

Expósito López, J., Olmedo Moreno, E. y Fernández-Cano, A. (2004). Patrones metodológicos en la investigación española sobre evaluación de programas educativos. *RELIEVE*, v. 10, n. 2, p. 185-209.
http://www.uv.es/RELIEVE/v10n2/RELIEVEv10n2_2.htm



**Revista ELectrónica de Investigación
y EValuación Educativa**

PATRONES METODOLÓGICOS EN LA INVESTIGACIÓN ESPAÑOLA SOBRE EVALUACIÓN DE PROGRAMAS EDUCATIVOS

*(Methodological patterns in the spanish research on evaluation of
educational programs)*

por

[Article record](#)

[About authors](#)

[HTML format](#)

Jorge Expósito López (jorgeel@ugr.es)
Eva Olmedo Moreno (emolmedo@ugr.es)
Antonio Fernández-Cano (afcano@ugr.es)

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

This study tries to obtaining some inferences about the field of Program Evaluation, synthesising some patterns about research methodologies used by them. Concretely, foundation epistemological paradigm, specific design and its validity threats, the evaluation model, sample techniques, instruments for data collection, data analysis techniques, the evaluative agents and other different methodological characteristics in use. To this respect, we have obtained indicators about 208 studies published in Spain from 1975 year to 2000, mainly papers from educational scientific journals. These studies were achieved from retrieval in several data base and completed with other crossed-references.

Keywords

Program-Evaluation, Educational Programs, Research Methodology, Spanish Production, Research Analysis, Scientometric Review.

Resumen

Este estudio pretende realizar inferencias sobre el campo de la Evaluación de Programas Educativos (*E.P.E.*), sintetizando ciertos patrones relativos a la metodología de sus investigaciones. En concreto, sobre el paradigma de fundamentación epistemológica, la metodología de investigación específica, el modelo evaluativo empleado, las técnicas de muestreo, los instrumentos empleados para la recogida de información y las técnicas de análisis de datos, el diseños de investigación, las amenazas a su validez, los agentes evaluadores y otras características metodológicas consideradas. Para ello, se revisaron 208 trabajos realizados y/o publicados en España desde el año 1975 al 2000, publicados la mayoría como artículos en revistas científicas especializadas. Obtenidos a partir de una búsqueda en diversas bases de datos, complementada con una búsqueda manual de referencias cruzadas

Descriptoros

Evaluación de programas, programas educativos, metodología de la investigación, producción española, análisis de la investigación, revisión cienciométrica.

INTRODUCCIÓN

La investigación sobre evaluación de programas educativos hay que considerarla co-

mo parte constituyente de la investigación educativa general, concretamente de aquella que genéricamente se ha venido denominando investigación evaluativa (Alvira, 1991).

Como queda reflejado en el desarrollo normativo, que vertebra el sistema educativo español, existe desde la *Ley General de Educación de 1970* una preocupación creciente por la evaluación, la investigación evaluativa y la evaluación de programas. Preocupación, constatada por expertos como Pérez Juste (2000), que determina que en la actualidad en cualquier ámbito educativo, no puede entenderse una intervención sin un programa claro y detallado. Ni éste, sin una evaluación que permita su valoración para la mejora.

Partiendo, por tanto, de la necesaria intervención educativa por programas como mecanismo de comprensión profunda para la mejora (Álvarez Rojo, 1994), la Evaluación de Programas Educativos (desde ahora abreviadamente: *E.P.E.*) se ha de convertir, al igual que la programación de sus acciones, en una actividad básica de todo profesional de la educación.

Por ello, se hace necesaria la apertura de este amplio y complejo campo de la *E.P.E.* a los intereses marcados por todos los miembros constituyentes de las comunidades educativas, ampliando las temáticas a tratar o estableciendo metodologías y procedimientos complementaristas, que permitan un posicionamiento más científico de los educadores, tanto en su fundamentación teórica como en su práctica derivada.

Aún así, en 1997 y 1998, diversos autores consideran la *E.P.E.* un campo de investigación en plena vigencia, pero todavía con limitaciones propias de disciplinas incipientes o en proceso de reestructuración (Slavin, 1997; Klein, 1997; Pogrow, 1998).

Siendo precisamente, en etapas de replanteamiento y estructuración de un campo de indagación, muy considerados y necesarios los estudios de síntesis de cualquier tipo;

como queda patente en la dedicación de estudios similares con más de veinticinco años, como los realizados por Scriven (1974), Stufflebeam (1974 y 1978) o Cook y Gruder (1978). Y más recientemente, de forma contextualizada al caso español, el completado por De la Orden y Martínez de Toda (1992).

Dentro de la metodología de síntesis, disponemos del método cuantitativo, que permite la obtención de visiones globales sobre un campo, empleando una perspectiva eminentemente descriptiva y realizando un esfuerzo de estructuración del mismo.

Los estudios de síntesis cuantitativa ofrecen, pues, un aumento de la comprensión dinámica de un campo de investigación. En el caso de la *E.P.E.* española, como muestra el proceso de revisión de la literatura, no existen trabajos disponibles como el planteado aquí, de ahí la necesidad de estudios de este tipo.

Es precisamente el propósito del presente trabajo ofrecer una visión comprensiva global de este campo, sin llegar a realizar explicaciones profundas en cuanto a posibles variaciones en base a modelos evaluativos y/o paradigmas de evaluación. Informaciones que delimitan una interesante línea de investigación a continuar en futuros trabajos.

2. ESTUDIO EMPÍRICO.

2.1. Revisión de la literatura y pertinencia del estudio.

El proceso de revisión de la literatura muestra la existencia de estudios de síntesis sobre las tendencias de la investigación educativa española en general (Pérez Carbonell *et. al.*, 1999; Fernández Cano, 1997; Fernández *et. al.*, 2002), sobre gestión y organización de la investigación pedagógica (Pineda, 1985; Colás *et al.*, 1995) o análisis de ámbitos temáticos y metodológicos de la producción investigadora en algunas disciplinas educativas (Álvarez Rojo, 1993 y 1997).

Por otro lado, estudios de síntesis cuantitativa de la producción realizada por centros o departamentos de educación (Colás, García y Rodríguez, 1995; Bartolomé y Sancho, 1997), por áreas de conocimiento (Galante y Sanchidrián, 1996; Benayas, 1997; Fernández Cano *et. al.*, 2000; Fernández Cano y Expósito, 2001; Expósito y Fernández Cano, 2002), o por su formato o fuente de publicación (Bueno, 2001; Torralbo, 2001).

Sin embargo, se comprueba una escasez de realizaciones referida a la *E.P.E.*, ya señalada por Álvarez Rojo y Hernández (1997), que permite vislumbrar algunas de las consideraciones a tener en cuenta en la pertinencia de la realización de este trabajo. Residiendo ésta además de su pertinencia ocasional, en la situación del propio campo disciplinar que algunos autores, como (Fernández Cano y Bueno, 1998), califican de asentamiento y/o replanteamiento.

Parece por tanto justificada la indagación específica sobre el tópico de la *E.P.E.*, dentro de un contexto bien acotado, como es el caso de la investigación educativa española.

2.2. Problema de investigación.

La cuestión a indagar en este estudio se centra en la caracterización de la investigación sobre la *E.P.E.* en el ámbito español en un periodo diacrónico que abarca desde el año 1975 al 2000, en sus aspectos de fundamentación epistemológica, metodología de investigación empleada, modelo evaluativo empleado, las técnicas de muestreo, los instrumentos empleados para la recogida de información y las técnicas de análisis de datos, el diseño de investigación y las amenazas a su validez, los agentes evaluadores y otras características metodológicas relevantes.

2.3. Diseño de investigación.

Para el desarrollo empírico de este estudio se recurre a un diseño eminentemente descriptivo, con una cierta pretensión censal

limitada por los problemas derivados de acceso a la totalidad de la población. A través de un análisis de documentos, que informan de evaluaciones concretas de programas específicos, se deriva una síntesis de la investigación sobre *E.P.E.* en el ámbito y periodos indicados.

2.4. Población y muestra.

La unidad básica de análisis es el informe de evaluación de un programa educativo. Realizándose una búsqueda en diversas bases de datos (*I.S.O.C.*, *Psyclit*, *Biblioteca Nacional*, *E.R.I.C.* y *S.S.C.I.*) por medio de descriptores como “investigación + evaluación + programa + educación `español””. Búsqueda que, complementada con una búsqueda manual, arrojó un total de 1798 registros.

De estos registros, en base a la información contenida en su título y/o resumen, se seleccionaron 300 mediante un procedimiento triangulado de revisión de expertos. La concordancia entre revisores estuvo por encima del 90%, resolviéndose las discrepancias conjuntamente o acudiendo a un tercer revisor.

