

9.3. RESULTADOS SOBRE CITACIÓN EN LA INVESTIGACIÓN SOBRE E.P.

El cuarto objetivo cuantitativo del estudio referido a la descripción de las pautas de citación en los trabajos revisados se ha cumplido en tanto a los datos reflejados en el capítulo seis y discutidos aquí de forma general.

9.3.1. Desarrollo diacrónico de la citación.

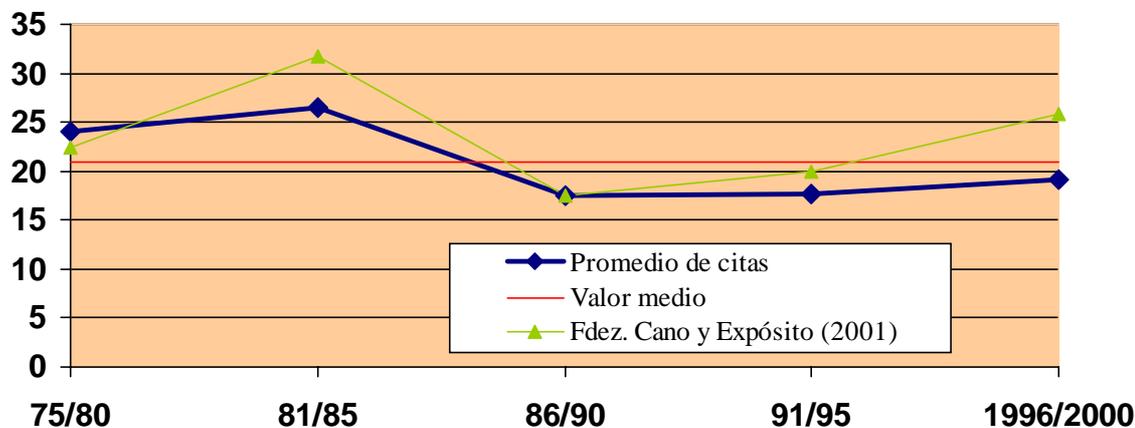
Los datos obtenidos muestran la citación de 3821 trabajos para los 208 considerados como muestra, lo que supone un promedio de 18,37 citas por trabajo.

Este valor es cercano al establecido en estudios similares (Fernández Cano y Expósito, 2001) que arrojan valores medios entorno a 21 citas por trabajo. Situando la revisión de la publicación de la *Revista de Investigación Educativa* (Bueno, 2001), fuente de publicación nuclear en la investigación sobre E.P., el promedio de citas desde el año 1983 al 2000 entre 21 y 27,5 por trabajo.

El desarrollo diacrónico del patrón del promedio de citaciones realizadas adopta una distribución típica de crestas y senos en la primera mitad del periodo diacrónico considerado (1775/1987), pero en la segunda presenta unos valores delimitados en el intervalo que abarca 13 a 27 citas realizadas, estabilizándose cada vez más entorno al valor medio cercano a 21 citas en cada trabajo.

El patrón que detectó Garfield (1994) y corrobora Fernández Cano (1995, p. 224) muestra un progresivo aumento en el número de citas según pasan los años.

Gráfico 9.3. Comparación del desarrollo diacrónico del promedio de citación de datos experimentales y estudios afines.



9.3.2. Antigüedad de la citación.

El estudio del promedio de la antigüedad de las citas se obtiene un valor global de 15,23 años, propio según el índice de Garfield (1983) de las disciplinas de corte humanístico. Muy semejante a la obtenida por Torralbo (2001) de 12,8 años.

Sin embargo, en otros estudios afines, como Bueno (2001) o Fernández Cano y Expósito (2001) se obtienen valores mucho más compacto entorno a 8,5 años.

La alta variabilidad de la antigüedad con respecto a la media alcanza algo más de 26 años. Este alto valor queda sin duda determinado por la inclusión en la citación en los estudios sobre E.P. a estudios clásicos de la Investigación Educativa en cuanto a su cierta antigüedad, que aunque sólo se empleen para obtener una perspectiva diacrónica sobre ciertos aspectos concretos hacen envejecer la antigüedad general de la citación.

La asidua citación a los clásicos en una disciplina, aún en los trabajos de corte más empírico y actual, puede estar fundamentada en diversos criterios. El primero de ellos sería la poca actividad global en la materia que obligaría a citar a escasos y antiguos trabajos, no siendo éste el caso como ha quedado patente en la discusión previa de datos sobre la producción. Otra de las posibles razones, más plausible por su mejor adecuación a los datos aquí reflejados, es la que relaciona la citación con el individualismo productivo, tanto a escala personal como institucional y departamental. Este aislamiento originaría un cierto desconocimiento de los trabajos más rabiosamente actuales y una dependencia endémica a la citación clásica.

Así mismo, pese a que la E.P. es un área de conocimiento claramente delimitada y en un periodo de madurez productiva, todavía adolece de ciertos rasgos de las disciplinas emergentes como puede ser cierta “inconsistencia disciplinar”. En la que la falta de ciertos criterios evaluativos estándares hace necesaria una reconstitución o replanteamiento histórico en cada trabajo, ofreciendo la citación a clásicos un ámbito de seguridad epistemológica no considerado para las realizaciones más noveles.

Atendiendo al porcentaje de la antigüedad de las citas en los intervalos propuestos por Garfield (1979) para la caracterización de las disciplinas científicas, y comparando los datos obtenidos en esta Tesis Doctoral (datos experimentales) con los ofrecidos en trabajos similares y considerados aquí como criterios comparativos para la discusión de resultados, se obtienen los siguientes valores.

Tanto en los valores experimentales de esta Tesis, como en su comparación con un trabajo de las mismas características (Fernández Cano y Expósito, 2001) referido a E.P. y los obtenidos de la publicación de Investigaciones educativas en la Revista de Investigación Educativa se obtiene valores muy semejantes, propios de las Ciencias Sociales. Que mediante la

adaptación del índice de Price, permite delimitar la proporción de citas con una antigüedad menor de cinco años que es tan sólo el 19,1 % de todas las realizadas.

