanto a la metodología y diseño de evaluación a emplear, descritas en la Tabla 2.2.

Las evaluaciones científicas se basan en modelos científicos clásicos en los que se intenta recoger informaciones válidas y fiables, para que con un análisis de tipo estadístico comprobar si se cumplen las hipótesis de partida.

Los modelos interpretativos son más abiertos y flexibles, su pretensión no es la generalización, sino el proporcionar un conocimiento comprensivo desde la óptica de los implicados en el desarrollo del programa evaluado.

Los procesos de tipo crítico se fundamentan en procesos dialécticos, sólo aplicables en un contexto determinado, en el que se buscan explicaciones consensuadas, válidas y prácticas con una finalidad transformadora.

En definitiva, estas concepciones teóricas que pueden servir de marco orientativo y delimitador del proceso evaluativo no deben ser excluyentes ni herméticas. Debe ser el evaluador/es los que decidan y utilicen aquellas técnicas, procedimientos y concepciones que se adecuen más al problema y contexto de evaluación planteado.

Las aproximaciones complementaristas o eclécticas pretenden mediante un conocimiento teórico-práctico generar teorías o resolver problemas. En sus prácticas evaluativas se incorporan los valores y los supuestos contextuales, así como la labor del evaluador como “factor humano”.

Tabla 2.1. Modelos para la evaluación de programas.

DENOMINACIÓN DEL MODELO		FINALIDAD	CONTENIDO	UNIDADES EVALUACION	TOMA DECISIONES	PAPEL EVALUADOR
OBJETIVISTAS	TAYLER	Prescripción	Producto	Beneficiarios del programa	Autoridad que encarga la evaluación	Externo
	CRONBACH		Proceso Producto			
	STUFFLEBEAN		C: Contexto I: Inputs, Diseño Planificación P: Proceso P: Producto			
	SCRIVEN Sin referencia a objetivos		Contexto Proceso Producto		Contrato autoridad/ evaluador	
SUBJETIVISTAS	STAKE Diseño respondente	Descripción, relato completo y holístico del programa	Proceso Producto	Sujetos y centros	Contrato evaluador/ clientes	Cooperación
	PARLET Y HAMILTON Modelo Iluminativo	Prescripción	Proceso Contexto	Centros e instituciones	Participantes Autoridad: Evaluador/ clientes	Cooperación
	McDONALD Evaluación Democrática	Transformar				
CRITICO	PARLEY COLAS McKINNEY BROWN				Todos los participantes	Implicación y compromiso

Tabla 2.2. Metodologías y diseños en los distintos modelos de evaluación de programas.

Aspectos metodológicos		TIPO DE MODELO EVALUATIVO		
		<i>OBJETIVISTAS</i>	<i>SUBJETIVISTAS</i>	<i>CRITICOS</i>
Metodología de la evaluación	Criterio evaluación	Grado de productividad. Grado de eficacia resultados/coste.	Comprensión y valoración de los procesos y los resultados del programa.	Grado de transformación de los participantes en el programa.
	Función evaluación	Valorar la productividad para conseguir los objetivos marcados.	Centrada también en la valoración de los efectos secundarios, a largo plazo, así como los procesos explícitos y ocultos.	Análisis profundo de las circunstancias personales, sociales, políticas y económicas que rodean a los individuos en las propias acciones.
	Contenidos evaluación	Inputs y resultados según esas entradas.	Procesos, con la intención de reformular las acciones educativas.	Tanto el proceso de la toma de decisiones, como el resultado de las acciones derivadas de éstas.
	Métodos evaluación	Científicos clásicos (hipotético – deductivo) Se valora la rigurosidad, capacidad de replicación y posibilidades de generalización.	Centrados en entender y valorar las situaciones particulares o singulares, no busca generalización.	Carácter cíclico y no predeterminado, se altera durante el proceso.
Diseño de la evaluación	Diseño	Estructurado, prefijado en el proyecto inicial.	Abierto, flexible y emergente.	Dialéctico, se genera mediante la reflexión y discusión crítica.
	Selección muestreo	Representativo estadísticamente.	Ajustado al tipo y cantidad de información, no prefijado.	Propio grupo que aborda la investigación (no generalizable).

Instrum.	recogida datos	Válidos y fiables.	Carácter abierto en interpretación y enfoque.	Técnicas cuantitativas, pero sobre todo cualitativas.
	Análisis datos	Estadístico.	Proceso cíclico interactivo para delimitar el problema y la hipótesis.	Proceso cíclico grupal. Alto índice de abstracción (relacionan factores personales, sociales...).
	Factores claves	Validez interna y externa.	Preocupación por la credibilidad de resultados (triangulación, réplica paso a paso).	Explicaciones validadas consensualmente.
	Finalidad	Comprobar si se cumple la hipótesis de partida.	Proporcionar conocimiento y comprensión desde la óptica de los implicados.	Dar una explicación práctica y transformadora.

Actualmente hay una clara tendencia a la búsqueda de la complementariedad de los diferentes enfoques, evitando la oposición y descalificación, para acogerse a esta posición ecléctica, que vinculada a mayores dosis de pragmatismo, entiende que la relación entre paradigmas y métodos es meramente instrumental, siendo la metodología una simple recolección de técnicas y procedimientos (Cook y Reichardt, 1986). Pareciendo más que evidente, que cualquier problema puede ser abordado desde diversas perspectivas.

Esta complementariedad se genera en aspectos funcionales de los procesos evaluativos. Siendo la propia necesidad social la que ha solicitado de la investigación evaluativa argumentos firmes para la resolución de sus problemas reales, recurriendo si fuera necesario a planteamientos pragmáticos y heterodoxos frente a consideraciones más cercanas a la filosofía de la ciencia.

Es por tanto, de esta necesidad, de dónde surge la disposición para utilizar todos los planteamientos, metodologías e instrumentos necesarios para resolver dichas situaciones en un contexto plural.

Como argumenta De Miguel (1999), la consecuencia lógica de toda evolución hacia una evaluación pluralista ha tenido claras repercusiones sobre los paradigmas de investigación, siendo la evaluación de programas uno de los ámbitos de la investigación social donde se hace más patente la superación del debate entre metodología cualitativa y cuantitativa (Cook, 1997).

Las características especiales y específicas de los programas obligan a abordar los procesos evaluativos con planteamientos integradores y enfoques multimétodos, dejando la polémica paradigmática en un segundo plano, como ya planteaban (Fetterman, Kaftarian y Wandersman, 1966), frente a la necesidad de avanzar en el conocimiento de la realidad social.

Así, se ha venido consolidando una corriente epistemológica que intenta promover como marco teórico para este tipo de investigación un paradigma orientado hacia la explicación social realista (Pawson y Tilley, 1997) siguiendo la tradición que sitúa la explicación científica entre el positivismo y el relativismo (Sayer, 1984).

2.6. CONSIDERACIONES METODOLÓGICAS EN LA EVALUACIÓN DE PROGRAMAS EDUCATIVOS.

2.6.1. Los diseños de investigación evaluativa para su aplicación en evaluación de programas.

La evaluación de programas es un proceso de generación de explicaciones útiles sobre una determinada intervención innovadora. Que en un sentido crítico de su utilidad puede emplearse para mejorar los procesos de acción

educativa, razón por la que en el ámbito metodológico, puede entenderse como un proceso de investigación evaluativa (Tejedor, García-Varcárcel y Rodríguez, 1994; Pérez Juste, 1995; Amézcuca, 1996; Martínez Mediano, 1996 y Tejedor, 2000).

Se entiende como diseño a un plan o esquema del investigador que recoge las estrategias que se utilizarán en el estudio. En los diseños evaluativos se fijan qué medidas y cuando serán tomadas en el curso del proceso de evaluación.

Estos planes, conllevan una resolución teórica determinada en los principios que fundamenten el proceso investigativo reflejados en el modelo teórico seleccionado, aspectos especificados en el anterior epígrafe de este mismo capítulo.

Anguera (1992) define al diseño de investigación como estrategia integral de proceso y en consecuencia el curso de acción o secuencia de decisiones acerca de cómo recoger, ordenar y analizar los datos, siempre subordinado al cumplimiento de los fines del programa.

En su concreción, la variedad de diseños de investigación y de intentos de clasificar éstos es amplísima, pero es cierto que dadas las peculiares características de la investigación educativa hay diseños más adecuados para investigar la intervención mediante programas. Aunque su tratamiento minucioso supondría una tarea compleja, ardua e inútil, dada la abundancia de manuales monográficos, se exponen a continuación unas clasificaciones generales en las que se puedan fundamentar los datos de esta investigación en este aspecto concreto.

Fitz-Gibbon y Morris (1987) hacen referencia en su propuesta a siete modalidades de diseño:

- Diseños pretest post-test con grupo de control.
- Diseño post-test (únicamente) con grupo de control.

- Diseño pretest post-test con grupo de control no equivalente.
- Diseño de series temporales con un único grupo de control (experimental solamente).
- Diseño de series temporales con grupo de control no equivalente.
- Diseño antes-después.

Es ésta una clasificación simple, en la que se olvidan determinadas estrategias que Tejedor diferencia (1994) y actualiza (2000).

Por lo que para ofrecer una perspectiva completa, todas estas propuestas de ordenación son asimiladas y contextualizadas aquí en los diseños clásicos de investigación, señalados por Colás y Rebollo (1993), para la evaluación de programas:

a. Diseños para la investigación cuantitativa, agrupados en descriptivos, correlacionales y experimentales:

a.1. *Diseños descriptivos* que tienen como principal objeto la descripción sistemática de hechos y características de una población o área de interés de forma objetiva y comprobable. Poseen gran capacidad para identificar aspectos relevantes o sugerir factores explicativos de situaciones.

Entre ellos destacan los de tipo Descriptivos - comparativos (Tipo survey) con diversa temporalidad: Longitudinales secuenciales (diacrónico) y transversal (sincrónico).

a.2. *Diseños correlacionales*, que trabajando en entornos naturales y sin manipulación previa de variables, tratan de investigar las relaciones entre ellas.

Son una amplia familia tanto en la forma de proceder como en los objetivos científicos que pretenden alcanzar. Entre ellos destacan los que analizan la correlación y regresión (discontinuidad en la regresión), los estudios predictivos y las técnicas de clasificación, entre las que se pueden considerar los análisis factoriales (simplificar y organizar estructuras), los análisis discriminantes (diferenciar grupos por perfiles) y los análisis de conglomerados (clasificar sujetos en grupos). Así como los diseños causales.

a.3. *Diseños técnicos experimentales*, que integran numerosas modalidades, pero que Pereda (1987) diferencia por la manipulación directa de la variable independiente o tratamiento y por el control de las posibles variables contaminadoras. Atendiendo a criterios sobre su validez interna y externa se clasifican en:

- Preexperimentales: Pretest con un solo grupo, Pretest - posttest con un solo grupo y Diseño de comparación estática.
- Cuasiexperimentales: Pretest posttest con grupo de control no equivalente y de Series temporales con pretest y posttest con un solo grupo o con grupo de control
- Experimentales (intergrupo e intragrupo): Pretest con grupo de control, Pretest posttest con grupo de control, N=1 (línea base), Intrasujetos y Diseños Complejos.

b. Diseños para la investigación cualitativa que siguen una secuencia interactiva flexible y abierta, propia de esta metodología, entre las fases de recogida de datos, hipótesis, muestreo y elaboración de teorías.