Para ser seleccionado, un informe, debía cumplir las siguientes condiciones:

- Referirse a programas educativos (en cualquier ámbito).
- Estar implementado o publicados en España.
- Las evaluaciones debían estar referidas a programas implementados total o parcialmente.
- Los procesos de investigación debían poseer unas características mínimas de rigor científico; a saber: introducción, método, análisis de datos y ciertas decisiones/conclusiones justificadas sobre el programa.

Tras la recuperación de los mencionados informes se seleccionaron, al responder plenamente a las características establecidas, mediante un proceso de revisión de su contenido en profundidad, un total de 208 que

configuraron la muestra productora de datos. Informes mayoritariamente (n >93%) publicados como artículos en revistas científicas españolas del campo de la educación.

2.5. Instrumento y análisis de datos.

Para la recogida de datos se construye un instrumento por medio del programa *Microsoft Excel*, que además de facilitar el registro y su gestión, permite la realización de análisis preliminares de datos a nivel descriptivo.

Las categorías de registro consideradas se obtienen en el informe o mediante inferencia manifiesta tras la lectura del texto. Considerándose las reflejadas en la siguiente tabla I.

La validez de este instrumento queda reforzada en su contenido, mediante la valoración de expertos sobre la representatividad de los campos que lo componen, y por su uso, al formar parte de un instrumento usado en trabajos similares (Expósito, 2000; Torralbo, 2001). E incluso, en su factibilidad para el registro, al ensayarlo previamente en un taller de prácticas con alumnos de 5º curso de Pedagogía.

En cuanto a su fiabilidad, utilizándolo en un estudio piloto con el 10% de la muestra, se obtienen valores de estabilidad de las adscripciones al emplearlo en dos ocasiones diferentes, y de concordancia entre dos recolectores superior al 90 % de los registros.

Tabla 1. Categorías de registro consideradas.

INDICADOR		CATEGORÍAS ADMISIBLES
1. Paradigma.		Científico/Interpretativo/Crítico/Complementarista
2. Metodología.		Cuantitativa/Cualitativa/Mixta
3. Modelo evaluativo.		Objetivistas/Subjetivistas/Críticos/Eclecticos
4. Población y Muestra.	4.1. Descripción de la población.	Indica/Describe/No Indica o No describe.
	4.2. Técnicas de muestreo.	Abierta, para categorización a posteriori.
	4.3. Edad y nivel académico.	Intervalos y categorización a posteriori.
5. Instrumentos de recogida de datos.		Abierta, para categorización a posteriori.
6. Técnicas de análisis de datos.		Abierta, para categorización a posteriori.
7. Diseños de investigación.		Abierta, para categorización a posteriori.
8. Amenazas a la validez del diseño.		Plantea amenazas y su control/ Plantea amenazas, pero no su control/No plantea amenazas, ni su control.
9. Agentes evaluadores		Externos/ Implementadores/ Usuarios/ Mixtos.

2.6. Análisis de resultados.

Los análisis de datos realizados sobre los paradigmas de fundamentación, metodología y modelos empleados en los procesos evaluativos, se basan en el empleo de estadísticos de corte descriptivo, como las frecuencias y porcentajes, y su

representación tabulada y gráfica.

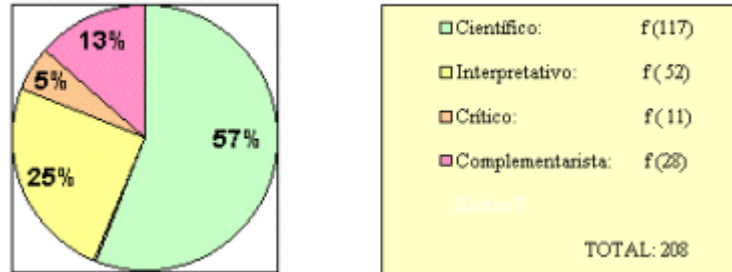
2.6.1. Paradigmas y/o enfoque de fundamentación.

Siguiendo los criterios de categorización, se obtienen las frecuencias y porcentajes de

los paradigmas y/o enfoques de fundamentación general de los distintos trabajos que se

muestran en la siguiente representación gráfica I.

Gráfico I. Frecuencias y porcentajes de paradigmas de fundamentación



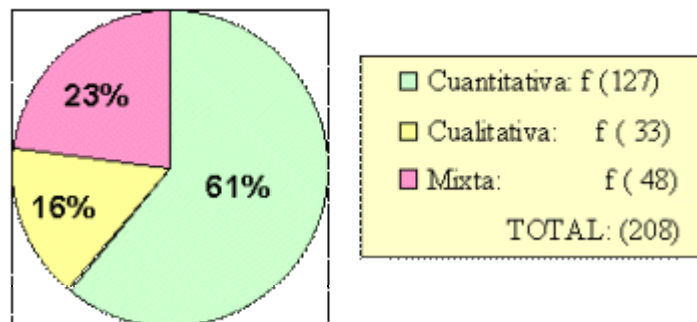
Una mayoría absoluta, del 57%, informan de la fundamentación de evaluaciones en el paradigma científico. A continuación, pero con una frecuencia de prácticamente la mitad y un porcentaje del 25%, se encuentran los estudios con un carácter eminentemente interpretativo humanísticos. Teniendo de nuevo que dividir por la mitad la frecuencia, para acercarse al 13,5% de los estudios fundamentados en paradigmas con planteamientos de tipo complementarista o eclécticos.

Para finalizar con los basados en desarrollos socio-críticos que tan sólo son 11 de los 208 trabajos revisados, ofreciendo un escaso 5,3% del total de ellos.

2.6.2. Metodologías empleadas.

De los 208 estudios considerados 127 emplean una metodología cuantitativa, lo que supone una amplia mayoría de más del 61 % (Ver Gráfico 2).

Gráfico 2. Frecuencias y porcentajes de la metodología empleada



Con frecuencias menores de la mitad, se encuentran los que emplean consideraciones metodológicas de carácter sincrético o mixto, que suponen un 23%; y los que recurren a técnicas cualitativas que tan sólo suponen, como muestra la representación gráfica, algo más del 15%.

Para el análisis de datos que corresponden a esta variable, se parte de los modelos para la evaluación de programas, especificados en el apartado en el que se describen las variables de este estudio. A partir de esta clasificación de carácter teórico, se comprueba el ajuste de la distribución de los datos obtenidos, para la posible categorización de los resultados, como muestra la Tabla II.

2.6.3. Modelos evaluativos empleados.

El porcentaje más alto corresponde a la categoría de los modelos evaluativos de carácter objetivista, positivistas o científicos con un 43,4%. De todos ellos y con una diferencia amplia, el más usual es el *Modelo Basado en Objetivos*; quedando las evaluaciones basadas en modelos sin referencia a objetivos y planificaciones educativas, con una escasa o nula frecuencia y porcentaje del total de los trabajos.

En segundo lugar, con un mismo porcentaje de algo más del 21%, se encuentran los modelos subjetivistas o interpretativos y los mixtos o eclécticos. El modelo *C.I.P.P.* es el que acumula mayor frecuencia, seguidos por el *Iluminativo* y *Respondente*. Obtienen escasas frecuencias, que suponen un escaso porcentaje las evaluaciones que siguen otros modelos y aquellas que pese a declarar un modelo de carácter mixto, no especifican a cual se refieren.

También son escasos los trabajos evaluativos basados en modelos críticos o emancipatorios, con un escaso porcentaje muy similar al de aquellos estudios en los que no se registra el modelo evaluativo empleado o es imposible inferir éste por las características técnicas del informe de investigación empleado.

Cotejando en un sentido comparativo los datos globales sobre los modelos evaluativos empleados y los obtenidos en el epígrafe de los paradigmas de fundamentación teórica en los que cada uno de estos modelos se desarrolla, tanto en frecuencia como en porcentaje, se obtienen los valores que se muestran en la siguiente Tabla III.

Salvo en los modelos críticos fundamentados en principios paradigmáticos homónimos que poseen idéntica frecuencia, el resto muestra ciertas diferencias, que hacen considerar un estudio pormenorizado de estas relaciones.

Cada uno de los tipos de modelos evaluativos considerados acumula, con cierta lógica, su mayor frecuencia en el paradigma en que se origina (objetivista/científico, subjetivista/interpretativo, críticos/crítico y ecléctico/complementarista).

El nivel de congruencia supone, en la mayoría de casos, porcentajes del total de trabajos considerados superiores al 60%. Siendo significativa la alta concordancia entre los modelos críticos y sus principios de teóricos de fundamentación (90,1%) y muy dispersa la adecuación de frecuencias del paradigma complementarista en modelos objetivistas, subjetivistas y eclécticos, con valores cercanos y semejantes.