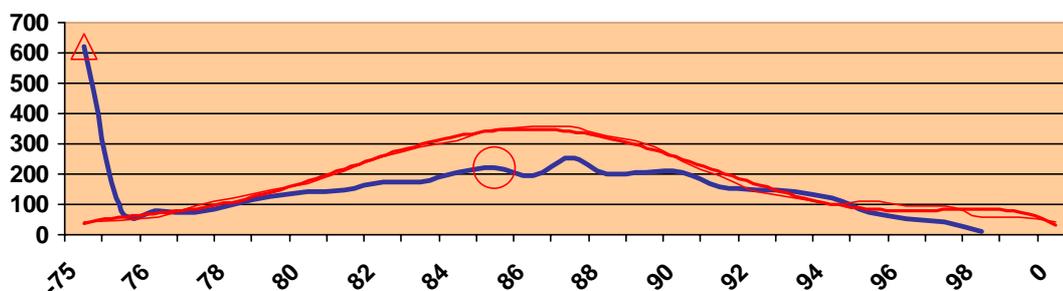
Tabla 9.5. Antigüedad de las citas.

Caracterización disciplinar (Garfield, 1979)	0/5 años	6/10 años	+10 años
	Ciencias duras	Ciencias Sociales	Humanidades
	%		
DATOS EXPERIMENTALES	19,1	55,6	25,3
Bueno (2001)	15,29	53,6	31,07
Fernández Cano y Expósito (2001)	17,5	54	28,5

Profundizando en la discusión de la antigüedad de las citas realizadas en los trabajos revisados y a tendiendo a los años en los que se publicaron los trabajos citados, se puede comprobar una distribución en la que se destacan dos aspectos. El primero se refiere a la gran citación a clásicos publicados con anterioridad a 1975, primer margen del intervalo diacrónico considerado. Y el segundo su ajuste a la normalidad en una distribución en que los valores más habituales se acumulan en los valores diacrónicos medios.

Esta disposición parece lógica puesto que el valor medio del intervalo diacrónico está en 12,5 años (es decir quedaría entre 1986 y 1987) y la moda de antigüedad es el año 1987.

Gráfico 9.4. Frecuencias de los años de edición de las citas.



Es decir, se siguen demostrando los parámetros de citación mencionados anteriormente, de citación a clásicos en E.P. y un promedio antigüedad entorno a diez años. Sin embargo, en un pormenorizado estudio de la antigüedad lo largo del intervalo diacrónico 1975-2000, se comprueba una disposición a citar cada vez a un mayor porcentaje de trabajos con cierta actualidad, siempre dentro de la década de producción del propio trabajo. Esto supondría en valores numéricos una tímida disminución de la antigüedad promedio por debajo de diez años.

Acogiéndose al postulado de Garfield (1979) en cuanto a la caracterización de las disciplinas científicas por la antigüedad de las citaciones, en investigación sobre Evaluación de Programas se estaría produciendo un “fortalecimiento disciplinar” o “tecnificación” en cuanto a su constitución como campo de investigación con características propias.

Sin embargo, la citación de literatura “obsoleta”, mayormente libros en lugar de revistas, es un hallazgo preocupante sobre la baja calidad científica en este campo de investigación.

9.3.3. Citación por fuentes de edición y lengua empleada.

Tanto por los datos obtenidos en este trabajo de revisión como los declarados en trabajos afines, se puede observar que con porcentajes muy similares los trabajos más citados en la investigación sobre E.P. son los libros y manuales (51,3 %), seguidos por la revistas (34,3 %) y en último lugar con un porcentaje menor (13,8 %) la literatura gris.

Tabla 9.6. Fuentes de edición de las citas.

TIPO DE FUENTES DE EDICIÓN		% Libros	% Revistas	% Lit. Gris
DATOS EXPERIMENTALES	Citación	51,3	34,3	13,8
	Productividad	1,92	87,98	10,1
Bueno (2001)		61,5	35,5	3
Torralbo (2001)		54	33,6	12,4
Fernández Cano y Expósito (2001)		49	37	14

Sin embargo, los patrones de citación más deseable para una dinámica línea de investigación en una disciplina consistente deberían estar encabezados por las revistas especializadas, principal fuente de publicación de informes de investigación teórico-práctica. Seguidos por los libros de texto, más dedicados a las grandes revisiones o fundamentaciones teóricas. Y en último lugar a la literatura gris, que permite una operativa comunicación de los últimos y más novedosos hallazgos de carácter eminentemente práctico.

El cumplimiento de este patrón deseable en los datos experimentales referidos a la productividad pero no en los de citación, apoya la caracterización de la investigación sobre E.P. como una línea con un patrón productivo típico de las líneas calientes de investigación, pero con unos hábitos de citación propios de áreas con poca claridad disciplinar.

La caracterización de la citación también responde, tanto en los datos experimentales como comparativos de trabajos afines, a patrones de citación fundamentalmente a trabajos realizados en inglés, seguidos por los realizados en español con un 9% menos. Siendo destacable la escasa citación a trabajos en otras lenguas, un escaso 5%, referidas en la distribución de datos al francés y sobre todo lenguas co-oficiales en las diversas Comunidades Autónomas.

Tabla 9.7. Idioma de edición de las citas.

IDIOMA DE LAS CITAS	% Español	% Inglés	% Otros
DATOS DE EXPERIMENTALES	42,9	51,5	5,2
Bueno (2001)	42,5	55	3,5
Torrallbo (2001)	38,5	53,7	8,3
Fernández Cano y Expósito (2001)	39	57	4

La interpretación de estos valores lleva a considerar dos aspectos concretos en la investigación sobre E.P.:

El primero de ellos, como en la mayoría de disciplinas científicas, la preponderancia del uso de la lengua inglesa como principal canal de comunicación de los hallazgos científicos (*lingua franca*). Especialmente en una era de globalización, en la que la actividad científica, como cualquier otra actividad social, queda inmersa en una rápida y necesaria dinámica de hipercomunicación.