Se basan fundamentalmente en técnicas *observacionales* y pueden clasificarse de modos muy diversos teniendo en cuenta diversos aspectos. Atendiendo al número de sujetos sobre los que se focaliza la atención pueden

destacarse; estudio de casos, estudios de grupo, estudios de comunidades o etnográficos y estudios de situaciones o contextos.

c. Diseños de investigación desde la perspectiva crítica, muy vinculados a la acción e intervención, como el de investigación- acción. Con procesos cíclicos de naturaleza dialéctica en los que los implicados se convierten en agentes activos, se realizan interpretaciones colectivas de las situaciones para la mejora.

Los planteamientos establecidos para la selección de un diseño en una determinada situación investigativa son muy diversos. En la tradición experimental clásica se ha venido ateniendo a criterios exclusivamente metodológicos relacionados con cuestiones como su precisión, simplicidad o validez.

Colás y Rebollo (1993) adecuan estos criterios a las peculiaridades en la evaluación de programas, contemplando como criterios para su selección el análisis previo de las cuestiones a evaluar y el examen de la situación real donde se va aplicar. Para seleccionar el diseño que mejor se adapte a la situación y contexto específico (Álvarez Rojo, 1991).

Pero en ocasiones, la complejidad de los contextos sociales en los que deben aplicarse estos diseños de investigación hace que ninguno de ellos se adapte completamente o respete las condiciones de éstos, por lo que autores como Fernández Ballesteros (1996) propugnan una necesaria complementariedad metodológica en estos casos.

Además de las cuestiones referidas al diseño de investigación, en los procesos de evaluación de programas hay que tomar decisiones sobre aspectos referidos a las fuentes de información de las que se extraerán las informaciones, qué instrumentos se emplearán para ello, y a qué tipo de análisis se someterán estos datos. Todo ellos bajo la perspectiva del control de los posibles sesgos que se puedan asumir como corresponde a los procesos de investigación educativa.

2.6.2. Fuentes de datos.

Un problema asociado al diseño y que se relaciona a los instrumentos de recogida de información es la delimitación de las fuentes de datos y los lugares de los que se extraen las informaciones.

Cuando se aborda la evaluación de programas se siente la tentación de evaluar a todos y cada uno de los individuos sobre los que actúa éste, lo que no siempre es posible. Apareciendo la extracción de muestras como mecanismos plausibles que permiten según Som (1996) una mayor economía de costos y temporal, el tratamiento de espectros más amplios de informaciones, una mayor fiabilidad y una mejor calidad general de los trabajos.

Existen formas muy diversas de seleccionar la muestra, que se pueden agrupar de forma muy general en métodos probabilísticos y no probabilísticos. Los primeros son importantes en la evaluación de programas, ya que cubren extensas poblaciones, pudiendo ofrecer resultados válidos y fiables y con capacidad de generalización. Mientras que los segundos, y en concreto el intencional muy utilizado en los enfoques cualitativos, se hace recomendable cuando se desea obtener informaciones extensas y profundas sobre una determinada cuestión.

El tipo de muestreo que se puede plantear para la selección de la muestra en los procesos de investigación, siguiendo a Gaviria (2000) que los relaciona específicamente con la evaluación de programas institucionales, planteando además de las numerosas ventajas que supone la extracción de muestras, los problemas derivados de su representatividad poblacional.

Para este autor, los procesos y técnicas concretas de selección de la muestra deben partir de un establecimiento claro de la población objetivo y una delimitación de las poblaciones implícita y explícitamente excluidas. Así como del establecimiento de un tamaño muestral, no determinado por el tamaño de la

población, sino por el error que se está dispuesto a sumir, la desviación típica de la variable medida y la correlación intra-clase en el caso de muestreos por conglomerados.

Igualmente asume la necesidad del establecimiento de ponderaciones en los muestreos por dominios o subdivisiones de la muestra (estratos o conglomerados) para que éstos posean una representación real y acorde con su tamaño específico.

2.6.3. Técnicas e instrumentos de recogida de información.

El esfuerzo económico, temporal y personal realizado en la implementación de un programa, debe ser evaluado. Lo que supone, para Fernández Ballesteros (1996), el uso de instrumentos y procedimientos de recogida de información y el análisis y la interpretación de los datos obtenidos.

El empleo de estas técnicas, queda determinada para Cook (1984), por la necesidad de su complementariedad como medio de captación de un número de informaciones lo suficientemente amplio o completo, como para permitir la descripción de los complejos contextos sociales en los que tiene lugar el hecho educativo.

Complementariedad que a su vez facilita la triangulación de técnicas y procedimientos, necesarios como para que estas informaciones posean una fiabilidad adecuada a procesos de carácter investigativo (Denzin, 1970). Esta complementariedad como enfoque más acertado debe llevarse a cabo empleando de forma complementaria enfoques, métodos y técnicas también en el análisis de datos (Lizasoain y Joaristi, 2000).

Existen numerosas clasificaciones atendiendo a diversos criterios. Fernández Ballesteros (1996) categoriza estos procedimientos para la recogida de

información en técnicas cualitativas y cuantitativas. Existiendo evaluaciones en las que se intenta integrar datos desde ambas perspectivas, aunque sólo sea porque por separado ambos comparten el mismo reduccionismo en su intento de descubrir la realidad (Dávila, 1994). Cuantitativa para extraer datos aprovechables, y cualitativa para organizar las cantidades obtenidas, pues es imposible dejar a un lado lo subjetivo incluso en el plan más estructurado, como suponía Patton (1984).

En el mismo trabajo de Fernández Ballesteros se describen los principales instrumentos de recogida de información en evaluación de programas atendiendo al sujeto de la evaluación. Por su parte Colás y Rebollo (1993), basándose en lo expuesto por Pourtois y Desmest (1988), describen los instrumentos de recogida de información de una manera aplicada, según las cuestiones que se pretenden evaluar.

Aquí se intentan integrar ambas perspectivas resumiéndolas en la presentación de técnicas sintetizadas en la tabla 2.3.y 2.4.

Esta diversidad de técnicas no debe llevar a la confusión, pues hay que aplicar todas aquellas que sean útiles sin más limitación que las restricciones derivadas de sus propias características, basándose los criterios de selección de una técnica o instrumento concreto en algunas de sus características como:

- Aplicabilidad (que el procedimiento se pueda llevar a cabo en el programa).
- Adecuación (a las condiciones del objeto de estudio e hipótesis de trabajo).
- Rigor (conocer las condiciones propias de aplicación).
- Competencia técnica del evaluador/es y asequibilidad de las técnicas.
- Obtención de la máxima riqueza y precisión de los datos.
- Rentabilidad temporal, de esfuerzo y económica.

- Adaptabilidad (dar respuesta a las expectativas del personal, cliente o audiencias a las que se dirige).

Y sobre todo, teniendo en cuenta las dimensiones evaluativas de las que se desea o necesita extraer información, como la inteligencia; las aptitudes particulares y personalidad, conductas y acciones; el conocimiento; situaciones y procesos de intervención o dimensiones sociales del grupo, valores, intereses y roles; unas pruebas serán más adecuadas que otras.

Tabla 2.3. Técnicas de recogida de información en evaluación de programas (I)

	TÉCNICAS Y PROCEDIMIENTOS	SUJETO DE EVALUACIÓN	FORMA Y PROCESO DE APLICACIÓN	ÁMBITO DE APLICACIÓN
C U A N T I T A T I V A S	Entrevista estructurada	Responsables. Gestores.	Entrevistador experto, delimita el propósito y objetivos, que convierte en cuestiones.	Evaluación diagnóstica. Evaluación producto.
	Encuesta/ Cuestionario	Gestores. Usuarios.	Personal experto que define objetivos, diseña la prueba, decide la muestra de aplicación, codifica e interpreta las informaciones, aplicada por Encuestadores entrenados.	Evaluación diagnóstica Evaluación producto (aptitudes, percepciones o posiciones personales).
	Técnicas de autoinforme	Usuarios.	Autoaplicación.	Evaluación formativa o procesual.
	Observación	Contextos de aplicación de programas.	Observadores expertos registran aspectos concretos de las conductas o del contexto.	Evaluación formativa o procesual. Evaluación diagnóstica. Evaluación del producto (Interpretación de éxito y fracaso de acciones).
	Indicadores sociales, económicos y de salud	Población General. Población Objetivo.	Personal contratado recoge información en plantillas estructuradas.	Evaluación diagnóstica.

	Test	Población Objetivo. Usuarios.	Personal experto que selecciona la prueba, aplica y corrige en condiciones y normas específicas.	Evaluación diagnóstica. Evaluación producto.
	Escalas de actitud	Población Objetivo. Usuarios.	Personal experto que determina la actitud a medir, redacta y elabora la prueba, y analiza e interpreta los datos.	Evaluación diagnóstica. Evaluación producto.
	Fuentes documentales y Datos de archivo sobre usuarios o el programa	Usuarios. Documentación escrita sobre el programa.	Personal contratado recoge información en plantillas estructuradas. Evaluadores expertos.	Evaluación diagnóstica. Evaluación personal sobre realización de actividades Evaluación de productos.

Tabla 2.4. Técnicas de recogida de información en evaluación de programas (II)

	TÉCNICAS Y PROCED.	SUJETO DE EVALUACIÓN	FORMA Y PROCESO DE APLICACIÓN	ÁMBITO DE APLICACIÓN
C U A L I T A T I V A S	Grupos de discusión	Responsables. Gestores. Usuarios.	Moderador con dominio de técnicas de animación.	Evaluación diagnóstica. Evaluación procesual. Evaluación de productos.
	Observación participante	Contexto de aplicación de los programas.	Agente de campo u observador participante que se introduce en el contexto de aplicación del programa como un miembro más del mismo.	Evaluación diagnóstica. Evaluación formativa o procesual. Evaluación de productos.
	Analizadores sociales y técnicas sociométricas	Población objetivo.	Obtención de analizadores en cuanto a claves explicativas del funcionamiento de los grupos.	Evaluación diagnóstica del grupo. Evaluación procesual (dimensiones educativas, clima social y guía de acciones).
	Método Delphi	Expertos. Gestores. Responsables.	Grupos de expertos a los que el evaluador envía cuestionarios y resultados de los mismos.	Evaluación diagnóstica. Evaluación de productos.

2.6.4. Análisis e interpretación de datos.

Todo el proceso de evaluación caracterizado hasta aquí requiere un continuo procesamiento de informaciones, que si no se organizan y tratan de la forma correcta pueden convertirse en una tarea ardua, derivar en interpretaciones incorrectas y conseguir que la evaluación sea infructuosa.