Tabla II. Modelos para la evaluación de programas.

MODELO		f	fp	%	%p
OBJETIVISTAS	Basado en objetivos	83	96	39,9	46,2
	<i>Planificación evaluativa</i>	13		6,3	
SUBJETIVISTAS	<i>Respondente</i>	19	45	9,1	21,6
	<i>Iluminativo</i>	21		10,1	
	<i>Sin metas, artístico</i>	5		2,4	
CRÍTICOS	<i>Contrapuesto, Democrático</i>	12	12	5,8	5,8
ECLÉCTICOS	<i>CIPP de Stufflebeam</i> * ¹	37	46	17,8	22,1
	<i>Mixtos sin especificar</i>	9		4,3	
N.R. o N.I.		9	9	4,3	4,3
TOTAL		208	208	100	100

*¹: Modelo centrado en la evaluación de los contextos, entradas, procesos y productos de los programas.

Tabla III. Frecuencias y porcentajes reconcordancia de los modelos evaluativos empleados en relación con los paradigmas de fundamentación declarados.

		TIPOS DE MODELOS EVALUATIVOS					
		Objetivistas	Subjetivistas	Críticos	Eclécticos	N.R. o N.I.	TOTAL
PARADIGMAS	Científico	83 (70,9%)	3 (2,6%)	0	28 (23,9%)	3 (2,6%)	117 (100%)
	Interpretativo	6 (11,5%)	33 (63,5%)	2 (3,8%)	7 (13,5%)	4 (7,7%)	52 (100%)
	Crítico	0	1 (9,1%)	10 (90,1%)	0	0	11 (100%)
	Complementarista	7 (25%)	8 (28,6%)	0	11 (39,3%)	2 (7,1%)	28 (100%)
	TOTAL	96	45	12	46	9	208

Por tanto, en los trabajos revisados, se puede describir una concordancia global entre modelos evaluativos y paradigmas de fundamentación, aunque no en su totalidad, puesto que incluso se registran casos que se podrían denominar casi paradójicos al plantear actuaciones basadas en ciertos modelos y principios paradigmáticos poco congruentes.

2.6.3. Población y muestra.

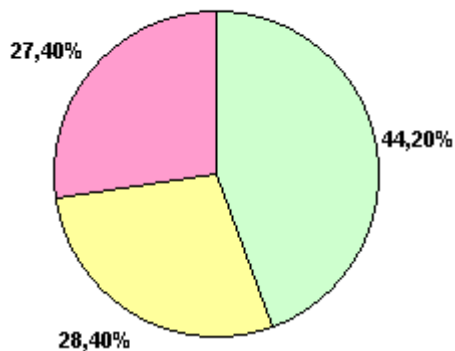
Se pretende caracterizar las poblaciones y muestras empleadas en los estudios de revisión

tomados como elementos maestres de este estudio, de forma limitada a los datos que los informes revisados ofrecen.

2.6.3.1. Descripción de la población.

El total de los resultados reflejados en la tabla IV y representados en el gráfico 3, se refieren a los 208 estudios revisados. La mayor frecuencia, 92 estudios, se refiere a aquellos en los que tan sólo se indica la población de la que se extrae la muestra tratada de manera superficial, suponiendo algo más del 44%.

Gráfico 3. Descripción de la población.



Un porcentaje mucho menor del 28,4% es el referido a los trabajos en los que si se des-

Tabla IV. Descripción de la población.

POBLACIÓN	f	fa	%	%a
Indica	92	92	44,2	44,2
Describe	59	151	28,4	72,6
NR o NI	57	208	27,4	100,0
TOTAL	208		100	

cribe en algún aspecto fundamental la población. Este porcentaje es muy similar al de los

estudios en los que no se registra o indican la población objeto de estudio.

Considerando globalmente las frecuencias de, los estudios en los que se refleja algún aspecto referido a su población, ya sea indicándolos o describiéndolos, se observa una frecuencia acumulada del 72, 6%, quedando tan sólo el 27,4% para los mencionados estudios que no registran aspecto alguno.

2.6.3.2. Técnicas de muestreo.

En la siguiente tabla se registra la frecuencia y porcentajes de las técnicas de selección

muestral empleada en la totalidad de los 208 trabajos revisados, agrupándolas parcialmente en categorías, dependiendo si su empleo supone un proceso de aleatorización o no en la selección.

Se puede comprobar que frecuencia total es mayor que el número de estudios revisados que asciende a 208. Este excedente de técnicas, se debe a que en dos estudios se recurre a dos técnicas diferentes de selección muestral.

Tabla V. Tipo y técnicas de selección muestral.

TIPO	TÉCNICA	f	fa	fp	%	%a	%p
Aleatorio	Aleatoria simple	13	13	23	6,2	6,2	10,9
	Estratificado aleatorio	1	14		0,5	6,7	
	Conglomerados aleatorizados	8	22		3,8	10,5	
	Sistemático	1	23		0,5	11	
No aleatoria	Estratificado por cuotas	3	26	101	1,4	12,4	47,9
	Conglomerados no aleatorizados	21	47		10	22,4	
	Conveniencia	40	87		19	41,4	
	Intencional	36	123		17,1	58,6	
Otras	Mixta	1	124	6	0,5	59	2,8
	Estudios censales	5	129		2,4	61,4	
NR o NI		81	210	81	38,6	100	38,6
TOTAL		210		211	100		100

En un estudio parcial se comprueba que las técnicas no aleatorias poseen la frecuencia más alta, suponiendo casi la mitad del total de ésta, pues alcanzan el 47, 9%. Sin embargo las técnicas aleatorizadas tan sólo suponen el 10, 9% del total.

La categoría “otras”, constituida por las técnicas de selección mixtas y aquellos casos en los que no se realiza selección alguna de muestra, puesto que son estudios censales, suponen un escaso 2, 8%. Siendo sin embargo muy alto (38, 6%), el de aquellos estudios en los que pese a extraer una muestra, no registran, indican o especifican el tipo y técnica de selección empleada.

La selección por conveniencia e intencional de la muestra son las técnicas más ampliamente utilizadas, tanto dentro de las técnicas no aleatorias como del total de éstas, alcanzando porcentajes muy similares del 19 y 17, 1 % respectivamente. También alcanzan un porcentaje comparativamente considerable, 10%, la selección por conglomerados no aleatorizados.

Hay que descender hasta el cuarto lugar en esta clasificación, para encontrar con un porcentaje menor, una técnica de carácter aleatorio. La selección aleatoria simple alcanza el 6, 2 % del total y a continuación la realizada por conglomerados aleatorizados con

un porcentaje del 3, 8% que supone ya casi la mitad.

Frecuencias muy bajas del total de la distribución se pueden observar en los estudios censales, en los que recurren a los estratos aleatorizados, a los procesos aleatorios sistematizados o a técnicas mixtas para la selección de la muestra.

2.6.3.3. Edades y Nivel académico de las muestras.

El registro y análisis de estos datos presentan las dificultades derivadas de la falta de claridad y unanimidad en el establecimiento explícito de los intervalos de edad de las investigaciones revisadas.

Se establecen unos intervalos en años en los que registrar los estudios atendiendo a la edad de sus sujetos muestrales. Considerado tal delimitación, además de por su ajuste a la

distribución de datos obtenida, por la semejanza a las edades correspondientes en las que se cursan los diferentes niveles establecidos en nuestro sistema educativo actual y estudiados en el siguiente epígrafe de este capítulo.

En la tabla VI, se muestra una frecuencia total mayor que la de los estudios revisados (234>208) debida a que en algunos estudios se consideran más de un intervalo de edad.

Aunque tan sólo por una diferencia de diez de los estudios revisados, una mayoría del 52% no registra o indica explícitamente la edad de la muestra considerada. Siendo la frecuencia más habitual o moda la referida a sujetos entre 7 y 12 años. Es decir, a sujetos que, según su edad, cursan estudios de Educación Primaria.

Tabla VI. Edades de las muestras en intervalos de años.

EDAD (años)	f	fa	%	%a
Hasta 6	9	9	3,8	3,8
De 7 a 12	42	51	17,9	22
De 13 a 18	35	86	15	37
De 19 a 24	8	94	3,4	40,4
Más de 24	12	106	5,1	45,6
Múltiples	6	112	2,6	48
NR o NI [*]	122	234	52	100
TOTAL	234		100	

[*] En tres casos no se registra la edad porque no es pertinente.

Con un porcentaje cercano al del intervalo anterior y que asciende hasta el 15% estarían los alumnos entre 13 y 18 años.