Esto lleva a la obligación del empleo del inglés para la comunicación internacional de los hallazgos de la producción nacional, habida cuenta que además la revistas de investigación más potentes en el ámbito internacional, en la mayoría de disciplinas, están editadas en este idioma.

En segundo lugar, conjugando el empleo mayoritario de la citación a trabajos en lengua inglesa con la realizada a libros, denota una cierta dependencia o colonialismo de citación con respecto al mundo anglosajón, en los principios teóricos y epistemológicos de la investigación sobre E.P., inferencia que probablemente pudiera ampliarse a la totalidad de la Investigación Educativa.

9.3.4. Citación personal o de autores.

En el total de los 208 trabajos, tomados como muestra y productores de datos experimentales, se comprueba que se citan un total de 3821 trabajos, lo que supone un promedio cercano a la realización de 20 citas en cada trabajo revisado.

A su vez, en estos 3821 trabajos se referencian nominalmente a un total de 6336 autores, que determinan un promedio de autoría en los trabajos citados de 1,7 autores por trabajo. Resultando este valor muy compatible con el índice de colaboración detectado en los 208 trabajos considerados como muestra y que ascendía a un promedio de 2,1 autores por trabajo.

Ambos datos, que sumarían una considerable muestra compuesta por más de cuatro mil elementos, permite establecer este índice de colaboración con una alta fiabilidad.

La discusión de los datos ofrecidos por tal cantidad de información se inicia con una clasificación que aunque no posee una base firme y se basa en criterios cuasi discrecionales, es usualmente empleada como indicador en los estudios de revisión cuantitativa (Fernández Cano y Bueno, 1999), por la facilidad de interpretación que ofrece.

La citación recae en su mayor medida (57,8 %) sobre autores que sólo reciben una o dos citas, que podrían calificarse como autores eventuales o transitorios en la materia, siendo muy escaso el porcentaje de autores que recibe más de diez citas. Aspectos que hacen que la citación se ajuste al patrón teórico propuesto en la Ley de Lotka, que deja la citación en manos de unos pocos autores.

Tabla 9.8. Tabla de citación.

CITAS EN...	Pequeños citados (1-2 citas)	Medianos citados (3-9 citas)	Grandes citados (+10 citas)
DATOS DE EXPERIMENTALES (Missing: 2,1)	57,8	26,9	10,5
Expósito (2000)	93,15	6,61	0,24

M.E.C.	39
Feuerstein, R.	28
Gargallo, B./ Novak, J.D.	25
Sussman, S.	22
De Ketele, J.M.	21
Prieto Sánchez, M.D.	20
Fernández Ballesteros, R./ Rivas, F./ Sternberg, R.J.	19
Rand, Y./ Vygotski, L.S.	18
Hoffman, M.D./ Piaget, J.	17
Stufflebeam, D.L.	16
Bloom, B.S./ Calafat, A. / Kagan, J.	15
Bruner, J.S./ Meichenbaum, D.H./ Shinkield, A.J./	14
Trianes, M.V./ Witkin, H.A.	13
Driesen, G./ Hernández, P./ Luria, A.R./ O.M.S./	12
Álvarez Rojo, V./ Campbell, D.T./ Matson, J.L./Pérez Juste, R.	11
Arnáiz Sánchez, P./Bandura, A./Castillejo, J.L./Colás Bravo, M.P./Cook, Th.D./Dueñas, M.L./Garaigordobil, M./Pelechano, V./Posner, G./Sanz Oro, R./Slavin, R.E./Tejedor, F./Verdugo, M.A./Wehman, P.H.	10

Por tanto se puede afirmar que en la investigación sobre E.P. no existen autores de “obligada citación” o clásicos en el campo. Y como ha quedado reflejado con anterioridad, al destacarse la lengua inglesa como la principalmente citada, entre los diez primeros autores más citados tan solo se encuentran cuatro españoles.

A modo de conclusión, tras el análisis de datos cuantitativos realizado, la producción científica sobre Evaluación de Programas Educativos en el periodo que abarca desde el año 1975 hasta el 2000 queda caracterizada fundamentalmente por los aspectos que se resumen en la siguiente tabla 9.9.

Tabla 9.9. Resumen de hallazgos cuantitativos de la investigación sobre E.P.

PRODUCCIÓN	Desarrollo diacrónico	<i>Patrón de crecimiento en momento logístico</i>
	Índice de colaboración	<i>2,2 autores/trabajo</i>
	Género y lugar de firma	<i>No sesgos en la producción.</i>
	Instituciones productoras	<i>Sistema educativo: Universidades.</i>
	Área productora	<i>Dpto. M.I.D.E.</i>
	Fuentes de publicación	<i>Revistas científicas: R.I.E.</i>
CITACIÓN	Promedio citas	<i>21 citas/trabajo en ligero crecimiento</i>
	Promedio antigüedad / variabilidad	<i>15,23 años / 26,62 años</i>
	Fuentes más citadas	<i>Libros</i>
	Idioma más citado	<i>Inglés</i>
	Autores	<i>Mayoría transitorios, no especialistas.</i>

Hallazgos que permiten realizar las siguientes inferencias evaluativas, como queda reflejado en la tabla 9.10, que caracteriza, desde la perspectiva cuantitativa, a la investigación española sobre evaluación de programas educativos, como un campo disciplinar productivo y prometedor, aunque adolece de ciertas dificultades comunes en la investigación educativa, como son la baja tasa de colaboración entre autores e instituciones, la escasez de especialización y la ausencia de características propias de la ciencia puntera en cuanto a citación e impacto de los hallazgos con fuentes de prestigio internacional.

Tabla 9.10. Resumen de inferencias evaluativas tras ajuste a patrones cientimétricos de la investigación española sobre E.P.E. (según modelo de Fernández Cano, 2002).