Por tanto, las principales funciones asignadas a los procesos de análisis de datos son:

- Describir de forma comprensiva los datos brutos.
- Ayudar a decidir si la relación entre datos se puede admitir con confianza o azar.
- Estimar magnitudes, cantidades de cambio o diferencias.
- Determinar si los efectos que aparecen en la información alcanzan un nivel relevante en el contexto de aplicación del programa.

Con todo ello, el propósito de la interpretación es combinar los resultados obtenidos con los valores asumidos, en criterios y patrones, para elaborar conclusiones juicios y recomendaciones. En este proceso de interpretación influyen tanto los métodos lógicos empíricos, propios de las ciencias estadísticas, como los valores del investigador (Worthen y Sanders, 1987).

Pero antes de hacer explícito cualquier procedimiento concreto para el análisis de datos es necesario realizar una serie de consideraciones preliminares acerca de la perspectiva de análisis a emplear, establecer unas fases ordenadas acordes con éste enfoque, y seleccionar las técnicas y procedimientos más adecuados a los datos de los que se disponga.

Existen diversas formas de realizar los procesos de análisis de datos, determinadas por ejemplo por la naturaleza cuantitativa o cualitativa de éstos. En

líneas generales, basándose en la propuesta fundamentada de Fernández Ballesteros (1996) un procedimiento de análisis de datos posee las siguientes fases:

1. Preparación y organización de los datos, que en datos de tipo cuantitativo consiste en el vaciado y organización de la información en variables y categorías, y en los de tipo cualitativo en la determinación de las unidades básicas de análisis.

En esta etapa es muy importante la creación de matrices de datos bien organizadas y en caso necesario el establecimiento de las escalas de medida que serán necesarias para analizar los datos.

2. Introducción de los datos en un procesador informático, que permite con el apoyo del *software* adecuado, una enorme posibilidad de análisis. Existiendo programas adecuados tanto para el tratamiento de datos cuantitativos (*S.P.S.S., B.M.D.P. o Statistics*) y cualitativos (*The ethnograph, Aquad o Nudist Vivo*).

3. Aplicación de estándares para fijar categorías y definiciones comunes. Categorización y conceptualización de los datos cualitativos.

4. Selección del tipo de análisis a realizar dependiendo de la naturaleza de los datos y del tipo de conclusiones a las que se quiere llegar.

5. Aplicación de los primeros resultados. En los procesos cuantitativos se suele comenzar por un análisis descriptivo de las variables por medio de estadísticos como porcentajes, frecuencias o moda.

6. Análisis inferencial de datos, que permite la extrapolación de los datos de la muestra a la población.

En los análisis cuantitativos de datos, se realiza un contraste de hipótesis para determinar si un resultado se debe al azar al cumplir ciertas hipótesis, se utilizan estadísticos muestrales para estimar y determinar su valor probable en la población y para determinar la significación estadística de los resultados obtenidos.

En los procesos cualitativos se realiza una inferencia con los datos obtenidos para elaborar hipótesis o proposiciones lógicas que expliquen diversos aspectos de la realidad descrita.

En ambos procesos, ya sean referidos a datos de naturaleza cuantitativa o cualitativa, la finalidad es el establecimiento de teorías, explicaciones o razonamientos fundamentados que permitan la toma de decisiones acerca del programa evaluado.

A estas fases primarias puede añadirse fases de carácter secundario propias de los procesos metaanalíticos. En los que se trabaja con resúmenes de estudios o investigaciones primarias, recuperados tras la revisión de la literatura, y seleccionados ateniéndose a unos criterios específicos.

Esta metodología de revisión puede considerarse como una respuesta a la necesidad de desarrollar procedimientos objetivos y cuantificables en la revisión evaluativa de la investigación (Glass, 1976; Wolf, 1987 y Sánchez Meca y Ato, 1989).

La interpretación sucede al análisis de datos y supone la asignación de significados otorgándole sentido a los resultados obtenidos. Ésta debe partir de una lectura lo más exacta posible de los datos, incluso en los casos más negativos, pues de ellos se pueden extraer jugosas e interesantes conclusiones.

Los datos deben contrastarse con los objetivos propuestos en la evaluación y con los extraídos en situaciones similares, intentando conectarlos

con explicaciones generales ofrecidas por teorías explicativas. Todas las interpretaciones realizadas deben a su vez estar fundamentadas y avaladas por los datos obtenidos.

Una última consideración a tener en cuenta en el análisis de datos y la discusión de sus resultados es la contemplación de las distintas perspectivas de los agentes implicados y relacionados con la evaluación del programa. Este aspecto no sólo debe considerarse en la comunicación final de conclusiones, atendiendo a criterios estéticos y de contenido del informe, sino que debe originarse en el propio proceso de tratamiento de datos previo, mediante el uso de procedimientos adecuados como son los grupos de discusión.

2.7. AVANCES Y NUEVAS PERSPECTIVAS EN LA INVESTIGACIÓN SOBRE EVALUACIÓN DE PROGRAMAS. LOS ESTUDIOS META-EVALUATIVOS Y DE SÍNTESIS.

La evolución y desarrollo de la investigación sobre evaluación de programas se encuentra ligada al de la propia investigación educativa en general.

Desde el punto de vista epistemológico, las aproximaciones eclécticas o integradoras suponen un avance en la superación de los principios excluyentes. En el sentido que ofrece Merton (citado en Wittrock, 1989) las nuevas teorías no sustituyen a las anteriores sino que las complementan, permitiendo prestar atención a hechos y fenómenos que podrían pasar desapercibidos desde un único punto de vista.

Este principio permite ofrecer una visión cada vez más completa y adecuada de las realidades examinadas. Sobre todo teniendo en cuenta que el objeto de conocimiento de las ciencias educativas se centra en contextos y realidades de enorme complejidad humana y social en ámbitos interculturales. Por lo que en estos estudios se requieren el empleo de diseños complementarios, cada

vez más aceptados y empleados por los investigadores, como medios adecuados de investigación de estos entornos.

Desde un punto de vista eminentemente empírico se ha producido la superación del debate *cuantitativo vs. cualitativo* (Reichardt y Rallis, 1994 y Cook, 1997). Permitiendo la metodología de base cuantitativa la búsqueda de relaciones entre las variables bajo precisas condiciones de control, aleatorización, medida y diseño, para alcanzar mejores niveles de generalización; mientras que los métodos cualitativos ofrecen mejores y más detalladas descripciones de los procesos. Consiguiendo con una conjugación de ambos estudios muy compactos.

Es desde estas nuevas consideraciones teóricas y procesos de investigación desarrollados, desde donde se han ido considerando una mayor participación de los destinatarios y afectados por los programas evaluados tanto como informantes de los procesos como su conversión en agentes activos de ellos. Ofreciendo informaciones más profundas desde la perspectiva interpretativo humanística y una ética de participación más democrática.

Esta evolución de la investigación sobre E.P. como realidad compleja ha derivado en la aplicación de análisis de datos más sofisticados que permiten alcanzar unas interpretaciones de mayor calado, riqueza y especificidad. Jugando en estos procesos un papel crucial la aplicación de nuevas tecnologías que permiten la gestión de grandes cantidades de informaciones y el tratamiento de múltiples datos por medio de variados análisis.

La consideración de procesos de investigación amplios como medio de explicación de contextos complejos ha desarrollado la atribución de nuevas funciones a la evaluación de los programas mediante los que se interviene en dichos ámbitos. Así, como indica Chelimnsky (1997) o De Miguel (1999), además de funciones como rendición de cuentas (*accountability*) o mejora del programa (*improvement*), las evaluaciones también comienzan a considerarse generadoras

de conocimientos y explicaciones de los problemas sociales, y por tanto determinantes para promover sus cambios y mejoras.

Al igual que en el resto de la investigación educativa las nuevas tecnologías aplicadas al proceso de búsqueda, registro y tratamiento de informaciones basados en redes informáticas, desarrollo y disponibilidad del *hardware* o aplicabilidad de *software* específico, han supuesto una revolución al poner en manos de los investigadores educativos potentes medios que facilitan enormemente los procesos de investigación en Evaluación de Programas. Implicaciones que incluso afectan a aspectos posteriores al propio proceso de investigación como es agilizar la difusión de los hallazgos y facilitar mayores ámbitos de impacto.

Como investigación en el ámbito social de carácter eminentemente práctico, son numerosos los factores sociales que han influenciado el desarrollo de la investigación sobre evaluación de programas, siendo uno de los de mayor importancia el derivado de la mayor preocupación social por aspectos sobre la cualidad de la educación. Preocupación que ha hecho que las instituciones desarrollen procesos de intervención por programas de espectro más amplio y de aplicación a contextos específicos por medio de mecanismos en los que la cultura evaluativa se acepta y considera como mecanismo ineludible para la consecución de la ansiada calidad educativa.

Uno de los mayores avances alcanzado en E.P. supone un refinamiento metodológico, como es el caso de las investigaciones mediante metaanálisis y estudios de síntesis, que permiten realizar inferencias desde supuestos más amplios ofreciendo una mayor extensión de los resultados obtenidos.

Los estudios de síntesis de carácter secundario se distinguen por el empleo como fuentes de datos de investigaciones primarias sobre un tópico determinado, para por medio del empleo de estadísticos que permitan realizar síntesis de los resultados obtenidos, permitir y facilitar el replanteamiento del

campo de investigación de una disciplina al ofrecer visiones globales y replantear nuevas líneas de actuación.

La pertinencia de estudios metaevaluativos y de síntesis es usualmente señalada e indicada por autores en el ámbito internacional (Scriven, 1974; Stufflebean, 1978, 1981; Shadish, Cook y Levinton, 1991) y nacional (Bartolomé, 1990; De la Orden y Martínez de Toda, 1992 o Fernández Cano y Bueno, 1998).

El desarrollo actual de los servicios evaluativos, hasta alcanzar un grado de profesionalización (Madaus *et. al.*, 1983), es una muestra de la patente preocupación social por la educación en la que el empleo de estos servicios se muestra como aval de su calidad. Sin embargo, a pesar de las muchas razones para ello no ha existido tradicionalmente una cultura evaluativa, sino más bien como señalaba Scriven (1974) una resistencia a ser evaluado y una escasez de revisión de los estudios de evaluación.

Si evaluación es la emisión de un juicio sobre el valor o mérito de algo, metaevaluación es enjuiciar la propia evaluación. De la Orden y Martínez de Toda (1992) conceptualizan los procesos de metaevaluación como sistema de garantía y responsabilización ante la sociedad de los servicios evaluativos ofrecidos, basándose en el desarrollo específico de las normas establecidas por el *Joint Committee* (1998).

Así los procesos metaevaluativos pueden adoptar diferentes funciones o utilidades en la investigación educativa:

- La primera de ellas hace referencia a la metaevaluación como evaluación de la evaluación (Stufflebean, 1981; Landsheere, 1982 y *Joint Committee*, 1998) incluso en el sentido de revisión para el cumplimiento de los procesos evaluativos desde una perspectiva de profesionalidad (Stufflebeam, 1987).

- Esta valoración de la evaluación para atestiguar su calidad, puede proceder de una perspectiva sumativa (Cook y Gruder, 1978) o formativa (Scwandt y Halpern, 1988), convirtiéndose la metaevaluación en una herramienta de mejora de la evaluación mediante un proceso de identificación de sus sesgos.