Cuantitativamente inferiores son las frecuencias o estudios en el resto de los intervalos de edad considerados para definir su muestra. Así los porcentajes obtenidos para muestras con más de 24 años superan escasamente el 5% del total de los estudios. Y las referidas a alumnos hasta 6 años, entre 19 y

24, o con intervalos de edad amplios y múltiples, siempre están por debajo del 4%.

En la siguiente tabla VII se categorizan los trabajos revisados, atendiendo al nivel académico declarado de la muestra que emplean en sus estudios. Estos niveles corresponden a los intervalos establecidos por el sistema educativo en la actualidad, adaptando a ellos los datos referidos a niveles dispuestos con anterioridad.

Además de los niveles básicos, se considera la categoría “otros”, que corresponde con los datos ofrecidos en aquellos estudios en los que en sus muestras se considera a padres de alumnos, cursos de postgrado para diversos niveles formativos; cursos en los que se integran sujetos con distintos niveles educativos y estudios en los que no es pertinente el

registro del nivel educativo, por no referirse su muestra a personas.

El mayor porcentaje de estudios declara un nivel académico para su muestra en la Educación Primaria con algo más del 28%. Seguidos por los que ésta cursa Educación Secundaria y estudios Universitarios, con un 19,2 y 14,6 % respectivamente.

Tabla VII. Nivel académico de las muestras de los estudios revisados.

NIVEL ACADÉMICO	f	fa	%	%a
Educación Infantil	17	17	7,8	7,8
Educación Primaria	63	80	28,8	36,6
Educación Secundaria	42	122	19,2	55,7
Formación Profesional	4	126	1,8	57,6
Universidad	32	158	14,6	72,2
Educación Especial y Compensatoria	19	177	8,7	80,9
Educación de Adultos	5	182	2,3	83,1
Otros (Educación de padres, cursos de postgrado, cursos inter-nivel y estudios en los que no es pertinente el registro del nivel educativo).	6	188	2,7	85,9
NR o NI	31	219	14,2	100
TOTAL	219 [**]		100	

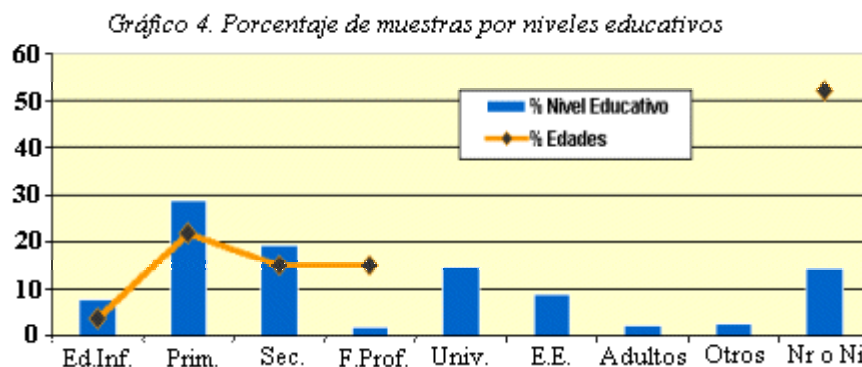
[**] La frecuencia total es mayor que el número de estudios revisados (219>208), porque 11 de ellos registran dos niveles y dos registran tres niveles educativos distintos

Frecuencias menores de veinte casos tienen los que toman sus muestras en la Educación Especial y Compensatoria o en la Educación Infantil, con porcentajes similares del 8,7 y 7,8 % del total.

El resto de categorías obtienen frecuencias por debajo de tres casos, lo que supone escasos porcentajes del total de los trabajos, quedando constituidas éstas por la macrocategoría “otros niveles educativos”, la Educación de Adultos y la Formación Profesional, que obtiene la *f* más baja.

De forma gráfica (Gráfico 4) se puede apreciar cierta semejanza entre los porcentajes de estudios cuyas muestras están en los intervalos hasta 6 años, 6 a 12 y 13 a 18; con los porcentajes de estudios cuyas muestras declaran cursar esos niveles educativos.

En el resto de los niveles educativos de las muestras no parece lógica la comparación en su porcentaje y la edad de los sujetos que las componen (línea azul interrumpida), puesto que no puede establecerse una relación unívoca entre niveles educativos y edades específicas para cursarlos.



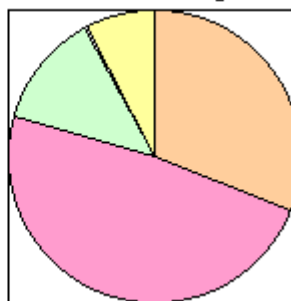
La mayoría de los estudios emplean instrumentos que no están estandarizados, sino que se elaboran de forma explícita para la propia investigación (Gráfico 5). Con una frecuencia menor, que suponen el 31,25% y 7,7% del total, encontramos a los trabajos que si emplean instrumentos estandarizados y aquellos estudios que basan la recogida de

datos en composiciones de instrumentos estandarizados + no estandarizados, categorizados como mixtos.

En total se registra un número de 483 instrumentos en los 208 trabajos revisados, lo que supone un índice medio de algo más de dos instrumentos por trabajo (2,3 instrumentos/trabajo).

Gráfico 5. Estandarización de los instrumentos de recogida de datos

Si:	31,25%
No:	48,21%
Mixtos:	13 %
NR o NI:	7,7 %



Para un estudio pormenorizado se recurre a la elaboración de la siguiente tabla (Tabla VIII), en la que se especifica aquellos instrumentos que suponen una fuente de acceso a la información directa e indirecta. Y se muestra la frecuencia y porcentaje del tipo de documento concreto empleado y diversas denominaciones similares o afines con las que se pueden encontrar en la distribución de datos registrada para esta variable.

En una primera aproximación se comprueba la escasez de fuentes indirectas empleada

y la amplia mayoría del uso de fuentes de acceso a la información directa, 7,9% frente a 83,8 % del total.

Esta macro-categoría de fuentes indirectas está constituida en su mayor parte por dos tipos de documentos: documentos de carácter más o menos oficial como libros de escolaridad, actas de evaluación o proyectos/memorias; y por otra por instrumentos de uso cotidiano, como fichas o historias clínicas. Todos ellos suponen un mecanismo para

el control/registro de los progresos de los sujetos.

El tipo de instrumento más utilizado es sin duda alguna el cuestionario o encuestas, pues

supone casi un 29% del total de instrumentos empleados. Siendo también usual el empleo de test, baterías o pruebas estándares, que alcanzan un 16,4%.

Tabla VIII. Tipos de instrumentos de recogida de datos.

FUENTE	DOCUMENTOS	OTRAS DENOMINACIONES	f	%
INDIRECTAS	Documentos	<i>Informes oficiales, actas, calificaciones, libros de escolaridad, fichas de alumnos, historias clínicas, proyectos, memorias.</i>	38	7,9
DIRECTAS	Notas de campo	<i>Anotaciones del alumno, cuadernos, registros anecdóticos, diarios, informes narrativos, cronogramas, mapas cognitivos.</i>	17	3,5
	Observación	<i>Listas de control, audio - vídeo grabaciones, parrillas de observación, dibujos, pruebas conductuales, sociométricas (tipo role - playing).</i>	49	10,1
	Entrevistas	<i>Protocolos, debates, grupos de discusión.</i>	38	7,9
	Escalas	<i>Likert, de clasificación, de valoración.</i>	51	10,6
	Cuestionarios	<i>Encuestas, Cuestionarios sociométricos.</i>	139	28,8
	Test	<i>Baterías, pruebas estandarizadas, rejilla.</i>	79	16,4
	Inventarios.		10	2,1
	Pruebas ad hoc	<i>Exámenes, ejercicios, pruebas inductivas, de rendimiento, de conocimiento.</i>	20	4,1
	Programas informáticos	<i>Bases de datos.</i>	2	0,4
NR o NI			40	8,3
TOTAL			483	100

Parece relevante hacer mención del número de casos en los que no se ha podido categorizar el tipo de instrumento empleado por la escasez de información ofrecida o lo segmentada de esta, que alcanza una frecuencia de 40, suponiendo un 8,3% del total de todos los instrumentos registrados.

Este uso, de determinados instrumentos como son por orden de frecuencia, cuestionarios (28,8%), tests (16,4%) y escalas (10,6%), se observa también en el trabajo de Bueno (2001), que revisa la producción investigativa publicada en la Revista de Investigación Educativa, incluso en porcentajes

muy similares (cuestionarios: 21,3%; tests: 13,26% y escalas: 10,6 %). Por lo que podría plantearse un empleo habitual en la I.E. Sin embargo, en otros trabajos como el de Torralbo (2001), referido a tesis doctorales en Educación Matemática, los instrumentos de mayor empleo son las pruebas *ad hoc*.

Los criterios empleados para denotar la validez de los instrumentos de recogida de datos, obtienen en el cómputo de la distribución obtenida, las frecuencias y porcentajes expresadas a continuación en la siguiente tabla.