PATRÓN	AJUSTE	VALORACIÓN
1. Crecimiento diacrónico.	Logístico.	+
2. Grado de colaboración interautores.	2,1 autores/estudio.	-
3. Tipologías de autores.	No grandes productores, exceso de ocasionalidad.	-
4. Sesgo en la producción por género.	No diferencias.	+
5. Grado de colaboración interautores.	Nulo.	-
6. Ajuste productividad institucional a la Ley de Bradford	Débil, multiplicativo no exponencial.	-
7. Fuente de publicación predominante	Revistas.	+
8. Ámbito de publicación de informes.	Mayoritariamente Nacional.	-
9. Ajuste de fuentes de a la Ley de Bradford	Débil, multiplicativo no exponencial.	-
10. Antigüedad de las citas.	15,2 años.	-
11. Fuente de citación predominante.	Libro.	-
12. Idioma mayoritario en citación.	Inglés.	-
13. Nacionalidad predominante de autores más citados.	Extranjera.	-
14. Verificación Ley de Lotka a autores.	Verificada.	+

Se establecen tres tipos de valoraciones: +(alta), =(media) y -(baja), en base al ajuste que denota.

9.4. RESULTADOS SOBRE ASPECTOS CONCEPTUALES EN LA INVESTIGACIÓN SOBRE E.P.

Como se presenta en el capítulo siete de este Tesis Doctoral referido al análisis de datos conceptuales, éstos quedan delimitados por la longitud y forma/contenido de sus títulos, las palabras claves que describen los procesos evaluativos, las áreas de contenidos que se establecen en las implementaciones de los programas y los soportes empleados para su desarrollo.

El análisis de la longitud de los títulos ofrece unos datos que muestran un moderado aumento de la longitud de los títulos en el periodo diacrónico considerado y su estabilización entorno a los 91 caracteres. Siguiendo las hipótesis de White (1981), que relacionan la longitud de los títulos con la complejidad de la materia a la que se refieren, la investigación sobre E.P. en el transcurso de este periodo, habría aumentado su complejidad como disciplina científica, encontrándose en un momento de estabilización.

En el mismo sentido los datos referentes a la presencia en los títulos del colon perfecto (representados gráficamente por los dos puntos), que relacionan su presencia con la complejidad de los títulos en los que se insertan (Dillon, 1980 y Perry, 1985), ofrecen un desarrollo similar, con un moderado crecimiento y periodo de nivelación entorno a valores medios.

Esta estabilización de la complejidad disciplinar es compatible con los datos referidos a productividad y citación, que ofrecen valores desde los que puede inferirse una caracterización semejante del desarrollo y del momento en el que se encuentra este campo de conocimiento en la actualidad, en crecimiento logístico.

Estableciéndose una cierta continuidad en el proceso de análisis, que permita el establecimiento de relaciones entre variables. A partir del análisis de contenido de los títulos, además de comprobar los aspectos a los que se refieren éstos, se delimitan las categorías que servirán para clasificar las palabras clave.

La mayor parte de los elementos constitutivos de los títulos analizados se refiere a aspectos relacionados con la necesidad de desarrollar e implementar los programas (30, 59%), seguido por aspectos relativos a los procesos de implementación y metodología de las evaluaciones (24, 81%). En menor grado al área de conocimiento a la que se refiere el programa (16, 79%), al contexto en el que se implementa (16, 23%) y a los destinatarios o usuarios (11, 56 %).

Es decir, los títulos hacen sobre todo referencia a la necesidad del desarrollo de los programas y a aspectos sobre el proceso y metodología de implementación/evaluación. Esto indica por una parte la permanente necesidad de justificar la intervención por medio de programas y la necesidad de su evaluación, y por otra, la importancia que se ofrece a los criterios técnicos de ésta.

Hay que considerar que la mayoría de la producción registrada está elaborada por autores pertenecientes a Centros Universitarios, principalmente asociados a Áreas o Departamentos de Métodos de Investigación y publicados en la Revista de Investigación Educativa. Rasgos que pueden determinar elaboraciones más preocupadas por justificar su intervención y desarrollar aspectos técnicos de ésta, que por consideraciones más prácticas relacionadas con el contexto o los usuarios de los programas evaluados.

El registro de las **palabras clave** que describen las investigaciones en las categorías desarrolladas a partir del análisis de contenido de los títulos (Ver epígrafe 7.2.2) permite apoyar y delimitar este sistema, ofreciéndole más amplio significado a cada una de éstas.

Con un proceso de análisis similar se establecen las **áreas de conocimiento** a las que se refieren las actuaciones desarrolladas durante las implementaciones de los programas. Clasificación que a su vez se emplea como sistema de agrupamiento de los contenidos desarrollados en tales procesos.

La mayor parte de los programas evaluados se refieren a disciplinas relacionadas con la Educación para la Salud, Consumo y Medio Ambiente (18,3%), Orientación ya sea de aspectos educativos, vocacionales y laborales (13,9%), psicoeducativa (13%) o familiar (12%) y Didáctica para la atención a las Necesidades Educativas Especiales (12%).

Estas áreas consideradas, determinan una agenda de investigación sobre E.P. muy afín a las líneas generales establecidas en la Investigación Educativa, determinadas por un momento de clara preocupación por la calidad del sistema educativo, reflejado incluso en el marco legal actual. En este sentido, la evaluación poseería el impacto sobre las decisiones políticas, señalado por (Carballo, 1992), jugando un papel primordial en los procesos de optimización (De la Orden, 2000) y la evaluación de programas una exigencia ineludible para la búsqueda progresiva y permanente de una mejor educación.

La revisión realizada por Velaz de Medrano *et. al.* (1996) sobre las investigaciones realizadas por el C.I.D.E. durante diez años, se referían en un 73% a la evaluación de programas, dedicando un 60% del total a la aplicación de programas curriculares de ámbito restringido, sobre áreas como las indicadas.