- Los procesos metaevaluativos poseen un carácter unificador al proveer procedimientos para resumir y sintetizar los resultados de varias valoraciones referidas al mismo programa o conjunto de programas (Stufflebean, 1981). Procesos que permiten ofrecer visiones amplias desde distintos planos como la función valorativa general de una organización o las realizaciones en un determinado campo o disciplina. Siendo muy útiles, sobre todo en procesos de reconstitución y reforma, los estudios de síntesis.

- Los contextos y ámbitos de intervención necesitan de intervenciones por programas cada vez más amplias y complejas, que a su vez delimitan procesos evaluativos de características similares, por lo que si cabe, la revisión metaevaluativa adquiere aún un sentido mayor.

- Estos procesos insertos en una dinámica constante de crítica constructiva, hacen trascender la práctica evaluativa particular hacia aspectos más amplios como vehículo potenciador de mejoras educativas (Worthen y Sanders, 1987).

La importancia y trascendencia que puede derivarse de cada una de estas funciones, hace que el desarrollo de procesos metaevaluativos y estudios de síntesis en la investigación sobre evaluación de programas parezca tan irrenunciable como el de la propia evaluación de éstos.

En algunas de estas características, como la complementariedad, su carácter de síntesis y el empleo de nuevas tecnologías, los procesos desarrollados

en esta Tesis Doctoral se basan en los nuevos desarrollos y pueden suponer un avance en el campo de la E.P.

Complementariedad que puede ser entendida desde múltiples aspectos teóricos y empíricos, como la consideración de análisis de aspectos cuantitativos, conceptuales y metodológicos de la investigación; el empleo de una metodología descriptiva desde una perspectiva sincrónica y diacrónica o los análisis de datos, que aunque eminentemente cuantitativos, se recurre al empleo de procesos cualitativos en aquellos aspectos en los que la naturaleza específica de los datos lo requiere. No existiendo precedentes en el campo de la E.P. de estudios de síntesis de este tipo.

También puede parecer novedoso el empleo de las nuevas tecnologías en todas las fases del proceso investigativo, desde la búsqueda por medio de enlaces electrónicos *on line* en diversas bases de datos mediante el empleo de motores de búsqueda basados en palabras clave, el registro de datos en bases digitales y tratamiento por medio del empleo de *software* específico, a la elaboración de informes de conclusiones.

CAPÍTULO III. CIENCIOMETRÍA Y EDUCACIÓN.

“Los indicadores cuantitativos pueden ser un instrumento útil para la evaluación de la investigación educativa. Pero se ha de poner el máximo cuidado para su uso correcto, evitando simplificaciones excesivas como inútiles refinamientos”.

(Maltrás, Quintanilla y Vidal, 1998)

3.1. INTRODUCCIÓN.

Uno de los apelativos que definen con mayor claridad la época actual es el que la denomina como *Era de las Comunicaciones*. Sea cual sea la disciplina en la que nos situemos, la cantidad de información generada y sus posibilidades de difusión son tan elevadas que se requiere un esfuerzo de estructuración (Galante y Sanchidrian, 1996). De hecho, ya en el segundo Congreso de la Ciencia, celebrado en Londres en 1931, se perfilaban lo que serían las directrices básicas de este análisis estructural de la información que posteriormente constituiría el cuerpo metodológico conocido con el nombre de *Bibliometría* o *Cienciometría*.

La importancia que la ciencia y la tecnología han llegado a adquirir en la sociedad, ha ido imponiendo la necesidad de su evaluación. Son numerosos los

recursos humanos y económicos que la sociedad dedica a la investigación en el desarrollo de trabajos científicos, como medio de su propia evolución y sostenimiento, que a su vez ésta exige un control y evaluación de sus tareas que justifiquen tal aporte. Siendo la metodología cuantitativa una de las más comúnmente utilizadas.

Se trata de estudiar científicamente la actividad de la investigación no para discernir su naturaleza o para relativizar su alcance, sino para favorecer su desarrollo.

La Cuantimetría o cuantimetría, también denominada Metodología Bibliométrica es ante todo un corpus metodológico inicialmente centrado en el análisis estructural y prognosis de la información, contenida principalmente en publicaciones de investigación científica.

López Piñero (1972) la denominaba "ciencia de la ciencia", considerándola como un auténtico programa de investigación de carácter interdisciplinar que interesaba a disciplinas complementarias: la información científica propia de los documentalistas, la historia de la ciencia y la sociología de la ciencia; dimensiones a las que se podría añadir la evaluación científica de las producciones, propia de la Metodología de la Investigación (Fernández Cano y Bueno, 1998).

Debido a su metodología los estudios cuantimétricos poseen un carácter eminentemente descriptivo, por lo que poseen gran importancia para obtener visiones globales de un determinado campo de estudio.

Se suele partir, en estos estudios, del análisis de la frecuencia de aspectos tales como: autores; fechas y fuentes de publicación; número, lengua, antigüedad y variabilidad de las citas. Para, mediante el cálculo de valores promedio y porcentajes, realizar una descripción diacrónica de la producción de una disciplina o de un campo determinado de una disciplina científica.

3.2. LA COMUNICACIÓN EN CIENCIA Y TECNOLOGÍA.

El desarrollo de la Cuantimetría ha sido posible gracias a un rasgo característico de la ciencia desde el siglo XVII: la comunicación científica (López López, 1996). Quedando así el desarrollo científico íntimamente asociado a su comunicabilidad.

Para Ziman (1980) la ciencia sólo puede existir en estrecha dependencia con la comunicación, ya que es esencial para establecer criterios de registro de resultados, observaciones o teorías que sirvan como referencia a los científicos. Facilitando a su vez la crítica, refutación y perfeccionamiento de los hechos y trabajos científicos.

Aunque es ésta una postura abierta a muchas críticas, Price (1973, p.69) también asocia la existencia, de la ciencia a la de la comunicación hasta tal punto que no podría existir sin ella, definiendo la ciencia como *“lo que se edita en las publicaciones científicas y al científico como el hombre que en algún momento de su vida ha colaborado escribiendo alguna de esas publicaciones”*.

Incluso desde una perspectiva sociológica la comunicación científica parte de la consideración de la ciencia como una organización dentro de la propia estructura social. Y en todo sistema organizativo de carácter social, como señala Gótor (1982), el establecimiento de un sistema de comunicación es inherente al propio desarrollo de la organización.

Pero en palabras de Carpintero y Peiró (1981), incluso la comunicación científica ofrece características significativamente diferentes a la comunicación en cualquier otra organización, puesto que al convertirse en la prueba más efectiva de las realizaciones y productos de los científicos, se ha convertido en un fin en sí misma.

Es decir, como planteaba Price (1973), el hecho y proceso científico quedaría incompleto sin la comunicación, mediante la publicación, de los hallazgos para su discusión y aceptación por la comunidad científica. Y por tanto una de las finalidades de la ciencia, debe ser su comunicación y su publicación.

Partiendo de este papel central asignado a la comunicación en la ciencia (Carpintero y Peiró, 1981), el estudio de dicha comunicación puede facilitar informaciones útiles acerca de su estructura organizativa. Convirtiéndose este planteamiento en uno de los pilares que ofrecen racionalidad al empleo de la metodología cientimétrica.

En este proceso de comunicación, que hace de la ciencia un conocimiento compartido, el análisis de los canales de comunicación es prueba y medida del desarrollo científico en un determinado contexto. Estos pueden poseer para Montoro y Carbonell (1989) naturalezas muy diversas adoptando desde aspectos informales como la correspondencia o entrevistas; o eminentemente formales como Tesis Doctorales, literaturas de congresos, informes o revistas de investigación.

Estas últimas convertidas en el vehículo por excelencia de la comunicación científica, han aumentado de forma ingente como canal comunicativo (Callon, Courtial y Penan, 1995).

Queda por tanto, patente la relación unívoca entre ciencia, comunicación y publicación en fuentes científicas, principalmente revistas especializadas. Siendo este el principal objeto de estudio de la Bibliometría/Cienciometría, que Garfield (1979a) define como la cuantificación de la información bibliográfica susceptible de ser analizada.

Esta asociación entre el trabajo científico y la necesidad de publicar en el mundo de la ciencia ligada a la industria necesita además de otra literatura tecnológica como son las patentes. Estas pese a soportar un gran volumen de la

producción científica mundial, quedan restringidas enormemente por lo poco accesible de su lenguaje.

3.3. LA CIENCIOMETRÍA. NOTAS HISTÓRICAS Y DELIMITACIÓN CONCEPTUAL.

El origen de esta disciplina surge de la convergencia de estudios de diversas disciplinas, en momentos y lugares distintos. Solapándose términos utilizados para definir este tipo de trabajos desde su origen, que algunos autores utilizan indistintamente, como son los de Ciencia de la Ciencia, Bibliografía Estadística, Bibliometría, Cienciometría o Infometría. No siendo más que fruto de su propio desarrollo, desde las aplicaciones esporádicas hasta la conceptualización metodológica de la disciplina, que cada vez se hace más amplia, tanto en su objeto como en el campo de conocimiento.

Desde la antigüedad se han venido realizando estudios bibliométricos, aunque sin el desarrollo lógico de una disciplina establecida. Méndez (1986), indica que el primero en utilizar un enfoque de este tipo fue Alphonse de Condolle, que en 1885, en su trabajo *Histoire des sciences et des savants depuis deux siecles* realiza un estudio sobre catorce países europeos y Estados Unidos, aplicando métodos matemáticos a factores relacionados con el desarrollo científico.

El primer estudio en el que podríamos determinar rasgos ciertamente bibliométricos es el de Cole y Eales, que en 1917, realizaron un análisis estadístico de las publicaciones sobre anatomía comparativa entre 1543 y 1860 con la distribución por países y campos, cómputos, categorías y las divisiones del reino animal; trabajos que según (Sancho, 1990) todavía tienen cierto interés bibliométrico.

Pero hasta 1923, no se realizan referencias literarias a éste tipo de estudios atribuyéndoles un carácter específico, mediante el término de *bibliografía estadística*. Referencias realizadas por Hulme en sus conferencias en la Universidad de Cambridge o en sus estudios de historia de la ciencia, en los que analiza autores y revistas referenciadas en el *International Catalogue of Scientific Literature* en el periodo 1910-1913, analizando la relación entre publicaciones científicas, condiciones económicas y crecimiento de la civilización moderna.

También Gross y Gross (1927), comienza a utilizar este término como primer paso diferenciador de una materia emergente, al analizar las referencias hechas en artículos de revistas sobre química indizadas en *The Journal of American Chemistry Society*. En 1926, el ruso Borichevski esboza en un artículo la constitución de un nuevo campo de investigación cuyo objeto es “el estudio de la naturaleza intrínseca de la ciencia”, considerándola como ciencia de la ciencia (*naukovodemia*).

