

INCIDENCIA DE LA EVALUACIÓN FORMATIVA EN EL PROCESO DIDÁCTICO

por
Concepción Felisa Abraira Fernández
Departamento de Matemáticas
Universidad de León

RESUMEN

En este artículo recogemos los resultados hallados en la evaluación del proceso, según el modelo CIPP, de un programa educativo diseñado para futuros maestros. Se aplicó con el fin primordial de mejorar los resultados cada vez más bajos que aquellos obtienen en matemáticas. Se llega a la conclusión de que la evaluación formativa en matemáticas contribuye a mejorar los resultados de los alumnos en cuanto a la variable de grupo *comportamiento de los alumnos respecto de la asignatura*. Ésta operacionaliza la variable de proceso *petición de ayuda*, necesaria para gran parte de los alumnos que quieren aprender significativamente.

Palabras clave: Aprendizaje. Educación/ matemática/ personalizada. Didáctica de la Matemática. Eficacia docente. Evaluación formativa/ de alumnos/ de proceso/ de programas. Formación de Maestros. Investigación cualitativa/ educativa/ evaluativa. Modelo CIPP.

ABSTRACT

In this paper, following the CIPP model, we gather the results found out in the evaluation process of an educational programme designed for the future teachers. It was applied with the uttermost target of improving the results increasingly lower of those obtained in Mathematics. We reach the conclusion that the formative evaluation in Mathematics contributes to the improvement of the results obtained by the students concerning the group variable *behaviour of students with the subject*. The latter makes operative the process variable of the help seeking, which is necessary for the students wishing to learn substantially.

Keywords: Assessment of students. CIPP Model. Didactics of Mathematics. Education individualised/ in Mathematics. Formative evaluation of students/ process/ programmes. Learning process. Teaching efficiency/ Training/ qualitative/ quantitative/ educational/ research.

1. INTRODUCCIÓN

Las últimas tendencias en psicología del aprendizaje en general, y en el de las matemáticas en particular (Ausubel, Novak y Hanesian, 1987; NCTM, 1990; y Novak y Gowin, 1988), coinciden en aceptar que *aprender* es una actividad personal e idiosincrásica que se produce por y en cada individuo de acuerdo con su situación, características, preferencias y conocimientos previos. En consecuencia, sólo aprenderá significativamente quién quiera, pueda y sepa hacerlo.

Por otra parte, aceptando para la educación el enfoque *personalizado* (García Hoz, 1986), es obvio que la principal —por no decir la única— misión de los profesores es ayudar a cada uno de sus alumnos a *aprender a aprender*, prestando más atención a los más débiles y a los que más lo necesitan y de la forma en que lo necesitan, ya que cada uno tiene su modo de aprender. De esta manera, se trasciende con mucho el papel de meros transmisores de conocimientos que hace años se les asignaba, y que en más de un caso —sobre todo en los niveles superiores de la educación— todavía perdura. Así pues, los profesores, como agentes de los programas educativos, han de tener en cuenta permanentemente en qué grado están presentes en sus destinatarios —los alumnos— la *motivación* necesaria para aprender, los *conocimientos previos* básicos para construir los nuevos y el *método de trabajo* que lo permita.

Los comentarios anteriores tienen una repercusión obvia, generalmente aceptada por los teóricos de la enseñanza y del aprendizaje y no tanto por los «prácticos»: no es suficiente con «enseñar» en el sentido clásico de «explicar la lección». Con todo el valor que tiene esta actuación de los profesores, su labor ha de ir mucho más allá: creemos que ésta debe consistir fundamentalmente en preparar el ambiente adecuado para que los alumnos aprendan, ayudar a cada uno según sus necesidades a situarse en el lugar adecuado para **construir** su propio conocimiento, y potenciar en ellos el deseo de hacerlo. Para ello es preciso aprovechar todas y cada una de las múltiples oportunidades que proporciona el trabajo «día a día», así como la interacción que se produce permanentemente entre alumnos y entre alumnos y profesor.