Por tanto, la investigación sobre E.P. pese a su caracterización con producciones frecuentemente de rasgos técnicos, es sin duda un ámbito de la investigación educativa que por su propia naturaleza aplicada, se encuentra muy conectado a las necesidades educativas y realidades sociales.

Paradójicamente, en estos nuevos ámbitos de intervención por medio de programas, los soportes materiales para desarrollarlos/evaluarlos siguen

respondiendo a formatos tradicionales. La mayoría se basa en procesos sin formato específico como clases ordinarias, conferencias o seminarios; seguidos en frecuencia por los tradicionales basados en el lápiz y papel. Siendo muy escasas las realizaciones apoyadas en formatos electrónicos y basadas en las nuevas tecnologías.

De forma general se pueden resumir la discusión de datos conceptuales por los valores más habituales en cada una de las variables consideradas, como se muestra en la siguiente tabla 9.10.

Tabla 9.11. Resumen de hallazgos conceptuales de la investigación sobre E.P.E.

TÍTULOS	Longitud y colon perfecto.	<i>Entorno a 91 caracteres: Aumento de la complejidad de la disciplina en momento de estabilización</i>
	Contenidos	<i>Sobre necesidad de programas y metodologías de la evaluación.</i>
PALABRAS CLAVE Sobre...		<input type="checkbox"/> Necesidad de implementación. <input type="checkbox"/> Metodología de la evaluación. <input type="checkbox"/> Contextos de aplicación <input type="checkbox"/> Destinatarios/usuarios <input type="checkbox"/> Áreas de Conocimiento
ÁREAS usuales	CONTENIDOS	
	Ed. Salud, Consumo y Medio Ambiente	<input type="checkbox"/> Prevención de drogodependencias <input type="checkbox"/> Planificación familiar
	Orientación	<input type="checkbox"/> Conocimiento del medio académico, social y laboral. <input type="checkbox"/> Educación para la carrera. <input type="checkbox"/> Educación social/familiar.
	Didácticas para la atención a N.E.E.	<input type="checkbox"/> Desarrollo de habilidades psicoeducativas. <input type="checkbox"/> Resolución de problemas.
SOPORTES	<i>Habitualmente sin especificar, el material más empleado tipo lápiz y papel.</i>	

9.12. Resumen de inferencias evaluativas tras ajuste a patrones conceptuales de la investigación española sobre E.P.E.

PATRÓN	AJUSTE	VALORACIÓN
1. Longitud de los títulos	Estabilizado	=
2. Complejidad de los títulos	Medio	=
3. Diversidad conceptual de los títulos	Limitado	-
4. Soportes.	Tradicional	-
5. Diversidad conceptual en palabras clave.	Alta	+

Se establecen tres tipos de valoraciones: +(alta), = (media) y – (baja), en base al ajuste que denota.

9.5. RESULTADOS DE ASPECTOS METODOLÓGICOS EN LAS INVESTIGACIONES SOBRE E.P.

La discusión de los datos de carácter metodológico de los estudios revisados en esta Tesis Doctoral se realiza en los mismos ámbitos en los que se agrupan las variables en el primer epígrafe del capítulo cinco.

9.5.1. Ámbitos de fundamentación teórica.

Las investigaciones sobre evaluación de programas se fundamentan mayoritariamente en el paradigma científico en más de un 56 % de los casos, ocupando un porcentaje menor (25%) aquellas que se fundamentan en posiciones humanístico interpretativas. El paradigma crítico es escasamente empleado, puesto que tan sólo ocupa el 5,3% de los casos.

El empleo de aproximaciones científicas determina sin duda, como señalan Arnal *et. al.* (1993), procesos de evaluación más viables para evaluar efectos finales y cuantificables, que para mejorar el desarrollo de la implementación de los programas.

Los ámbitos concretos de fundamentación teórica se reparten equitativamente con porcentajes similares entre las teorías que intentan explicar el ajuste social, las teorías de aprendizaje, las de desarrollo psicológico y las relacionadas con el desarrollo didáctico.

La distribución de los ámbitos de fundamentación teórica se dispersa en pequeños porcentajes entre diversas teorías específicas. Las más usuales (+10%) son el cognitivismo y las teorías didácticas, las referidas a la atención a las N.E.E y las que intentan explicar el ajuste y desarrollo sociofamiliar, afectivo y de valores. (Ver tabla 8.2 sobre Ámbitos de fundamentación teórica).

Aunque el paradigma científico es el más empleado, una amplia mayoría de evaluaciones, que alcanza el 63 %, no señala explícitamente las hipótesis de la investigación en su informe.

En cuanto a la metodología, la más ampliamente utilizada (61,1%) es de corte cuantitativo, que según los planteamientos de Pérez Juste (1997), es más apropiada para la obtención de visiones generales profundas. Quedando porcentajes mucho menores para la mixta (23,1%) y cualitativa (15,9%), más acorde con la evaluación de elementos contextuales.

El empleo prioritario de metodologías cuantitativas también es denotado por Álvarez Rojo (1993) refiriéndose a procesos evaluativos en el campo de la Orientación.

El porcentaje obtenido por las metodologías de corte mixto o ecléctico no cumple los principios que señalan Pérez Carbonell *et. al.* (1999), que abogan por el empleo y vigorización de estas aproximaciones, para conseguir objetivos múltiples y facilitar la triangulación. Aspectos a desarrollar en la investigación sobre E.P.

Estos rasgos de los ámbitos de fundamentación de las evaluaciones revisadas, determinan el uso mayoritario (46,2 %) de los modelos objetivistas para la evaluación de programas, especialmente por el aporte de frecuencia realizado por el modelo orientado a objetivos. Obteniendo idéntico porcentaje (21, 6%) las aproximaciones subjetivistas y eclécticas, y un escaso uso las de corte crítico (5, 8%).

Con cierta lógica de ajuste a los modelos más empleados, puesto que éstos ponen más énfasis en la evaluación de unas u otras facetas del programa, las más evaluadas se refieren a la consecución de los objetivos propuestos en el programa y reflejados en su plan de evaluación.