Tabla IX. Criterios de validez para los instrumentos de medida.

Validez	f	fa	%	%a
INSTRUMENTOS ESTANDARIZADOS ¹	92	92	37,9	37,9
Denomina genéricamente, sin especificar	25	117	10,3	48,1
Pruebas piloto	2	119	0,8	49,0
Contenido	7	126	2,9	51,9
Constructo	11	137	4,5	56,4
Criterio	2	139	0,8	57,2
NR o NI	104	243	42,8	100,0
TOTAL	243		100	

¹: Se asumen los criterios establecidos, sin comprobarlos.

El cociente entre el número total de criterios registrados y el de documentos revisados, de los que se obtiene tal distribución, ofrece el coeficiente de criterios de validez medio empleado en estos estudios, que muestra un valor por debajo de la unidad por trabajo revisado.

*Índice de criterios de validez*₁ = 139/208 = 0,66 *criterios por trabajo*

Sin embargo, la interpretación de este indicador global puede ofrecer una visión sesgada, puesto que el número de instrumentos registrados a los que se refieren tales criterios de validez, es mayor que el de trabajos revisados. Por ello, el cociente debe referirse al número total de instrumentos de recogida de datos empleados en todos los trabajos revisados por los criterios registrados.

*Índice de criterios de validez*₂ = 139/483 = 0,3 *criterios por instrumento*

Se puede comprobar que la mayor frecuencia o moda la obtienen aquellos que no trabajos en los que no se ofrecen criterios de validez sobre los instrumentos de medida em-

pleados, con un porcentaje del 42,8% del total registrado. Denotando una significativa escasez en los informes de investigación sobre el establecimiento de los criterios de validez de los instrumentos de registro de datos empleado, lo que supone sin duda una enorme debilidad instrumental en las investigaciones sobre *E.P.E.*

Los valores obtenidos en los índices globales empleados en el estudio descriptivo de los criterios de validez, son idénticos en los criterios de fiabilidad para los instrumentos de registro de datos: *1,17 criterios por trabajo* y *0,5 criterios por instrumento*.

En este mismo sentido, los valores en frecuencia y porcentaje de las categorías referidas a “instrumentos estandarizados” y a aquellos que no registran o indican criterios de fiabilidad, son también idénticos, como queda reflejado en la tabla VII y comprobado mediante la comparación con la tabla X. Sin embargo, el resto de la distribución, aunque con ciertas semejanzas adopta distintos valores.

Tabla X. Criterios de Fiabilidad para los instrumentos de medida.

Fiabilidad	f	fa	%	%a
INSTRUMENTOS ESTANDARIZADOS	92	92	37,9	37,9
Denomina genéricamente, sin especificar.	30	122	12,3	50,2
Consistencia interna	7	129	2,9	53,1
Estabilidad	4	133	1,6	54,7
Equivalencia	5	138	2,1	56,8
Concordancia entre observadores	1	139	0,4	57,2
NR o NI	104	243	42,8	100,0
TOTAL	243		100	

Si se consideran las frecuencias de aquellas categorías en las que realmente no se especifican criterios de validez, por asumirse en instrumentos estandarizados o por hacer mención general ($37,9 + 12,3 = 50,2\%$) y por no registrarse o indicarse ($42,8\%$), con respecto a la que si lo hace. Denotándose igualmente una significativa escasez en los informes de investigación sobre el establecimiento explícito tales criterios. Y por tanto unas inferencias evaluativas realmente negativas.

Por lo que de forma conjunta puede considerarse que en cuanto al establecimiento de criterios específicos de validez y fiabilidad, los instrumentos en investigación sobre *E.P.E.*, presentan una enorme debilidad metodológica.

2.6.5. Técnicas de análisis de datos.

Tras la revisión de la distribución de datos obtenida y la categorización de los mismos,

los estadísticos de carácter descriptivo empleados en los diversos análisis de datos realizados en las investigaciones revisadas como muestra de este estudio son los relacionados en la tabla XI.

Atendiendo al número total de estadísticos empleados por todos los estudios revisados y el número de éstos, se establece el coeficiente medio que indica su relación ($351 / 208 = 1,7$ estadísticos descriptivos/ trabajo).

Resulta un índice moderado, que no alcanza el empleo de dos estadísticos por trabajo, en una distribución de datos que abarca un intervalo con valores máximos de 9 estadísticos y mínimos de 0. Comparativamente importante es el número de trabajos en los que no se registra o indica los estadísticos descriptivos empleados ($f = 70$ y $\% = 19,9$), ya sea por la “no pertinencia” de su uso o por la escasez de informaciones que ofrece el informe de investigación.

Tabla XI. Estadísticos descriptivos.

TIPO	ESTADÍSTICOS	f	fa	fp	%	%a	%p
Distribuciones	Frecuencias	42	42	116	11,5	11,5	31,7
	Porcentajes	74	116		20,2	31,7	
Tendencia Central	Media	72	188	81	19,7	51,4	22,1
	Mediana	3	191		0,8	52,2	
	Moda	6	197		1,6	53,8	
Variabilidad	Desviación Típica	44	241	77	12	65,8	21
	Error Típico	3	244		0,8	66,7	
	Varianza	9	253		2,5	69,1	
	Puntuación Típica	4	257		1,1	70,2	
	Rango	5	262		1,4	71,6	
	Valores máximos y mínimos	10	272		2,7	74,3	
	Cuartiles y percentiles	2	274		0,5	74,9	
Índices	Asimetría	4	278	7	1,1	76	1,9
	Tamaño del Efecto	3	281		0,8	76,8	
NR o NI		70	351	70	19,1	95,9	19,1
Otros		15	366	15	4,1	100	4,1
TOTAL		366		366	100		100

Mediante el análisis parcial por tipos de estadísticos empleados, se puede comprobar que el mayor porcentaje se refiere a aquellos que estudian y delimitan las distribuciones de datos. Esta categoría, constituida por frecuencias y porcentajes, supone el 31,7% de los estadísticos empleados.

Son semejantes el número de estudios en los que emplean estadísticos de tendencia central y de variabilidad, pues ambas categorías están entorno al 21-22% del total. Y poco significativo es el uso de índices, como los de asimetría y dificultad, pues se recurre a su empleo en escasas ocasiones de la distribución de datos analizada.

El resto de estadísticos, reseñados en la tabla anterior, son escasamente empleados, pues obtienen frecuencias comparativamente bajas y porcentajes que no superan en ningún caso el 3% de los estadísticos registrados. Existiendo un preocupante bajo uso de un estadístico descriptivo como el tamaño del efecto (T.E.) que permite hacer comparaciones entre grupos de tratamiento y que según Fernández Cano *et al.* (2000) debiera ser un recurso analítico básico en evaluación de programas.

Utilizando un procedimiento semejante pero refiriéndose en este caso los datos a los estadísticos correlacionales y multivariados, se obtienen las frecuencias y porcentajes especificadas en la tabla XII.

Tabla XII. Estadísticos correlacionales y multivariados.

TIPO	ESTADÍSTICOS	f	fa	fp	%	%a	%p
Bivariados	r de Pearson	11	11	21	5,1	5,1	9,7
	C de contingencia	1	12		0,5	5,6	
	Phi	2	14		0,9	6,5	
	V de Cramer	1	15		0,5	6,9	
	Alfa de Cronbach	6	21		2,8	9,7	
Multivariados	Regresión múltiple	6	27	18	2,8	12,5	8,3
	Análisis Factorial	8	35		3,7	16,2	
	Análisis Discriminante	2	37		0,9	17,1	
	Análisis de Conglomerados	1	38		0,5	17,5	
	W Kendall	1	39		0,5	18	
NR o NI		178	217	178	82,4	100	82
TOTAL		217		216	100		100

Atendiendo al número total de estadísticos empleados por todos los estudios revisados y el número de éstos, se establece el coeficiente medio que indica su relación en 0,85 estadísticos correlacionales / trabajo. Obteniéndose un valor inferior al de los estadísticos descriptivos, que aparecen por tanto con una utilización más usual en la distribución de datos.

En este mismo sentido, todavía es mayor que en los estadísticos descriptivos, el número de trabajos en los que no se registra o indica los estadísticos empleados. Esta frecuencia es ampliamente mayor que la frecuencia acumulada por el resto de categorías referidas al empleo de algún estadístico correlacional concreto ($f = 178 > fa = 39$), ocupando la mayoría de los casos ($\% = 82$).