En consecuencia, para diseñar y desarrollar programas educativos ya no son suficientes las variables de *entrada* y de *contexto*, del mismo modo que para evaluarlos no bastan las de *producto*. Para un estudio integral de los fenómenos educativos es necesario contemplar las variables de *proceso* conjuntamente con las citadas, así como la interacción entre todas ellas. El paradigma que asume las ideas anteriores es el denominado *ecológico*, que aunque introducido por Doyle hace ya más de 20 años (Pérez Gómez, 1985), no tiene todavía en nuestro país la suficiente

traducción en lo que a investigaciones concretas se refiere. Dicho paradigma es el que va a encuadrar el presente trabajo. De acuerdo con él, al diseñar, ejecutar o evaluar programas es de todo punto necesario contemplar a cada alumno tanto como persona distinta de las demás, cuanto como miembro de una colectividad, lo que supone considerar también al grupo como unidad, naturalmente con el profesor formando parte de él.

Creemos que las líneas anteriores justifican debidamente la insuficiencia de los resultados finales obtenidos por los alumnos para evaluar los programas educativos. Además del *qué* ocurrió interesa conocer el *cómo*; junto a la *explicación* interesa la *comprensión* (Scriven, 1967; Weiss, 1990). En definitiva, una evaluación formativa de programas, con el propósito de perfeccionamiento y mejora tanto del propio programa como del objeto de éste —la formación matemática de los alumnos en nuestro caso— ha de contemplar variables *grupales de proceso*. Pues bien, en este artículo recogemos los resultados de una evaluación de un programa educativo, en relación con algunas de las variables del tipo considerado.

2. PLANTEAMIENTO DEL PROBLEMA

Los profesionales de la enseñanza de las matemáticas en el marco de la formación de maestros en la Universidad de León, desde hace un buen número de años venimos constatando la disminución del nivel de los alumnos, tanto en cantidad y calidad de conocimientos como en motivación y método de trabajo. Ciertamente, esta disminución no es exclusiva de nuestro contexto, pero siendo nuestros actuales alumnos futuros maestros, la situación en él es especialmente grave. Es, por tanto, urgente la búsqueda y adopción de alguna solución.

De acuerdo con numerosos teóricos, autoridades y profesionales del ámbito educativo (Bloom, Hastings y Madaus, 1975; Crooks, 1988; MEC, 1989; NCTM, 1990; Pérez Juste y García Ramos, 1989; Rico, 1993; Rosales, 1984; y Webb, 1993) que dan cuenta de la eficacia de la evaluación formativa para mejorar los resultados de los alumnos en los ámbitos afectivo, cognitivo y social, creímos útil diseñar y aplicar un programa destinado a los citados, que contemplase la evaluación en función formativa, además de la tradicional y necesaria función sumativa (como profesores responsables hemos de asignar una calificación a cada alumno). Nuestra idea de partida es que tal programa (Abraira y González, 1992) ayudaría a mejorar nuestra práctica profesional y por tanto la formación matemática de los alumnos. Para evaluar el programa usamos el modelo CIPP (Stufflebeam y Shinkfield, 1987) que contempla variables de entrada, contexto, proceso y producto. Razones obvias de espacio nos impiden reflejar aquí todos los resultados obtenidos, de modo que nos centraremos en las variables de proceso.

En nuestro contexto particular, estas variables pueden encuadrarse bajo la denominación *comportamiento de los alumnos a lo largo del curso* y corresponden a los acontecimientos que se producen en la clase durante la enseñanza, describiendo lo

que ocurre a los participantes (Rodríguez Diéguez, 1980). Dada la imposibilidad de contemplar **todo** lo que ocurre, en nuestro estudio nos hemos centrado en los alumnos, tomando el grupo como unidad de análisis y sin perder de vista la individualidad de sus componentes.