Pero pese a la importancia del uso de modelos científicos, son escasos los estudios en los que el establecimiento de hipótesis, propio de éstos, se realiza de forma explícita y justificada. Corriendo enormes riesgos evaluativos al darlas por supuestas o asumirlas tácitamente.

El análisis de las facetas evaluadas en los procesos investigativos, arroja la disposición en tres modelos diferenciados: Uno internalista, de corte cualitativo, cuyo interés se centra en la contextualización de entradas y procesos. Otro con una disposición más externalista, que fundamentada en el conductismo, evalúa las entradas y productos. Y por último, otro con un enfoque trascendentalista, con un interés focalizado en la evaluación de los impactos de los programas.

Los productos considerados/resultantes de estos procesos evaluativos se refieren sobre todo al desempeño escolar y académico de los individuos, especialmente a la variable rendimiento; a su promoción personal individual; y en menor grado a su ajuste social.

De forma general, los principios de fundamentación teórica de la investigación sobre evaluación de programas se asientan en modelos que Guba (1978) llama convencionales y naturalísticos, y que se podrían calificar de cierto conservadurismo epistemológico. Existiendo además cierta coherencia, Arnal *et. al.* (1992), entre los distintos elementos que constituyen los aspectos de fundamentación teórica.

En trabajos como el de Torralbo (1991), de revisión de investigaciones sobre enseñanza de las matemáticas, en un formato eminentemente academicista como es el de Tesis Doctorales y diferente de los revisados aquí, aunque también se basan en principios cognitivistas, declaran el mayor empleo del paradigma interpretativo humanístico y el uso más habitual de aproximaciones metodológicas de corte cualitativo y/o mixto.

9.5.2. Instrumentos para la recogida de información.

En las investigaciones revisadas la mayoría de las fuentes de información son directas, recogiendo los datos por medio de instrumentos no estandarizados (48,1 %) frente a los estandarizados (31, 25 %), entre los que destacan los cuestionarios.

Estos datos son congruentes a los obtenidos por Torralbo (2001) que obtiene un 54,1 % y 35, 6% de instrumentos no estandarizados y estandarizados respectivamente.

Los criterios para establecer su validez y fiabilidad no son informaciones que se declaren usualmente en los informes revisados, mayoritariamente por tratarse de instrumentos estandarizados en los que se asumen éstos o por establecerse genéricamente sin hacer especificación alguna. Así, aunque con porcentajes muy bajos de menos del 5% del total de la distribución, el criterio de validez más empleado es el constructo y el de validez la consistencia interna, que Velaz de Medrano *et. al.* (1996) refieren a aspectos relacionados con la minuciosidad de su elaboración.

Se establece un índice de 0,5 criterios de validez por instrumento considerado. Datos que ofrecen una perspectiva de la instrumentación en la investigación sobre E.P. que algunos autores como García Ramos (1992) califican de cierta debilidad, al faltar instrumentos con suficientes garantías de validez y fiabilidad.

Comparativamente en la revisión de investigaciones sobre enseñanza de las matemáticas, Torralbo (1991), la mayoría de instrumentos no son estandarizados sino de elaboración *ad hoc* para cada trabajo, basando su validez en criterios de contenido y su fiabilidad en su consistencia interna.

9.5.3. Población y muestra.

En las investigaciones revisadas, no parece ofrecerse importancia a la descripción de las poblaciones a las que se generalizarían los hallazgos, ya que en su mayoría (44,2 %) tan sólo se indica ésta de forma general.

Los procesos de selección de las muestras se suelen basar en procedimientos no aleatorios, sobre todo por conveniencia e intencionalidad, aunque también es habitual el empleo de procesos basados en conglomerados no aleatorizados.

Estas características son propias de la implementación/evaluación de programas sobre grupos previamente conformados, como son las clases y agrupamientos típicos del sistema educativo. Realizándose actuaciones específicas sobre poblaciones suficientemente caracterizadas, como para no tener que profundizar en su descripción.

Precisamente este origen poblacional determina que en la mayoría de las muestras ni siquiera se considere el género o posean composición mixta, carácter muy asumido en el Sistema Educativo público. Muestras que suelen tener intervalos de edad de 7 a 6 ó 13 a 18, con niveles consecuentes en Educación Primaria y Secundaria.

Los tamaños muestrales más usuales están entre 100 y 1000 unidades, cantidad muy acorde con el empleo mayoritario de cuestionarios como instrumentos de recogida de informaciones.

Todos estos datos sobre las poblaciones y criterios de selección de la muestra son muy compatibles con los ofrecidos por Torralbo (2001) en un campo distinto al de la Evaluación de Programas, con lo que con las debidas precauciones, no parece del todo inusual en toda la investigación educativa.

Todos estos aspectos determinan el planteamiento de dudas sobre la verdadera representatividad de las muestras, en los estudios revisados.

9.5.4. Diseños de investigación.

Los diseños empleados en E.P. son muy diversos destacando en la muestra considerada los denominados prefijados, sobre todo cuasiexperimentales con dos grupos (grupo de control equivalente) y pretest-posttest (26, 9%), y descriptivos tipo encuesta (20,7 %). Poseyendo este tipo de diseños una disposición temporal de tipo longitudinal.

El carácter de fundamentación científico de las investigaciones y el desarrollo de diseños de tipo casi experimental determinan unas condiciones evaluativas en las que el evaluador asume un rol externo al programa, como asesor de soluciones sobre su desarrollo y juez de sus resultados.

El empleo usual de diseños de tipo de corte descriptivo es igualmente congruente con los instrumentos más utilizados en la recogida de informaciones, que como se indicaba anteriormente se realiza por medio de cuestionarios, muy empleados en este tipo de estudios.