También en este mismo año, Lotka (1926) formula su conocida ley de la productividad de los autores científicos; y todavía en 1934, cuando Bradford realizó su trabajo sobre la distribución de artículos en revistas sobre geofísica aplicada y en investigaciones sobre lubricantes, ahora conocido como Ley de Bradford seguía utilizándose el término de bibliografía estadística para describir este tipo de trabajos de la ciencia de la ciencia. Sin embargo, con posterioridad, se consideró que no era un término satisfactorio para describir esta actividad científica, debido a que podía ser interpretado erróneamente como bibliografía sobre estadística.

Al margen de su denominación, el desarrollo de la disciplina continuó afianzándose, basándose en el carácter práctico de sus realizaciones. Pues es común la asociación del desarrollo de una ciencia, por empleo de bibliografía o el número de descubrimientos, con el desarrollo social y económico.

Y aunque el origen de su denominación específica se lo atribuyen algunos autores a Paul Olet, no es hasta 1969, cuando Alan Pritchard considerando tosco y poco descriptivo ese término, acuña el de *bibliometría*. Pritchard definió la bibliometría como la aplicación de la matemática y los métodos estadísticos para analizar el curso de la comunicación escrita y de una disciplina. Es decir, la aplicación de tratamientos cuantitativos a las propiedades del discurso escrito y los comportamientos típicos de éste.

Pero al abogar por una ciencia de la ciencia, ampliaba considerablemente la perspectiva de la Bibliometría, más restringida a recuentos de producciones en determinadas disciplinas o incluso a estudios más detallados. La ciencia de la ciencia debería ir más allá en la elaboración y aplicación de instrumentos estadísticos. Su finalidad es identificar las leyes y regularidades sobre el funcionamiento de la actividad científica en su globalidad.

El pleno desarrollo de esta disciplina está asociado a los trabajos de dos grandes pioneros, que realizaron sus trabajos de forma paralela. Price, que en 1963 y 1964, apoyándose en toda una serie de estudios de carácter bibliométrico anteriores, publica respectivamente: *Little Science, Big Science* y *Science since Babylon*. Y Garfield, que desde la aparición en 1963 del Science Citation Index, publicado por el *Institute for Scientific Information (ISI)* fundado por él mismo en Filadelfia, ha contribuido enormemente al desarrollo de este campo científico.

Así mismo, de gran importancia para la transformación profunda y diversificación fue la aparición, en 1979 de una fuente de publicación específica, como fue la revista *Scientometrics* que ofrecía una existencia visible a esta Área de conocimiento.

El término *Cientimetría o Cienciometría* surgió en la URSS y Europa Oriental (*Naukometriya*) y fue practicado especialmente en Hungría. Originalmente se refería a la aplicación de métodos cuantitativos a la ciencia y el progreso tecnológico (Egghe, 1988).

En 1935 aparece bajo la pluma de Ossowski la expresión *Nauka o nause*, designando un amplio programa de investigación que subraya la necesidad de impulsar estudios históricos, sociológicos y psicológicos consagrados al estudio de las ciencias. Pero el cambio hacia análisis eminentemente cuantitativos no se produce hasta que Dobrov entra en relación con Price y sistematiza las intuiciones de éste último. Creándose rápidamente un Instituto dedicado a la investigación sobre estos aspectos y apareciendo los primeros libros sobre la “nueva materia”.

Diferencialmente la bibliometría estudia la organización de los sectores científicos y tecnológicos a partir de las fuentes bibliográficas y patentes para identificar a los actores, sus relaciones y tendencias. Mientras que la cienciometría, por el contrario, se encarga de la evaluación de la producción científica mediante diversos indicadores numéricos como pueden ser de publicación o patentes, entre otros.

En las primeras definiciones y delimitaciones conceptuales se consideraba la cienciometría como “la medición del proceso informático”, donde el término informático no poseía las connotaciones actuales referidas a equipos informáticos, sino a la ciencia que estudia la estructura y las propiedades de la información científica y las leyes del proceso de comunicación (Mikhailov *et al*, 1984).

El significado ampliado que tiene hoy en día el término cienciometría, que sustituye al de *ciencia de la ciencia*, fue promovido por Price, quien impulsó este campo de la investigación en la década de los años sesenta desde la Universidad de Columbia, en Estados Unidos.

La cienciometría aplica técnicas bibliométricas, de carácter matemático y análisis estadístico, para investigar las características de las ciencias, tanto físicas, naturales como sociales. Pero la Cienciometría va más allá de las técnicas bibliométricas, pues también examina el desarrollo y las políticas científicas. Sus análisis cuantitativos consideran a la ciencia como una disciplina o actividad

eminentemente social, en algunos de sus múltiples aspectos económicos y políticos, por lo que podría considerarse como un instrumento de la sociología de la ciencia.

Pero este campo de conocimiento sigue en evolución con el desarrollo de nuevas consideraciones, como muestra la aparición de nuevas denominaciones como la Infometría, que amplían su campo de acción.

Algunos autores usan bibliometría e infometría como sinónimo (Rousseau, 1990), aunque otros consideran que la segunda incluye un área mucho más amplia.

La *infometría* se basa en las investigaciones de la bibliometría y cuantimetría, y comprende asuntos tales como el desarrollo de modelos teóricos y las medidas de información, para hallar regularidades en los datos asociados con la producción y el uso de la información registrada.

De acuerdo con Brookes (1990), el término infometría fue propuesto por primera vez por Nacke, director del Institut für Infometrie und Scientometrie de Bielfeld, Alemania en 1979. Término que fue inmediatamente adoptado por el VINITI de la URSS, cuyos estudios consideran que la infometría comprende a la bibliometría y la cuantimetría (Gorkova, 1986).

La infometría trata de la medición de todos los aspectos de la información, el almacenamiento y su recuperación, por lo que incluye la teoría matemática y la modelización. En sentido amplio estudia los aspectos cuantitativos de la información, no solamente la registrada como los registros bibliográficos, sino todos los aspectos de la comunicación formal o informal, oral o escrita.

Las aplicaciones prácticas de todos estos tipos de estudios evaluativos se dan en contextos tan variados como la recuperación de la información, la administración de archivos o bibliotecas, la historia de la ciencia, la política científica, la lingüística cuantitativa, la econometría. Y en cuanto a su contenido, absolutamente en todas las áreas o disciplinas científicas.

La cienciometría ha alcanzado su edad adulta siendo su crecimiento espectacular. Y aunque la variedad de sus orientaciones y aplicaciones podría haber puesto en tela de juicio su integridad como disciplina, ésta se ha mantenido, ya que los estudios cuantitativos comparten tres convicciones inamovibles:

La primera de ellas es que el estudio de las ciencias y de las técnicas pasa necesariamente por el análisis sistemático de las producciones literarias de los investigadores: ciertamente, la cienciometría no se limita exclusivamente a este objeto, pero le concede un lugar esencial.

La segunda es que los estudios cuantitativos, siempre que no constituyan un fin en sí, enriquecen la comprensión y la descripción de la dinámica de las *tecnociencias*.

Y la tercera es la prioridad absoluta y casi obsesiva que conceden a la concepción de instrumentos de análisis sólidos y fiables.

Actualmente las investigaciones de cienciometría y bibliometría se realizan con mayor intensidad en Estados Unidos, Canadá y Europa Oriental. Destacando también por su importancia las elaboradas en Japón y Europa Occidental.

3.4. LA CIENCIOMETRÍA EN ESPAÑA.

Aunque contemporáneamente a los estudios de Bradford, Ortega y Gasset plantea en su obra *La misión de bibliotecario* de 1935 (Riaño, 1991) la necesidad de una “estadística de las ideas” que permita facilitar la comprensión de su origen, su periodo de desarrollo y expansión y declive, no es hasta la década de los setenta en la que comienza a tomar cuerpo esta idea en nuestro país.

Durante el desarrollo de la década de los setenta se conforman tres grandes grupos de investigación desde los que comienza a desarrollarse esta metodología cientimétrica:

A principios de la década, y entorno a la figura de López Piñero y Terrada, pertenecientes a la Facultad de Medicina y al Instituto de Estudios Documentales e Históricos sobre la Ciencia de la Universidad de Valencia y C.S.I.C.¹¹, se comienzan a publicar trabajos de este tipo. Entre los que destacan *El análisis estadístico y sociométrico de la literatura científica* o *Hacia una Ciencia de la Ciencia* - Traducción de la obra de Price, *Little Science* - (López Piñero, 1972 y 1973) y también, un estudio desarrollado por Terrada (1973), que bajo el título *La literatura médica española contemporánea*, se centra en un análisis estadístico y sociométrico.

Éste es el germen de la cientimetría en nuestro país, publicándose varios artículos atendiendo a este enfoque por parte de discípulos de estos autores, como los realizados por Carpintero (1981) o Carpintero y Tortosa (1993), desde la Facultad de Filosofía y CC. E. de Valencia.

El segundo de los grupos mencionados, liderado por Pérez Álvarez-Ossorio y Lara Guitard desde el I.C.Y.T./C.I.N.D.O.C.¹² en Madrid, posee una perspectiva más dirigida al análisis de la política científica, de cuya línea de

¹¹ CSIC.: Centro Superior de Investigaciones Científicas.

investigación es muestra alguno de los trabajos realizados por autores como Gómez y Bordons (1996a,b).

El tercer grupo desde cuyos planteamientos se ha desarrollado esta disciplina en nuestro país, desarrolla trabajos desde la Universidad de Salamanca centrados en aspectos relativos a la difusión del conocimiento, como los de Maltrás y Quintanilla (1992), Vidal (1993) o Maltrás (1996).

Todos estos trabajos establecen el germen del desarrollo de la metodología de esta disciplina, determinando líneas de investigación todavía activas y consolidadas en la década de los noventa, como queda patente en las numerosa y diversa literatura producida por autores como Alonso y Gra (1994), Jiménez Contreras (1994) o López Yepes (1995).

En la Universidad de Granada, comienzan a desarrollarse algunos trabajos fundamentados en esta perspectiva. Como son los elaborados por autores como Delgado López-Cózar (1999), Fernández Cano y Bueno (1998, 1999) o Torralbo (2001). Siendo precisamente en este contexto concreto, en el que se desarrolla la línea de trabajo del grupo de investigación sobre Evaluación de la investigación y de programas educativos andaluces, desde el que parte el desarrollo de esta tesis doctoral.

Parece por tanto, que en la investigación española sobre este campo disciplinar, realizada desde estos centros universitarios, se pueden estar comenzando a establecer líneas de investigación con cierta especialización o cuanto menos con cierta especificidad.

¹² I.C.Y.T. /C.I.N.D.O.C.: Instituto de Información y Documentación en Ciencia y Tecnología,

3.5. INDICADORES BÁSICOS Y LEYES CIENTIMÉTRICAS.

Un *indicador* es un parámetro empleado en la evaluación de cualquier actividad, que puede ajustarse a patrones teóricos establecidos en investigaciones previas en enunciados o leyes de algún campo de la ciencia, en este caso la Cienciometría. La información científica puede por tanto adquirir una serie de características, mediante el uso de unos indicadores descritas mediante el establecimiento de las que se han venido en llamarse, quizás con alguna licencia lingüística, *Leyes Cientimétricas*.