En este resto de categorías, con un escaso 18% conjuntamente, se aprecia un mayor uso de los estadísticos correlacionales bivariados frente a los multivariados, aunque también con unas escasas diferencias. Indicando su análisis individual la mayor utilización del momento r de Pearson (5,1%) seguido por el Análisis Factorial (3,7%), El coeficiente alfa (α) de Cronbach y Regresión Múltiple, en un intervalo cada uno de entre (2,8). Obteniendo el resto de estadísticos porcentajes de

uso inferiores al 1%, que denotan escasa importancia cuantitativa.

Los resultados obtenidos para los estadísticos inferenciales, representados en la siguiente tabla, muestran un coeficiente medio de empleo de 0,52 estadísticos inferenciales / trabajo. Más bajo que el de los descriptivos, pero también que el de los estadísticos correlacionales/multivariados.

Analizando parcialmente las frecuencias y porcentajes en los distintos tipos en que se agrupan los estadísticos inferenciales, se puede observar una frecuencia más amplia en caso de los estadísticos paramétricos que suponen el 45% del total.

En orden decreciente de frecuencia, se encuentran a continuación, todos aquellos casos en los que no se registra, indica o especifica el estadístico inferencial empleado que alcanzan algo más del 36%. Y con un valor de frecuencia inferior a la mitad (13,8%) de los anteriores es el referido a los estadísticos inferenciales no paramétricos.

Comparativamente escasos son los valores obtenidos por los estadísticos de comparación múltiple y los empleados en la comprobación de la bondad de ajuste y verificación de supuestos, que suponen respectivamente el 2,7 y 2% del total.

Tabla XIII. Estadísticos inferenciales.

Tipo y estadístico empleado		f	fp	%	%p	
Paramétricos	<i>z normal</i>	10	134	3,4	45	
	<i>t de Student</i>	33		11,1		
	ANOVA	<i>Simple</i>		60		8,4
		<i>Factorial</i>		6		13,8
	ANCOVA	18		6		
	MANOVA	5		1,7		
	MANCOVA	2		0,7		
No paramétricos	<i>Kruskal-Wallis</i>	4	41	1,3	13,8	
	<i>Q de Cochran</i>	3		1		
	<i>T de Wilcoxon</i>	8		2,7		
	<i>U de Mann-Whitney</i>	6		2		
	<i>Chi cuadrado</i>	19		6,4		
Bondad de ajuste y verificación de supuestos.	<i>T de Barlett</i>	1	6	0,3	2	
	<i>Prueba de Kolmogorov-Smirnov</i>	1		0,3		
	<i>Prueba A de Sandler</i>	1		0,3		
	<i>Probabilidad de Mantel-Haenzel</i>	1		0,3		
	<i>T de Fisher</i>	2		0,7		
Comparación múltiple	<i>Método de Scheffé</i>	7	8	2,3	2,7	
	<i>Prueba de Tukey</i>	1		0,3		
NR o NI		109	109	36,6	36,6	
TOTAL		297	297	100	100	

En el análisis de cada estadístico por medio de su frecuencia individual la frecuencia y porcentaje más alto es el referido a los casos en los que no se especifica el estadístico, con el porcentaje anteriormente especificado. Seguido por el *ANOVA_{simple}* y la *t de Student* con un 20,2 y 11,1 % respectivamente.

El resto de porcentajes se encuentra enormemente disperso por los demás estadísticos

registrados, con valores inferiores en todos los casos al 10%.

Los estudios de la muestra revisada que realizan análisis de carácter cualitativo con los datos que han recogido son escasos, pues tan sólo suponen una *fa* de 34 del total de 210 registrados (16,2 %), como queda registrado en la Tabla XIV. Sin embargo, los casos en los que no se registra o indica este tipo de análisis alcanzan una frecuencia de 176 (83,8%).

Tabla XIV. Análisis cualitativos de datos.

Tipo		f	fa	%	%a
ANÁLISIS DE CONTENIDO	sin especificar	17	17	8,1	8,1
	del discurso	8	25	3,8	11,9
	de protocolo o documento	3	28	1,4	13,3
NARRACIÓN	análisis narrativo	1	29	0,5	13,8
ANÁLISIS DESEMPEÑO	análisis de tareas	3	32	1,4	15,2
SEMILOGÍA GRÁFICA	mapas de contenido	1	33	0,5	15,7
	diagramas de flujo	1	34	0,5	16,2
NR o NI		176	210	83,8	100,0
TOTAL		210		100	

Se obtiene un promedio de 0,16 análisis cualitativos por estudio en la muestra, siendo el promedio más bajo de todos los considerados hasta ahora y referidos tanto a los estadísticos descriptivos, correlacionales o multivariados, como a los de carácter inferencial.

Se puede comprobar que la mayoría de categorías se obtienen escasas frecuencias y porcentajes poco significativos (tan sólo el análisis de contenido que alcanza una %a del 13,3 %), mostrándose un escaso uso de procedimientos analíticos de naturaleza cualitativa en las investigaciones revisadas.

2.6.6. Diseños de Investigación.

En cuanto a los diseños de investigación, cuyas frecuencias de uso se registran en la Tabla XV, se muestra una clara disposición al empleo de los diseños de tipo cuasi-experimental (26,9%) y los de corte descriptivo (20,7%).

Los diseños pre-experimentales son empleados en semejante cuantía que los experimentales, ocupando cada uno de ellos un escaso 9,6% del total. Y en porcentajes similares se emplean las investigaciones basadas en diseños de investigación-acción y los diseños mixtos, pues obtienen el 6,3 y 5,8 % del total de casos respectivamente.

Tabla XV. Diseños de investigación.

TIPO	DISEÑOS	f	fp	%	%p
Descriptivo	Tipo encuesta o survey	26	43	12,5	20,7
	Estudios exploratorios	7		3,4	
	Estudios analíticos en profundidad	3		1,4	
	Estudios observacionales	0		-	
	Estudios sobre el desarrollo	7		3,4	
Correlacional	Estudios correlacionales/explicativos	1	1	0,5	0,5
Preexperimental	De un solo grupo con postest	5	20	2,4	9,6
	De un solo grupo con pretest-postest	15		7,2	
Cuasiexperimental	Con dos grupos y postest	10	56	4,8	26,9
	Con dos grupos y pretest-postest	33		15,9	
	Series temporales	3		1,4	
	Jerárquico	3		1,4	
	N=1	7		3,4	
Experimental	Con dos grupos (GE y GC) y postest	1	20	0,5	9,6
	Con dos grupos (GE y GC) y pretest-postest	12		5,8	
	Con múltiples grupos y pretest-postest	1		0,5	
	Diseños factoriales	6		2,9	
Cualitativo	Estudio de casos	3	8	1,4	3,8
	Estudio de casos múltiples	3		1,4	
	Estudios etnográficos	2		1	
Crítico/cooperativo	Investigación-acción	13	13	6,3	6,3
Flexible/Mixto		12	12	5,8	5,8
NR o NI		35	35	16,8	16,8
TOTAL		208		100	100

Los diseños cualitativos y estudios correlacionales son empleados escasamente, obteniendo frecuencias por debajo de diez casos y porcentajes escasos del total de la investigación registrada y analizada.

También es destacable el alto porcentaje, comparativamente con el resto de las categorías, de estudios en los que nos se registra o indica explícitamente el diseño de investigación empleado, ni se ofrece información suficiente como para inferirlo con garantías.

El diseño prototípico en la investigación sobre *E.P.E.* parece ser de tipo cuasiexperimental, concretamente de dos grupos con pretest y postest, seguido por los estudios de encuesta, usualmente limitados a la administración de un cuestionario a los sujetos incluidos en un programa.

Por tanto, puede constatarse la ausencia de diseños contrabalanceados muy útiles en la

comparación de tratamientos, y el abuso de diseños de tipo descriptivo, centrados en el estudio de alguna faceta concreta del programa evaluado. Destacando como inferencia evaluativa global el empleo de diseños de investigación, que con ciertas debilidades, puedan hacer cuestionables la búsqueda de relaciones causales.

2.6.7. Amenazas a la validez del diseño y mecanismos para su control.

Los factores y variables bajo las que se puede ver amenazado el diseño de las investigaciones son múltiples y variados, no especificando en la mayoría de los informes revisados su naturaleza concreta y las oportunas medidas destinadas a su control. Por ello, en el registro de estas variables, se ha optado por la simplificación dicotómica que permita un mejor y operativo tratamiento conjunto de datos.

Tabla XVI. Amenazas al diseño de investigación y control.

AGENTES EVALUADORES	f	fa	%	%a
Plantea amenazas y su control	25	25	12	12
Plantea amenazas pero no su control	11	36	5,3	17,3
No plantea amenazas ni su control	172	208	82,7	100
TOTAL	208		100	

Como puede apreciarse en los resultados expuestos en la tabla anterior (Tabla XIII), una alta frecuencia, que corresponde a más del 82% de los trabajos revisados, no plantea amenazas a su diseño de investigación y por tanto no considera mecanismos para controlar éstas. Siendo un escaso 17,3% de los casos, en los que se consideran posibles amenazas a sus diseños, y tan sólo en un 12% de éstos se plantean medidas para su control.