La intervención por programas y la evaluación de estos se suele realizar en contextos sociales de enorme complejidad a escala experimental, por ello abundan realizaciones en cuyo diseño no siempre es posible el control total de todas las variables, como los de tipo cuasiexperimental o descriptivos (Tejedor, 2000). Este principio, congruente con los datos obtenidos, obliga en cierto sentido a intentar controlar estas variables estableciendo diversos mecanismos para ello. Sin embargo, según estos mismos datos, en más del 82 % de los trabajos revisados no se hace referencia alguna a amenazas a su diseño de investigación, ni por lo tanto a medidas para controlarlas.

Estas características en cuanto a diseños de investigación, aunque puedan ser calificadas de cierta debilidad, son propias de la Investigación Social. En la disciplina concreta de Evaluación de Programas no está exenta de polémica, como muestran la sucesión de trabajos en ciertas publicaciones como *Educational Researcher*, para la que diversos autores (Slavin, 1997; Klein, 1997; Pogrow 1998 y 1999) que proponen diseños de investigación/evaluación en los que son muy pertinentes los grupos de control, y que por medio de su comparación en implementaciones prácticas “en competencia” y su revisión por expertos, permita delimitar cual ofrece mejores resultados en cada caso concreto y pueda ser considerado como un “diseño ejemplar”.

El abuso del empleo de diseños de corte descriptivo y el debate global sobre la debilidad de los diseños, pone en tela de juicio el planteamiento de cualquier relación causal establecida a partir de los hallazgos encontrados en algunos trabajos sobre investigación educativa.

Al igual que en otros aspectos ya tratados, la caracterización de los diseños de investigación en E.P. es similar a los valores obtenidos por Torralbo (2001) para las Tesis Doctorales sobre educación matemática.

9.5.5. Aspectos cronológicos y de implementación de los programas.

Las investigaciones sobre E.P. consideradas proponen seguimiento evaluativos sobre todo y exclusivamente durante la implementación de los programas evaluados (30,3%), siendo más escasas aquellas en las que éste seguimiento del programa se realiza durante y posteriormente a su implementación (22,6 %). El porcentaje de evaluaciones que sólo realiza un seguimiento a posteriori, es decir con un carácter eminentemente sumativo, alcanza un escaso 8,7%.

La descripción de los aspectos metodológicos de las evaluaciones descritos hasta aquí y el seguimiento que éstas hacen sobre todo durante la implementación del programa indica un rol de los evaluadores, como asesores externos mayormente procedentes de Centros Universitarios, más que implicados socialmente en el contexto de desarrollo/evaluación del programa, como experimentadores técnicos de procesos en los que confirmar sus teorías.

La duración específica de las actuaciones de implementación/evaluación, no es un dato de relevancia en los informes revisados, pues en más de un 58% de los casos no se ofrece esta información. Destacando entre los que sí lo hacen, las implementaciones puntuales medidas en un número determinado de horas y menores a cursos completos.

9.5.6. Tratamiento y análisis de datos.

Las unidades de análisis más empleadas en la recogida de informaciones, en las que destacaban las fuentes directas, son los sujetos implicados en la implementación de los programas (usuarios, padres o profesores) en un 75,5% de los casos, frente a las unidades grupales como Centros o Instituciones o los propios programas. Este origen o fuente de datos es normal ya que los productos más considerados en las evaluaciones estaban relacionados precisamente con la variable rendimiento de éstos.

La mayoría de las evaluaciones poseen un carácter descriptivo, por lo que el uso de estadísticos con esta naturaleza es habitual, empleándose un promedio de 1,7 estadísticos por evaluación. El empleo es usual y compensado, en porcentajes similares, tanto para los distribucionales, como los de tendencia central o variabilidad.

Estadísticos inferenciales se usan con un promedio inferior de 1,4 por trabajo, siendo en un 45% de tipo paramétrico, usualmente ANOVA simple y t de Student.

Sin embargo los estadísticos correlacionales y multivariados, así como los análisis cualitativos son poco usuales ya que entorno al 83% de las evaluaciones revisadas no se emplean. En aquellos casos que sí se utilizan, entre los primeros destaca el Coeficiente de Correlación r Pearson, y entre los segundos, el análisis de contenido.

El uso de criterios o estándares externos de comparación, que permitan valorar los resultados obtenidos, es muy poco usual; pues en más del 89% de los casos no se realiza. Esta característica de las evaluaciones denota por una parte la ausencia de líneas comunes que permitan criterios de comparación interoperativos entre trabajos similares; el escaso impacto de las evaluaciones; o la consideración de estas tareas investigativas, denotada en datos anteriores, de tareas solitarias personal e institucionalmente.

En el mismo sentido la triangulación no se emplea en más del 57 % de los casos. En aquellos que sí se realiza, se emplea una triangulación con un carácter que Denzin (1989) llama técnico como es la de instrumentos, frente a otros como el de las fuentes de información. Carácter técnico que de forma global parece que se asume en la investigación sobre E.P.

9.5.7. Decisiones tras la evaluación.

Considerando la evaluación como un proceso cuyo sentido es fundamentalmente la emisión de juicios sobre el programa evaluado para mejorarlo, como plantean Cook y Sadish (1986) en su teoría del valor de la evaluación, debe existir cierta coherencia, en el sentido que le atribuye Arnal *et. al.* (1992) entre los procesos de planteamiento y desarrollo del proceso evaluativo y las decisiones tomadas a partir de las consideraciones realizadas en él. Para ello se realiza la discusión conjunta de aspectos relativos a resultados de la evaluación, continuidad del programa y decisiones de modificación.

A partir de estas informaciones combinadas se determina que la situación más usual es aquella en la que se declaran resultados positivos tras la implementación del programa, pero no se registran informaciones sobre su modificación ni sobre su continuidad (31, 3%). Seguida por aquella en que tras establecer unos resultados positivos, se deciden realizar modificaciones de mejora para continuar implementando el programa en el futuro. (28,4 %). Resultando el resto de combinaciones con porcentajes mucho más bajos.