Mediante el empleo de una serie de indicadores bibliométricos y la comprobación de su ajuste a los patrones propuestos en las leyes cientimétricas se pretende, en este estudio, describir algunos aspectos de la producción científica en el campo de evaluación de programas en el ámbito y periodos estudiados, tales como la productividad y la citación.

En el primer número de *Newsletter* publicado por el Comité de Infometría de la Federación Internacional de Documentación, Rivicandra Rao consideró que “entre los diversos patrones de la bibliometría, los siguientes pueden ser estimados como significativos: Ley de Dispersión, Ley cuadrática inversa de la productividad científica o Ley de Lotka, Ley del mínimo esfuerzo o Ley de Zipf, Modelos logarítmicos y Proceso de la ventaja acumulativa de Price.

Teniendo en cuenta que el número de indicadores que puede emplearse para describir estos aspectos es enorme, a lo que habría que añadir las peculiaridades al referirse a autores personales, institucionales o editoriales, su enumeración se limita aquí a una descripción general de los empleados de manera específica. Por lo que la descripción realizada es de tipo general, remitiendo a referencias de manuales específicos.

Sancho (1990) ordena estos indicadores por su referencia a la calidad de la investigación establecida por jueces expertos, a la actividad o nivel de productividad científica de los autores, del impacto de su investigación mediante análisis de las citas; y de las asociaciones temáticas, por la citación, referencias y palabras comunes.

Por su parte, Callon, Courtial y Penan (1995) clasifican los análisis cuantitativos en dos categorías, atendiendo a si la referencia de los indicadores es la actividad o la relación entre los autores. Así, mientras los primeros proporcionan datos sobre el volumen e impacto de la investigación; los segundos rastrean los lazos e interacciones entre autores, instituciones y áreas de conocimiento.

Partiendo de estas clasificaciones y basándose en la ofrecida por Fernández Cano y Bueno (1998), se establece una priorización y ordenación de indicadores y leyes bibliométricas básica, ya experimentadas en trabajos previos (Expósito, 2000).

3.5.1. Indicadores de productividad.

Los indicadores de producción informan cuantitativamente sobre el tamaño de la producción científica de una determinada fuente institucional o personal. Aunque de forma general, puede referirse a la producción sobre una disciplina o tópico de investigación en un ámbito o periodo temporal concreto.

Los indicadores científicos de productividad pretenden medir la cantidad de investigación producida por los científicos. Suele referirse a la cantidad de publicaciones producidas por un autor, institución o país; aunque otras medidas incluyen la cantidad de investigadores por disciplina.

El empleo de informaciones sobre la producción suele partir de distribuciones de frecuencias y porcentajes empleando diversos tipos de intervalos. Si estos poseen naturaleza temporal, su estudio ofrecerá una visión diacrónica y si se limitan a un campo, ámbito o disciplina, una perspectiva sincrónica.

- *Productividad general:* Referida a cómputo del número y distribución de publicaciones de una disciplina, temática o ámbito geográfico es uno de los indicadores básicos en los estudios cientimétricos.

Aunque se ha encontrado correlación entre el número de publicaciones y la consistencia o calidad de ésta, el empleo exclusivo de indicadores cientimétricos puede limitar la consideración de su calidad (Bellavista *et al.* 1997).

- *Productividad de autores:* generalmente mensurada a partir del número de trabajos originales que publica. Para la aplicación de este indicador es necesario un número lo suficientemente completo de referencias y que éstas cubran un periodo diacrónico lo suficientemente amplio.

A partir de este indicador o asociados a él pueden aparecer otros que describan la colaboración y coautoría, ya sea por medio del establecimiento de promedios generales, estableciendo diferencias por género, lugares de firma.

Existen índices de proporcionalidad que relacionan el número de autores con el número de páginas del informe. Y también, que establecen los autores con una determinado porcentaje de la producción total considerada, como el índice de transitoriedad o de autores transitorios.

En este mismo sentido, el análisis de la productividad queda muy determinado por los supuestos teóricos establecidos en algunas leyes científicas, descritas a continuación, como la de crecimiento exponencial de Price o la Ley de Lotka.

- *Productividad institucional/Áreas de Conocimiento*: en los mismos términos que la referida a autores, pero partiendo del registro de la Institución o del Área de Conocimiento a la que éstos pertenecen.
- *Productividad de fuentes editoras*: los indicadores relativos al estudio de la productividad de las fuentes de publicación quedan muy determinados por su clasificación entorno a un tópico o disciplina. Suponiendo un principio teórico de relevante importancia la distribución propuesta por Bradford, referida al vehículo por excelencia de comunicación científica, la revista de investigación.

En general, todos estos indicadores e índices de productividad giran entorno a la idea de la hipótesis enunciada por Cole y Cole (1972), que atribuyen al filósofo Ortega y Gasset la producción/colaboración entre científicos. Está de acuerdo con un principio estratificado de la ciencia, opera con un criterio universal por el que los científicos que publican trabajos más significativos reciben el reconocimiento amplio que merecen, mientras que los que no publican trabajos de significación son ignorados.

Aunque también Turner y Chubin (1976, 1979) presentan una interpretación alternativa de tal hipótesis, que basada en una paradoja del Eclesiastés, asigna el 80% de la producción científica de los autores a sólo el 20% de éstos.

3.5.2. Indicadores de citación.

El análisis de citación suele referirse a patrones y frecuencias de las citas hechas y recibidas por autores, revistas, disciplinas o de las relaciones entre documentos citados.

En los estudios de citación se consideran dos supuestos básicos (Spinak, 1996): Que un artículo científico no es una entidad aislada, sino que está inmerso en una literatura sobre el tema del que trata, por lo que la obra del autor se construye sobre la obra de sus predecesores. Y que en el estudio del impacto de la investigación, a través de su nivel de citación, hay que tener en cuenta que estas citas se hacen no a la literatura usada sino a la publicada.

Este análisis, usualmente de la frecuencia y patrones a los que se ajustan las citas recibidas por autores, fuentes o disciplinas, es asumida por algunos autores como una rama de la propia bibliometría.

Su permite determinar el conjunto de autores o documentos que constituyen significativamente una disciplina, el promedio de referencias por documento, formas de la literatura citada (monografías, artículos, etc.) y la distribución geográfica o idiomática de los documentos citados, desde distribuciones diacrónicas o sincrónicas.

Algunos autores (Garfield, 1979) asumen el índice de citación como una caracterización de la excelencia del trabajo de un determinado autor, empleado incluso para la distinción de premios de gran prestigio internacional como los Premios Nobel.

Sin embargo, no faltan las críticas al análisis de la citación y al análisis cuantitativo general, fundamentas tanto en el desarrollo de métodos sin una investigación adecuada de sus supuestos (Brooks, 1985); como a problemas inherentes a citas incompletas, sesgadas o negativas (McRoberts y

McRoberts, 1989); a trabajos que rompiendo con paradigmas dominantes planteen cambios radicales sin la adecuada fundamentación (Lindsay, 1989), o a las dificultades derivadas del intento de operativizar las citaciones en sistemas de investigación no vinculados al ámbito anglo-americano (Fernández Cano y Bueno, 2002).

Pero tanto, críticos como proponentes de estos métodos están de acuerdo que el uso de indicadores, como la citación de un trabajo, no es aplicable para medir la calidad del trabajo de un individuo en particular. Ya que el trabajo científico como trabajo social, que describe hechos en no inmutablemente objetivos, mide su “calidad” en la medida que es importante para la comunidad científica y por lo tanto es “citable” por ella.

- *Citación general:* A partir del estudio de la distribución de citas se pueden establecer índices generales como promedio del número de citas, de su antigüedad o su variabilidad. Siendo, en ocasiones, de importancia relevante el análisis del desarrollo diacrónico de los mencionados índices.

- *Factores de impacto:* Referidos a la aceptación u uso de los hallazgos de un autor por otros, puede referirse a dos índices diferentes: Al que resulta del cociente de la cantidad de citaciones de un autor por el número de artículos que ha escrito. Y otro, más elaborado al factor de impacto de revistas.

En los mismos términos puede referirse a la importancia o impacto de una determinada fuente de publicación o revista dentro de una disciplina concreta al relacionar mediante su cociente, el número de citas que reciben las fuentes que se consideran núcleo de tal disciplina y la cantidad de artículos publicados por la fuente. Mediciones que pueden establecerse en periodos temporales arbitrarios.

- *Citación específica:* Estudio de citación por criterios/aspectos concretos como el tipo de fuentes citadas (libros, revistas, otros o literatura gris) o el idioma de las referencias a las que se citas.

Para ofrecer una visión descriptiva por medio de la metodología cientimétrica, además del empleo de unos indicadores básicos como los descritos, también se hace necesario comprobar el ajuste de la distribución de sus datos a los patrones teóricos ofrecidos en hipótesis, postulados y leyes básicos de tal metodología.

3.5.3. Ley de crecimiento exponencial de la literatura científica.

Dentro de los estudios sobre capacidad investigadora, existe una amplia tradición sobre medición de tal capacidad haciendo uso de la variable frecuencia de publicaciones o productividad (Fernández Cano, 1995).

Así, siempre ha sido convencionalmente aceptado un crecimiento exponencial de la producción científica, expresado como *Ley de desarrollo acelerado*, pero no es hasta 1956, en el que Price plasma este principio en la *Ley de Crecimiento Exponencial de la Ciencia*.

En ésta, Price constata que el nivel de crecimiento de la información científica es muy superior al de otros fenómenos sociales como pueden ser el de la población o nivel de su renta. De forma que, en un periodo temporal de entre diez y quince años tales informaciones se duplican.

Esta tasa se puede apreciar de manera similar en el número de revistas científicas, repertorios de resúmenes (ejemplificados en el gráfico 3.1), libros editados, bases de datos, número de investigadores y otros muchos indicadores de crecimiento científico.