Esta escasez en el planteamiento de amenazas de los diseños de investigación y la toma de medidas para su control, denota un seguimiento mecánico de los diseños previamente establecidos por parte de los investigadores, sin plantearse este tipo de cuestio-

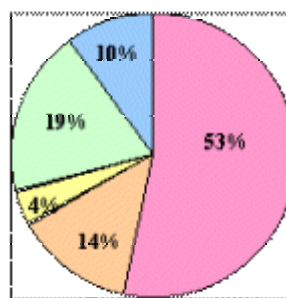
nes que serían proclives de control mediante técnicas adicionales.

2.6.8. Agentes evaluadores.

La racionalidad en la que se sustenta la consideración de esta variable de forma anexa a la de diseño de las investigaciones, queda patente al considerar ésta una información que puede arrojar cierta luz sobre cualquier diseño de investigación, en el campo de la evaluación de programas. Puesto que es una característica que facilita enormemente la delimitación epistemológica de su fundamentación, el modelo evaluativo empleado y en consecuencia los diseños de investigación que pudieran, al menos *a priori*, parecer más oportunos.

Gráfico 6. Agentes evaluadores

■ Agentes externos: f(111)
■ Implementadores: f(28)
■ Usuarios: f(8)
■ Mixto: f(40)
■ NR o NI: f(21)



Precisando los datos obtenidos en la distribución de datos, como queda reflejado en el gráfico 6, la mayoría de estudios (53,4%), están realizados por expertos/agentes externos a las instituciones u organismos a las que pertenecen los usuarios sobre los que se implementan los programas considerados.

Realmente revelador puede considerarse el dato, que en tan sólo el 3,8% del total de las evaluaciones se estime a los usuarios como principal agente evaluativo. Datos que están en consonancia con la metodología y los modelos evaluativos más empleados en las investigaciones sobre *E.P.E.*, como se describe con anterioridad en sus respectivos epígrafes. Puesto que las metodologías fundamentalmente cuantitativas y las evaluaciones desarrolladas mediante modelos de corte objetivista, plantean procesos evaluativos en los que los agentes externos juegan el principal papel.

3. CONCLUSIONES.

Las investigaciones sobre evaluación de programas se fundamentan mayoritariamente en el paradigma científico, ocupando unos porcentajes menores aquellos que se fundamentan en posiciones humanístico interpretativas o críticas.

El empleo de aproximaciones científicas determina sin duda, como señalan Arnal *et. al.* (1992), procesos de evaluación más viables para evaluar efectos finales y cuantificables, que para mejorar el desarrollo de la implementación de los programas. Sin embargo, compromete la necesaria diversidad

paradigmática y la búsqueda del complementarismo.

Estos rasgos de los ámbitos de fundamentación de las evaluaciones revisadas, determinan un uso mayoritario de modelos objetivistas para la evaluación de programas, especialmente por el aporte de frecuencia realizado por el modelo orientado a objetivos. Obteniendo un semejante y más escaso empleo las aproximaciones subjetivistas y eclécticas, y especialmente las de tipo socio - crítico.

En cuanto a la metodología más ampliamente utilizada 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 y cualitativa, más acorde con la evaluación de elementos contextuales.

Esta escasez en el uso de metodologías de corte mixto o ecléctico, determina el incumplimiento de uno de los principios que debiera tener la investigación sobre *E.P.E.*, señalados por Pérez Carbonell *et. al.* (1999), que abogan por el empleo y vigorización de estas aproximaciones, para conseguir objetivos múltiples y facilitar la triangulación.

Sin embargo, admiten 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.

En las investigaciones revisadas, no parece ofrecerse importancia a la descripción de las poblaciones a las que se generalizarían los hallazgos, ya que tan sólo se indica ésta de forma general.

Se emplean con mayor frecuencia las técnicas de selección muestral no aleatorizadas, usualmente de conveniencia e intencionales, para diseños de investigación muy diversos, destacando en la muestra considerada los denominados prefijados. Especialmente los cuasiexperimentales con dos y pretest-postest, y los descriptivos o de encuesta. Poseyendo una mayoritaria, y en ocasiones cuestionable, disposición temporal de tipo longitudinal.

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.

Las técnicas de análisis de datos se fundamentan en el empleo de estadísticos eminentemente descriptivos, especialmente los que facilitan los estudios distribucionales como son la *f* y %. Denotándose un escaso uso tanto de otros estadísticos descriptivos como el T.E., que parecerían muy pertinentes en el campo de la E.P.E. (Fernández Cano *et al.*, 2000), como de tipo correlacional, inferencial e incluso análisis de tipo cualitativo.

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.

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, también congruente con los datos obtenidos, obliga en cierto sentido a intentar controlar estas variables estableciendo diversos mecanismos para ello.

Sin embargo, se constata 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. Pese a que su mayoría está realizada por agentes externos a la implementación del programa, como evaluadores expertos.

Esta escasa preocupación por este tipo de criterios está generalizada independientemente del paradigma desde el que se opere, aunque en aproximaciones no científicas su ausencia es mucho mayor. Paradigmas desde los que también, con sus propias peculiaridades, pueden y deben considerarse aspectos referidos a la validez de sus diseños.

Aunque estas características, de los diseños de investigación, pudieran ser calificadas de cierta debilidad, son propias de la Investigación Social. No estando exentas de polémica en la investigación sobre E.P.E., como muestran la sucesión de trabajos en ciertas publicaciones como *Educational Researcher*, en los que diversos autores (Slavin, 1997; Klein, 1997; Pogrow 1998 y 1999) proponen diseños de investigación/evaluación con grupos de control, que por medio de su comparación en implementaciones prácticas “en competencia” y su revisión por expertos, permita delimitar aquel que ofrezca mejores resultados en cada caso concreto y pueda ser considerado como “diseño ejemplar”.

El uso de metodologías que pueden catalogarse con escaso rigor, ciertas debilidades en el empleo de diseños e instrumentos, y con una escasa preocupación por aspectos relacionados con la validez o fiabilidad de éstos, pudiera poner en tela de juicio el planteamiento de cualquier relación causal estable-

cida a partir de los hallazgos encontrados en algunos trabajos.

En consecuencia, y de forma general, podría considerarse una baja valoración de las realizaciones en este campo disciplinar, en su consideración global. Resultando preocupante que ciertas políticas y prácticas educativas, auspiciadas por programas previos, se hubieran podido implementar o se pretenden generalizar sin una evaluación rigurosa.

Pero no por ello deberíamos culpabilizar a la víctima (la propia E.P.E.), pues como plantea el profesor Álvarez Rojo (comunicación personal), no siempre ha sido fácil ganar el acceso a los escenarios educativos donde se puedan disponer de muestras amplias y representativas que permitiesen obtener diseños con menos amenazas a la validez. Respondiendo muchas de estas evaluaciones, más que al planteamiento por parte de los destinatarios o usuarios del programa evaluado, a la realización de investigaciones que parten del profesorado universitario, con dificultades de interacción con el profesorado, de acceso a determinados centros educativos o disposición de las aulas.

4. BIBLIOGRAFÍA.

Álvarez Rojo, V. (1993). Ámbitos temáticos y opciones metodológicas de la investigación en el campo de la Orientación. *Revista de Orientación Educativa y Vocacional*, 4 (5), 11-23.

Álvarez Rojo, V. (1994). La intervención por programas: ¿Una simple innovación tecnológica? *Revista de Investigación Educativa*, 23, 557-564.

Álvarez Rojo, V. y Hernández Fernández, J. (1997). El modelo de intervención por programas. Aportaciones para una revisión. *Revista de Investigación Educativa*, 16 (2), 79-123.

Alvira, F. (1991). *Metodología de evaluación de programas*. Madrid: CIS.

Arnal, J.; Del Rincón, D. y Latorre, A. (1992). *Investigación educativa*. Barcelona: Labor.

Bartolomé, M. y Sancho, J. M. (1997). La investigación educativa en la división de ciencias de la educación de la Universidad de Barcelona. *Revista de Educación*, 314, 269-283.

Benayas, J. (1997). La investigación en educación ambiental. Análisis de las tesis doctorales sobre educación ambiental leídas en España. En *Líneas de investigación ambiental*, J. Gutiérrez (Ed.) (pp.72-94). Granada: Proyecto Sur.

Bueno, A. (2001). *Evaluación de revistas científicas españolas del campo de la educación: El caso de la Revista de Investigación Educativa "R.I.E." (1983-2000)*. Tesis doctoral inédita.