Si bien esta segunda postura parece lógica tras el esfuerzo desarrollado para poner en marcha un proceso de investigación evaluativa. La primera de ellas confirma la caracterización de la investigación desde una perspectiva técnica externa, desarrollada con una relación poco real con el programa y los individuos implicados en él y con una consideración global experimental como proceso de prueba de ciertos modos y procesos de intervención.

Resulta curioso, aunque confirma este mismo planteamiento, que tan sólo un 2,4% y 1,4% declaren resultados negativos o nulos de los programas. Y comparativamente alto el porcentaje de casos en el que no se ofrece información alguna sobre resultados, continuidad o modificación del programa evaluado.

De forma general y atendiendo a los valores obtenidos con mayor frecuencia, los hallazgos encontrados en la investigación española sobre E.P. pueden resumirse en los aspectos señalados en la siguiente tabla 9.13 y las inferencias desarrolladas a partir de éstos hallazgos en la tabla 9.14.

Tabla 9.13. Resumen discusión de hallazgos metodológicos de la investigación sobre E.P.E.

FUNDAMENTACIÓN TEÓRICA	Paradigma.	<i>Científico(56,3%).</i>	
	Ámbitos de fundamentación.	<i>Ajuste social. Teorías del aprendizaje (cognitivismo)</i>	
	Hipótesis.	<i>No se explicitan</i>	
	Metodología.	<i>Cuantitativa (61%).</i>	
	Modelos evaluativos.	<i>Basados en objetivos (39%).</i>	
	Productos de la evaluación.	<i>Desempeño escolar, variable rendimiento.</i>	
INSTRUMENTOS DE MEDIDA	Estandarización.	<i>No(48,1%).</i>	
	Tipo.	<i>Cuestionarios(28,8%).</i>	
	Validez y fiabilidad.	<i>No se indica (42,8%).</i>	
POBLACIÓN Y MUESTRA	Población.	<i>Sólo indica (44,2%).</i>	
	Selección muestra.	<i>No aleatoria (47,9%).</i>	
DISEÑO DE INVESTIGACIÓN	Diseños.	<i>Descriptivos (20,7%.) Cuasiexperimentales (26,9%).</i>	
	Denominación.	<input type="checkbox"/> <i>Tipo encuesta o survey</i> <input type="checkbox"/> <i>Pretest posttest con G.E. Y G.C. equivalente.</i>	
	Temporalización.	<i>Longitudinal (63%).</i>	
	Agentes evaluadores.	<i>Externos (53,4%).</i>	
	Amenazas a la validez del diseño y control.	<i>No (82,7%).</i>	
	Seguimiento evaluativo.	<i>Durante su implementación (30,3%)</i>	
ASPECTOS CRONOLÓGICOS	Duración.	<i>Intervalos de horas (25,5%).</i>	
	Unidades de análisis.	<i>Sujetos</i>	
TRATAMIENTO Y ANÁLISIS DE DATOS	Estadísticos descriptivos.	<input type="checkbox"/> <i>Distribucionales: Porcentajes</i> <input type="checkbox"/> <i>Tendencia Central: Media</i> <input type="checkbox"/> <i>Variabilidad: Desviación Típica</i>	
	Estadísticos correlacionales y multivariados.	<i>Poco usuales</i>	
	Estadísticos inferenciales.	<i>Paramétricos: ANOVA y t de Student.</i>	
	Análisis cualitativo.	<i>Poco usual</i>	
	Valores y estándares externos de comparación.	<i>No</i>	
	Triangulación.	<i>Poco usuales</i>	
	Resultados.	<i>Positivos (80,8%)</i>	
DECISIONES TRAS LA EVALUACIÓN	Continuidad.	<i>No ofrece información/continuar (52,4/46,2%)</i>	
	Modificación.	<i>No ofrece información/ modificar (54,80/44,2%)</i>	

9.14. Resumen de inferencias evaluativas tras ajuste a patrones metodológicos de la investigación española sobre E.P.E. (según modelo de Fernández Cano, 2002a y 2002b)

PATRÓN DESEABLE	AJUSTE	VALORACIÓN
1. Diversidad paradigmática y búsqueda de complementarismo.	<i>Predominio enfoque científico.</i>	-
2. Unificación teórica, búsqueda de teorías comprensivas.	<i>Teorías unívocas.</i>	-
3. Hacia una metodología de la evaluación ecléctica.	<i>Predominio cuantitativa.</i>	-
4. Conjunción de modelos evaluativos sobre un mismo programa.	<i>Predominio modelo por objetivos.</i>	-
5. Enunciado, justificación e inducción de hipótesis.	<i>Bajo.</i>	-
6. Grado de preocupación por validez de los instrumentos.	<i>Escaso (9%).</i>	-
7. Grado de preocupación por la fiabilidad de los instrumentos.	<i>Escaso (7%).</i>	-
8. Uso de muestras representativas.	<i>Escaso.</i>	-
9. Descripción exhaustiva de la población.	<i>Baja.</i>	-
10. Uso de diseños consistentes.	<i>Bajo.</i>	-
11. Preocupación por el control de amenazas a la validez de los diseños de investigación.	<i>Baja.</i>	-
12. Uso de evaluadores múltiples y desde diversos posicionamientos.	<i>Escaso.</i>	-
13. Seguimiento post hoc tras la implementación.	<i>Medio.</i>	=
14. Tiempo de implementación.	<i>Reducido.</i>	-
15. Uso de procedimientos analíticos alternativos (T.E., porcentaje de varianza explicada, etc.)	<i>Nulo.</i>	-
16. Grado de sofisticación de procedimientos analíticos (multivariados, factoriales, etc.)	<i>Bajo.</i>	-
17. Uso de criterios o estándares externos de comparación.	<i>Bajo.</i>	-
18. Presencia de procesos de triangulación.	<i>Baja.</i>	-
19. Continuidad del programa.	<i>Baja.</i>	-
20. Modificación del programa.	<i>Escasas.</i>	-
Se establecen tres tipos de valoraciones: +(alta), = (media) y - (baja), en base al ajuste que denota.		