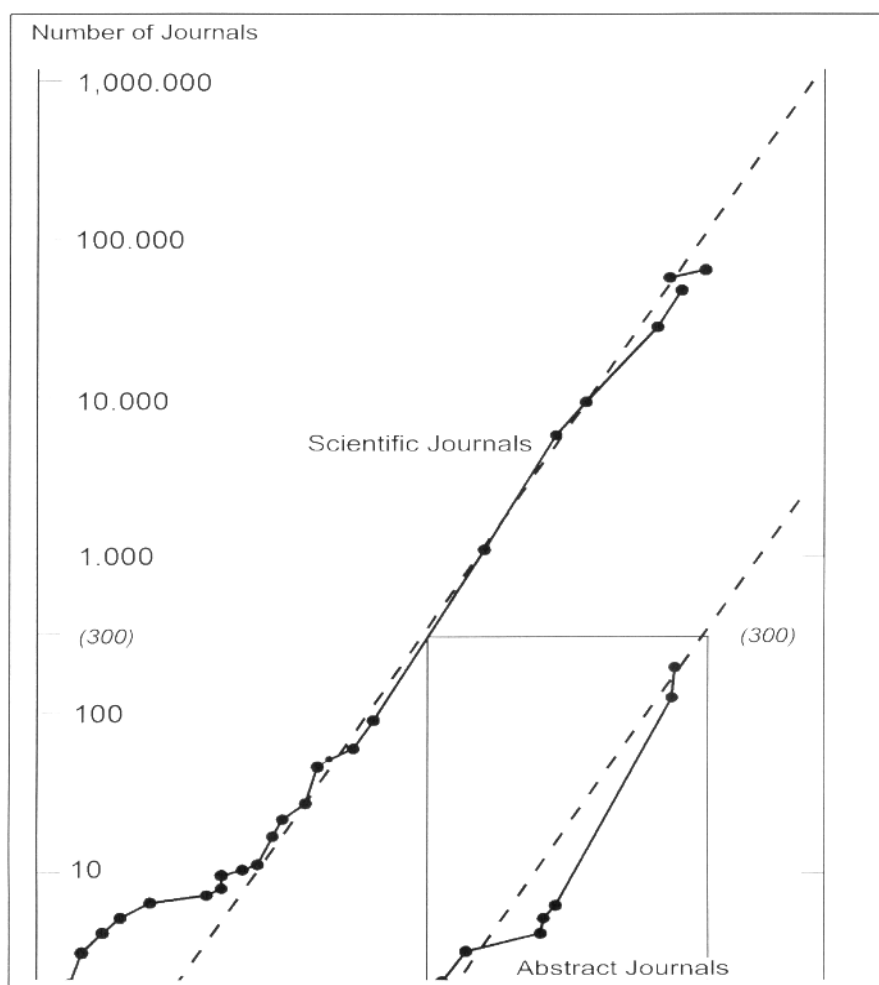


Gráfico 3.1. Número total de revistas científicas y de repertorio de resúmenes, en función de la fecha. Reproducido de D.J. de Solla Price (1965)

Sin embargo el nivel de crecimiento no puede ser indefinido sino que posee límites que Price explica relacionándolos con los recursos limitados de la sociedad para duplicar cada pocos años los recursos asignados a la ciencia. Así, como queda patente en el gráfico 3.2, la curva de crecimiento exponencial adquiere un punto de inflexión en el que se transforma un crecimiento logístico, más afín con la explicación de fenómenos naturales.

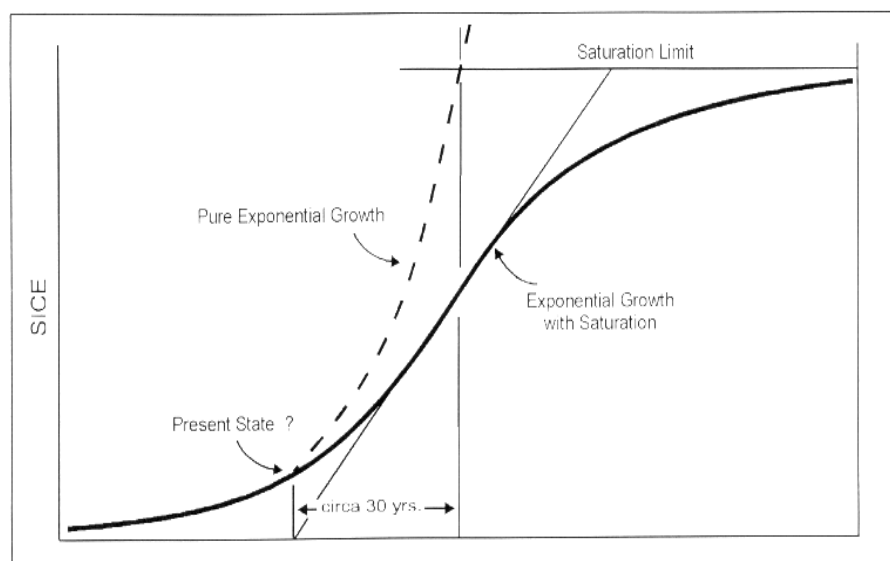


Gráfico 3.2. Forma General de la Curva Logística. Reproducido de D.J. de Solla Price (1965)

3.5.4. Ley de envejecimiento u obsolescencia de la literatura científica.

Mediante el estudio diacrónico de las referencias bibliográficas en distintas especialidades, Price observó que mientras el número de publicaciones se multiplica por dos aproximadamente cada 13,5 años, el número de citas que reciben tales publicaciones se divide por la mitad aproximadamente cada trece años.

Profundizando en la medida del envejecimiento, Burton y Kebler (en López Piñero, 1972) idean el concepto de semiperiodo, que referido al tiempo en que ha sido publicada la mitad de la literatura científica de una determinada disciplina, se define como la mediana de la distribución de frecuencias según el año de procedencia. Encontrando que la vida media de las referencias, en los años setenta, eran en Física 4,6; en Ingeniería Química 4,8; en Fisiología 7,2; en Matemáticas 10,5 y en Geología 11,8 años; diferenciaron entre literatura efímera y clásica.

Estos estudios de Price, basados en los primeros estudios sobre citación de Garfield, establecen diferencias en la citación a la literatura científica en una disciplina: Un 50% se distribuye entre la totalidad de la literatura anterior, y otro 50% en un número reducido de trabajos anteriores, producido por autores importantes en la disciplina o tópico investigativo (colegio invisible), de referencias que enlazan los artículos científicos de un año con los trabajos del colegio invisible, lo llamó Price frente de investigación.

Otra indicador típico de medida de la obsolescencia de la literatura científica es el *Índice de Price*¹³, que indica cuán recientes son las citas en un documento o disciplina, calculando la proporción del número de referencias que no poseen más de cinco años de antigüedad.

3.5.5. Leyes de dispersión de la literatura científica.

3.5.5.1. Ley de Bradford.

Formulada por Bradford en 1934 en un trabajo clásico, aunque con escaso impacto hasta que la publicó en su libro *Documentación* (1948).

Este autor plantea una preocupación por la duplicación de esfuerzos, y la insuficiencia de cobertura por parte de las agencias de indización y resúmenes, de los artículos de revistas científicas. Y en un esfuerzo de optimización, analiza la producción de artículos en Geofísica y Lubricación.

Mediante la representación gráfica semilogarítmica de las frecuencias de los artículos por revistas, comprueba la relación entre el logaritmo de la acumulación de éstas y la cantidad de artículos que se ajustaba a una recta, lo que le llevó a enunciar la siguiente formulación teórica o verbal:

¹³ No confundir con el índice de inmediatez de una revista científica, que se refiere al número de citas realizadas en año n a artículos publicados en este año, dividido por el número de artículos publicados sólo en ese año. (Para tasas de impacto por citación, véase Fernández Cano, 1995, pp.226-230).

Si las revistas científicas se ordenan en secuencia decreciente de productividad de artículos sobre un tema dado, éstas pueden dividirse en un núcleo dedicadas más en particular al tema y varias zonas concéntricas que con el mismo número de artículos que el núcleo presentan un número de revistas con relación 1: n : n² ...

(Bradford, 1934:75)

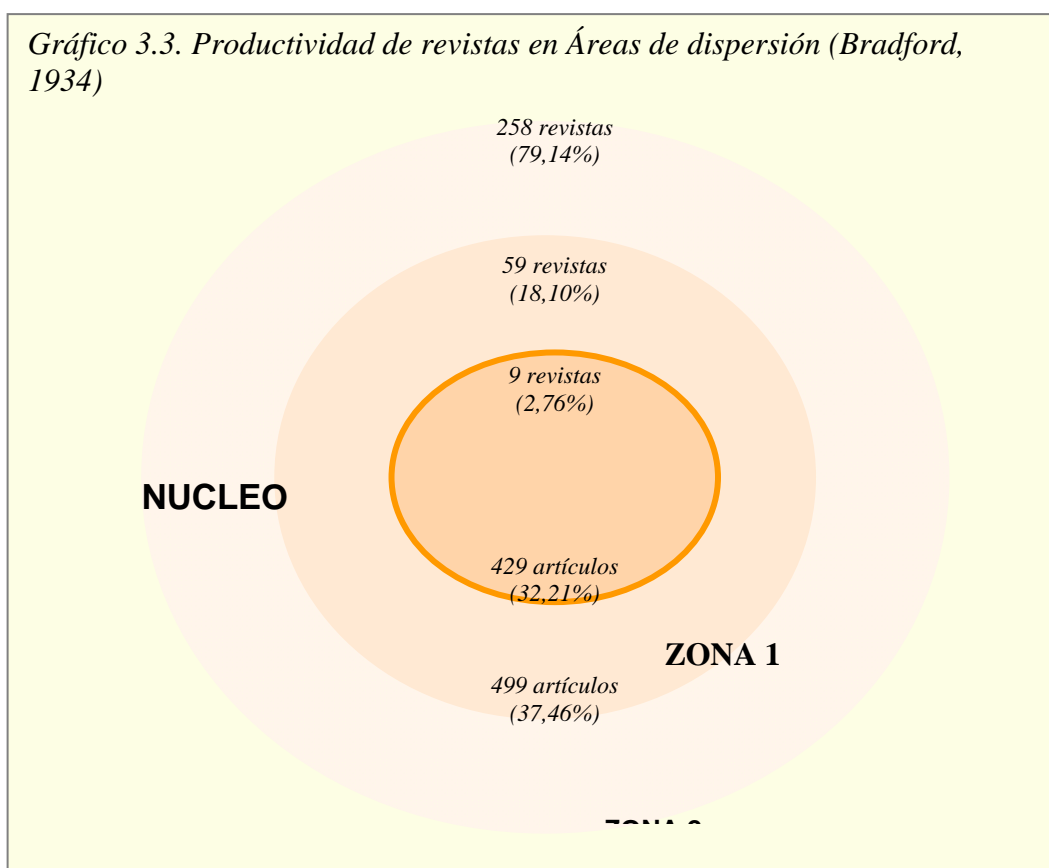
Esta formulación ha generado bastante literatura, pues existen ciertas incongruencias entre la formulación teórica y su expresión gráfica, que Bradford no advirtió.

Vickery (1948) comprobó tomando datos de 1600 publicaciones, que la parte rectilínea de la gráfica sólo se ajustaba a la parte superior de la distribución. Demostrándose en estudios posteriores que la distribución de Bradford no era logarítmica en toda su extensión, sino potencial en este núcleo y exponencial en las zonas exteriores. Incluso Ferreiro (1993) determina otros valores de composición de la distribución.

Estas desavenencias entre el modelo teórico o verbal y gráfico han generado una gran diversidad de modelos matemáticos, identificando Quiu en 1990 hasta veintidós. Algunos ofrecen más consistencia a la ley gráfica (Brookes, 1968 o Wilkinson, 1972) y otros a la teórica (Leimkuhler, 1967).

A principio de los sesenta, el matemático inglés Kendall demostró que la dispersión propuesta por Bradford es en realidad un caso particular de una distribución elaborada por Zipf (1935) que expresa cierta propiedad universal, inherente a cualquier lengua del mundo, conforme a la cual la parte más importante de cualquier texto queda conformada por unas pocas palabras de uso frecuente. Cualidad que algunos autores no sólo consideran aplicable a palabras, sino también a segmentos más amplios como artículos de revistas, por lo que la Ley de Bradford sería un caso particular de la Ley de Zipf.