Colás, P. (2000). *Nuevos retos en la investigación educativa*. Conferencia presentada en la II Jornadas de Psicopedagogía. F. CC. de la Educación. Universidad de Granada. (Documento interno).

Colás, P.; García, E. y Rodríguez, G. (1995). Organización de la investigación en el área MIDE de las universidades españolas. *Revista de Investigación Educativa*, 25, 45-78.

Cook, T.D. y Gruder, C.L. (1978). Metaevaluation research. *Evaluation Quarterly*, 2, 5 - 51.

De la Orden, A. y Martínez De Toda, M.J. (1992). Metaevaluación educativa. *Bordón*, 43 (4), 517-527.

Expósito, J. (2000). *La investigación sobre evaluación de programas educativos en España 1975-2000. Aspectos cuantitativos*. Granada: Autor.

Expósito, J. y Fernández Cano, A. (2002). La productividad en la investigación educativa española sobre evaluación de programas. *Revista de Investigación Educativa*, 20 (1), 113-129.

Fernández, J. (2002). Líneas de investigación en la psicología española. *Papeles del Psicólogo*, 81, 3-10.

Fernández Cano, A. (1997). Evaluación de la investigación educativa española: Una revisión integrativa de realizaciones en 25 años. *Revista Española de Pedagogía*, 207, 279-301.

- Fernández Cano, A. y Bueno, A. (1998). Síntesis de estudios bibliométricos españoles en educación. Una dimensión evaluativa. *Revista Española de Documentación Científica*, 21 (3), 269-285.
- Fernández Cano, A. y Expósito, J. (2001). Patrones de citación de la investigación española sobre evaluación de programas educativos (1975-2000). *Revista Española de Documentación Científica*, 24 (3), 289-305.
- Fernández Cano, A. Lorite, J. y Machuca, M. (2000). Exploración cientimétrica de la producción española en logopedia educativa. *Revista Logopedia, Foniatría y Audiología*, 20 (4), 197-203.
- Galante, R. y Sanchidrian, C. (1996). Evaluación de los trabajos de grado en historia de la educación desde una perspectiva bibliométrica (1980-1990). *Revista de Investigación Educativa*, 14, 37-59.
- Klein, S. (1997). Response: A system of expert panels and design competitions. Complementary federal approaches to find, develop, and share promising and exemplary products and programs. *Educational Researcher*, 26 (6), 12-20.
- Pérez Carbonell, A.; Jornet, J.M.; Suárez, J. y González Such, J. (1999). Algunas notas sobre la tendencia en evaluación de programas en nuestro país. En *Actas de Nuevas realidades educativas, nuevas necesidades metodológicas*, (pp. 77-81). Málaga: AIDI-PE.
- Pérez Juste, R. (1997). La evaluación del material impreso como elemento sustancial en la evaluación de programas. Ponencia presentada en las *Jornadas Internacionales sobre el material impreso en la Enseñanza a Distancia*. Madrid: U.N.E.D.
- Pérez Juste, R. (2000). La evaluación de programas educativos. Conceptos básicos, planteamientos generales y problemática. *Revista de Investigación Educativa*, 18 (2), 261-287.
- Pineda, J.M. (1985). Estudio bibliométrico de las publicaciones periódicas en España (1949-1976). *Historia de la Educación*, 5, 423-430.
- Pogrow, S. (1998). What is an exemplary program, and why should anyone care? A reaction to Slavin & Klein. *Educational Researcher*, 27 (7), 22-29.
- Pogrow, S. (1999). Rejoinder: Consistent large gains and high levels of achievement are the best measures of program quality?. Pogrow responds to Slavin. *Educational Researcher*, 28 (8), 24-31.
- Scriven, M.S. (1974). Standards for the evaluation of educational programs and products. En *Evaluating educational programs and products*, G. Boric (Ed.). Englewood Cliffs: Educational Technology Publications.
- Slavin, R.E. (1997). Design competitions. A proposal for a new federal role in educational research and development. *Educational Researcher*, 26 (1), 22-28.
- Stufflebean, D.L. (1974). *Metaevaluation*. The Evaluation Center. Kalamazoo, MI: Western Michigan University Press.
- Tejedor, F. J. (2000). El diseño y los diseños en la evaluación de programas. *Revista de Investigación Educativa*, 18 (2), 319-339.
- Torrallbo, M. (2001). *Análisis cientimétrico, conceptual y metodológico de las tesis doctorales españolas en educación*.

NOTAS

[1] En tres casos no se registra la edad porque no es pertinente.

[2] La frecuencia total es mayor que el número de estudios revisados (219>208), porque 11 de ellos registran dos niveles y dos registran tres niveles educativos distintos.

ABOUT THE AUTHORS / SOBRE LOS AUTORES

Jorge Expósito López (jorgeel@ugr.es). Doctor en Ciencias de la Educación, Maestro de Educación Primaria y Miembro del grupo de investigación "Evaluación de la investigación y de programas educativos andaluces". Sus área de interés son: Evaluación de Programas Educativos, Centros y Profesores e Impacto de la investigación educativa sobre la práctica docente.

Eva Olmedo Moreno (emolmedo@ugr.es). Doctora en Ciencias de la Educación: Profesora asociada del Departamento de Métodos de Investigación y Diagnóstico en la Educación. Sus áreas de interés son: Estrategias de aprendizaje y estudios sobre género e investigación educativa

Antonio Fernández Cano (afcano@platon.ugr.es): Es Profesor Titular en el Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universidad de Granada (España), cuya dirección es Facultad de Ciencias de la Educación, Campus Universitario "La Cartuja". 18071 Granada. Sus principales líneas de trabajo son "*Evaluación de la investigación*" y "*Metodología de investigación*".

Expósito López, J., Olmedo Moreno, E. y Fernández-Cano, A. (2004). Patrones metodológicos en la investigación española sobre evaluación de programas educativos. *RELIEVE*, v. 10, n. 2, p. 185-209. http://www.uv.es/RELIEVE/v10n2/RELIEVEv10n2_2.htm

ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	Expósito, Jorge, Olmedo, Eva y Fernández Cano, Antonio (2004). Patrones metodológicos en la investigación española sobre evaluación de programas educativos. <i>Revista Electrónica de Investigación y Evaluación Educativa</i> , v. 10, n. 2. http://www.uv.es/RELIEVE/v10n2/RELIEVEv10n2_2.htm Consultado en (poner fecha).
Title / Título	Patrones metodológicos en la investigación española sobre evaluación de programas educativos [<i>Methodological patterns in the spanish research on evaluation of educational programs</i>]
Authors / Autores	Jorge Expósito, Eva Olmedo y Antonio Fernández Cano
Review / Revista	Revista Electrónica de Investigación y Evaluación Educativa (RELIEVE), v. 10, n. 2
ISSN	1134-4032
Publication date/ Fecha de publicación	2004 (Reception Date: 2003 Oct 10; Approval Date: 2004 May 16; Publication Date: 2004 July 1)
Abstract / Resumen	<p><i>This study tries to obtaining some inferences about the field of program evaluation, synthesizing some patterns about research methodologies. Concretely, ones that: foundation epistemological paradigm, specific design, evaluative model in use, sample techniques, instruments for information collecting, procedures of data analysis, research designs, threats to validity and evaluation agents. To this respect, we have obtained indicators about 208 studies published in Spain from 1975 year to 2000, mainly papers from educational scientific journals. These studies were retrieved from several data base and completed with other crossed-references.</i></p> <p>Este estudio pretende realizar inferencias sobre el campo de la Evaluación de Programas Educativos (E.P.E.), sintetizando ciertos patrones relativos a la metodología de sus investigaciones. En concreto, sobre el paradigma de fundamentación epistemológica, la metodología de investigación específica, el modelo evaluativo empleado, las técnicas de muestreo, los instrumentos empleados para la recogida de información y las técnicas de análisis de datos, el diseños de investigación y las amenazas a su validez, así como los agentes evaluadores. Para ello, se revisaron 208 trabajos realizados y/o publicados en España desde el año 1975 al 2000, editados la mayoría como artículos en revistas científicas especializadas, y que se obtuvieron a partir de una búsqueda en diversas bases de datos, complementada con una búsqueda/recuperación manual de referencias cruzadas.</p>
Keywords Descriptores	<p><i>Program-Evaluation, Educational Programs, Research Methodology, Spanish Production, Research Analysis, Scientometric Review</i></p> <p>Evaluación de programas, programas educativos, metodología de la investigación, producción española, análisis de la investigación, revisión cienciométrica.</p>
Institution / Institución	Universidad de Granada (España)
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Español (Title, abstract and keywords in english)

Revista ELectrónica de Investigación y EValuación Educativa (RELIEVE)

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).