En cualquier caso lo importante de las propuestas de Bradford es la idea de distribución o dispersión de fuentes por zonas de diferente intensidad productiva con respecto a un tema determinado.



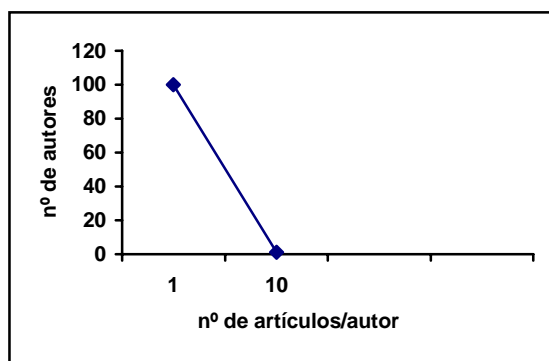
Garfield (1979) simplifica una ley elaborada por Bradford (1950), la *Ley de Bradford* en; las revistas ordenadas en orden decreciente de la calidad de los artículos (calculada ésta por la frecuencia de citación), sobre un tema dado, tales revistas podrían dividirse en un núcleo, que recibe la mayor parte de las citaciones, y varias zonas concéntricas que, conteniendo el mismo número de artículos del núcleo, siguen una sucesión de productividad ($1, 1/a, 1/a^2 \dots$), donde a se refiere al número de artículos.

3.5.5.2. Ley de Lotka

La cuantimetría ha generado ciertos patrones sobre productividad, uno de los cuales es la *Ley de Lotka*, que viene a decir que la productividad de autores sigue una ley cuadrática inversa, o lo que es lo mismo, que el número de autores que hacen n contribuciones es alrededor de $1/n^2$ de los que hacen una sola (Lotka, 1926).

Gráfica 3.4. Función gráfica de la Ley de Lotka, para 100 autores y tabular para 2000.

Tomada de Fernández Cano (1995, pp. 51-52).



Autores	Artículos/autor
Si 2000	1
500 (2000/2x2)	2
222 (2000/3x3)	3
125 (2000/4x4)	4
80 (2000/5x5)	5

Esta es una ley fundamental de la cuantimetría verificada en estudios de productividad en psicología y pedagogía (Budd, 1991), aunque no se tiene constancia de verificación en la variable *impacto por conteo de citas*.

Aunque en algunos estudios españoles (Blasco y Benavent, 1993) no se verifica, en general, en distribuciones amplias parece ajustarse bastante bien a datos empíricos de número de artículos publicados, excepto en sus extremos.

Su interpretación recuerda que la mayoría de la producción científica acerca de un tópico de investigación se realiza por un número significativamente escaso de autores consagrados como expertos en ese campo.

3.6. ANÁLISIS CIENTIMÉTRICOS EN ESTUDIOS EDUCATIVOS. UTILIDAD DE LOS INDICADORES MÁS USUALES Y DIFICULTADES DE EMPLEO.

Los estudios bibliométricos permiten construir visiones generales y sistemáticas de un campo, en nuestro caso de cómo se ha venido acometiendo en el contexto español la investigación acerca de la evaluación de programas educativos. Aunque no se debe sacralizar sus indicadores pues poseen abundantes limitaciones (Gómez y Bordons, 1996). Sin embargo, por su carácter descriptivo pueden ser útiles como estudios pre-evaluativos o como mecanismos complementarios y correctores de otras aproximaciones.

Las dificultades asociadas al empleo de este tipo de indicadores las sitúa Maltrás (1996) en la propia naturaleza de su origen, pues la literatura científica no tiene entre sus funciones prioritarias la evaluación externa de la ciencia. No careciendo la propia disciplina, en la realización de análisis estadísticos y sociométricos de la literatura científica, de puntos débiles asociados a una insuficiente base empírica; al empleo de métodos en ocasiones improvisados y poco precisos; explicaciones poco consistentes, por su superficialidad o por su osadía; y en general, por la audiencia de líneas de investigación lo suficientemente claras.

En España en los años sesenta y setenta la actitud de rechazo por parte de la comunidad científica hacia el uso de estos indicadores, Terrada y Piñero (1990) la relacionan con el desconocimiento de la actividad científica y objetivos de la Cienciometría y con una ideología poco evaluativa al considerar una pérdida de privilegios sociales del científico la revisión/evaluación de su trabajo. Y de forma muy importante por la asimilación de principios y valores ajenos, inculcados por el dominio científico anglosajón, que bloqueando la tradición científica propia parece dejar como opción una resignación al dominio y globalidad científica de las grandes potencias.

En el polo opuesto se sitúa una actitud realmente positiva hacia el análisis cientimétrico que permita el estudio “realmente objetivo” de la realidad científica, quizás con cierta ingenuidad acerca de la posibilidad de una ciencia absolutamente neutra.

Así, en los últimos años las ideas básicas de la Cienciometría han recibido múltiples críticas, y sus métodos se han sometido a continuos cambios, debido sobre todo a diversidad de trabajos sobre el tópico y a la ampliación de su base empírica. Siendo éstas características propias de una disciplina emergente, a la que más que criticar su valor se critica el uso inadecuado que puede hacerse de ella.

En cuanto a los estudios cientimétricos en investigación educativa realizados en España, la mayoría se refieren a investigaciones universitarias o al estudio de los artículos publicados en revistas determinadas, siendo la Orientación una de las disciplinas más atendidas. No existiendo ningún estudio referido a la evaluación de programas (Fernández Cano y Bueno, 1998).

De entre la gran diversidad de indicadores bibliométricos utilizados en Cienciometría, en nuestro contexto los más utilizados se refieren a la productividad, ya sea individual o institucional, y a la citación.

Algunas de las dificultades que podemos encontrar para realizar análisis cientimétricos en educación, además de las propias de la disciplina, se refieren a cuestiones acerca de la búsqueda, acceso y tratamiento de la información.

Las revistas de investigación educativa españolas no están indizadas en índices internacionales (SSCI, Eas y CIJE)¹⁴, siendo escasas las que poseen un prestigio científico internacional, y tan sólo a escala nacional.

¹⁴ S.S.C.I.: Social Sciences Citation Index (Índice de citación en Ciencias Sociales).
C.I.J.E.: Current Index to Journal in Education (Índice actualizado de Revistas Educativas).
E.A.: Educational Abstracts (Resúmenes de investigación educativa).

En este mismo sentido, tradicionalmente los trabajos de investigación educativos no han estado registrados en base de datos, o incipientemente lo han estado en algunas poco consistentes y diseminadas.

El registro de los datos e informaciones necesarios para elaborar estudio cuantitativos no es en ocasiones todo lo riguroso que debiera o puede encontrarse con las dificultades y sesgos como los señalados por Spinak (1996) en los análisis de citas:

- La auto citación puede ensombrecer los índices de citación en una disciplina, tanto como los sesgos y propensión de autores y revistas a citar a trabajos, idiomas o autores determinados (Campbell, 1990).
- La autoría múltiple, los homógrafos por el empleo de un solo apellido o iniciales y sinónimos dificultan el registro de este tipo de datos.
- Suelen darse enormes variaciones y errores en el registro de las referencias bibliográficas, incluso realizándose algunas de manera implícita o mediante el uso de epónimos.
- En ocasiones no existen o se siguen parámetros claros en el registro de informaciones, existiendo variaciones en la forma de registro según las disciplinas o los países (Murugesan y Moravssik, 1978; Irvine y Martin, 1985).

En general, la mayoría de problemas y dificultades se derivan de la escasa consideración del registro de la información como un valor científico importante (Fernández Cano, 2000). Y por lo tanto la realización de esta tarea de forma muy laxa.

3.7. EL ARTÍCULO DE REVISTA CIENTÍFICA COMO FUENTE DE INFORMACIÓN.

Una de las principales formas de comunicación de los hallazgos científicos a la comunidad es la publicación de un informe/artículo en una revista especializada. Formato de edición que adoptan la mayoría de informes de investigación identificados como posible muestra en esta Tesis, como se indica en las secciones dedicadas al estudio empírico, y que recuperados, son la principal fuente de datos.

El origen histórico de la publicación de revistas científicas hay que buscarlo en la necesidad de comunicación entre los científicos para mejorar tanto el desarrollo de sus trabajos como la difusión de éstos. Comunicación que con un origen epistolar derivó en una comunicación pública, organizada y formal asociadas al desarrollo de los colegios profesionales.

Este inicio gremial de las revistas ha hecho, en palabras de Crane (1972), que se tornen en un *lobby* epistemológico, símbolo de un colegio invisible de investigadores que comparten intereses comunes. Cuestión que puede ser enormemente positiva al permitir el desarrollo de disciplinas específicas; y no tanto, si endogámicamente limitan la publicación en aspectos temáticos, metodológicos o de autoría.

Las dos revistas más antiguas, que datan de 1665, fueron *Journal des Savants* de París¹⁵ y *Philosophical Transactions of the Royal Society of Medicine* de Edimburgo.

La funcionalidad de estas fuentes parte de la superación de las limitaciones que ofrecían las cartas, folletos y libros como vehículos de

¹⁵ “El Diario de los sabios”, ya recomendada por Feijoo en su Teatro Crítico Universal (1726-40/1988, p. 12).

comunicación científica, con perspectivas sociales y didácticas, que no epistémicas; aunque su uso como fuente de publicación de nuevos conocimientos es relativamente reciente.

En la actualidad, con una visión ciertamente papirocentrica, el punto final de cualquier investigación es su publicación en una revista científica, como medio de reivindicación de la propiedad intelectual de los trabajos. Hecho, que condicionado y mediado usualmente por una revisión previa de expertos o revisión por pares, es tomado como una declaración y reconocimiento de la comunidad científica de la originalidad y valor del trabajo publicado.

Esta condición, que transformó estas fuentes de un artefacto social en una técnica de acumulación de la información, ha planteado que la motivación inicial para leer sus contenidos parezca, como plantea Fernández Cano (1995b) menor que la motivación para publicar en ellas.

Pero para publicar en una fuente determinada, además de la revisión del contenido de los informes/artículos, se requiere la adopción de unas condiciones, estructura de los trabajos, y adecuación de las características del trabajo a aspectos propios de la fuente de publicación en la que se editan, establecidas por los Consejos Editoriales de éstas.

Este aspecto, que en cierto sentido puede procurar una cierta homogeneidad de formatos que facilitarían su revisión y registro en trabajos de revisión; a su vez provocan una serie de limitaciones relacionadas con el estilo de publicación en el que se tiende en numerosas ocasiones a la realización de síntesis, por el limitado espacio ofrecido a cada trabajo, que puede dificultar la apreciación de matices y cuestiones fundamentales para la comprensión de las investigaciones desde una perspectiva global.

En algunos casos, la necesidad de publicación como reconocimiento de la labor científica, como en la investigación educativa y en el ámbito español, presenta serias dificultades por la escasez de fuentes con un reconocido prestigio internacional.