Las dimensiones constituyentes de la variable citada fueron aquellas a las que previsiblemente afectaría en mayor grado la evaluación formativa: i) asistencia a clases ordinarias, ii) participación en clase mediante preguntas o dudas planteadas al resto de la clase y/o al profesor, iii) asistencia a tutorías y/o clases de recuperación, iv) realización de pruebas formativas, v) realización de pruebas sumativas, vi) motivos de los alumnos para la revisión de exámenes. Entendimos que estos tópicos conjuntamente operacionalizaban la denominada *petición de ayuda* (Newman, 1990). En consecuencia, el problema quedaría planteado de forma general del siguiente modo: **¿ayuda la evaluación formativa, como componente del proceso de enseñanza/aprendizaje, a mejorar el comportamiento de los alumnos respecto de la asignatura?** El planteamiento operativo se obtendría subdividiéndolo en varios subproblemas, correspondientes cada uno de ellos, a cada una de las dimensiones anteriormente explicitadas.

3. PROCEDIMIENTO

Para poder constatar la existencia de «mejora» es preciso comparar y establecer criterios que permitan identificar cuándo y porqué se produce aquella. Nuestro criterio fue el comportamiento respecto de la asignatura y la referencia un grupo de alumnos en el que no se contempló la vertiente formativa de la evaluación. La acción consistió en construir dos programas educativos, que aplicamos en las mismas condiciones, idénticos en cuanto a objetivos, contenido, método y material. Uno de ellos (PF, el que suponía la innovación) incluía el uso de la evaluación *formativa* (pruebas formativas frecuentes, retroalimentación informativa y correctiva mediante informes individuales personalizados, seguimiento permanente del trabajo en clase y análisis de tareas), mientras que el otro (PS, el programa tradicional, que queríamos mejorar) incluiría sólo la vertiente *sumativa* (pruebas sumativas y calificación correspondiente).

3.1. Población y Muestra

La población estaba constituida por los alumnos de la asignatura Matemáticas I, de primer curso de la Diplomatura en Profesorado de EGB de la Universidad de León, de tamaño 522.

La muestra estuvo formada por los alumnos que cursaban la asignatura por primera vez en dos de los cuatro grupos (según orden alfabético sin tener en cuenta la especialidad) que constituían el curso primero.

Al azar, a cada uno de estos dos grupos le asignamos uno de los programas.

Llamamos GF al grupo que siguió PF (evaluación formativa y sumativa), y GS al que lo hizo con PS (sólo evaluación sumativa). Al iniciar los programas GF estaba constituido por 91 alumnos y GS por 78. Al finalizar el curso, GF tenía 52 alumnos y GS 32.

Antes de aplicar los programas, los grupos GF y GS eran idénticos en términos estadísticos, en aquellas variables que teóricamente podrían potenciar —positiva o negativamente— el impacto de la evaluación formativa. Éstas eran: conocimientos previos necesarios para comprender la asignatura, madurez para el razonamiento matemático, método de trabajo y estado opinático/actitudinal (como variables *de alumno*), y tamaño del grupo, composición del grupo por especialidades, horario, número total de asignaturas, disponibilidad de espacios, disponibilidad de material bibliográfico, perspectivas profesionales y resultados de matemáticas en cursos precedentes (como variables *de grupo*). Por razones de posible falta de objetividad hemos prescindido de las variables *de profesor*.

3.2. Contraste de hipótesis

Al ser los grupos inicialmente iguales, y ser aplicados los programas en idénticas condiciones, las diferencias que se produjesen en el proceso se deberían predominantemente al uso de la evaluación formativa en GF (es obvio que sólo podríamos decir «se deberían exclusivamente» en una situación experimental pura, que claramente no es la nuestra).

En definitiva, para poder contestar afirmativamente a nuestro problema, —aceptar nuestra hipótesis— hemos tratado de confirmar la superioridad de GF frente a GS en las variables de proceso consideradas.

Llamando X_F y X_S a los valores obtenidos por una variable X dada, en los grupos GF y GS respectivamente, la hipótesis nula sería $X_F = X_S$. En los casos en que pudimos rechazarla, aceptamos la hipótesis de trabajo $X_F > X_S$.

3.3. Recogida, tratamiento y análisis de datos

Prácticamente todas las variables de proceso consideradas eran cualitativas, de modo que sólo pudimos «medirlas» en escalas nominales. En consecuencia, usamos fundamentalmente metodología cualitativa con técnicas observacionales y de entrevista. Registramos los datos (frecuencias de asistencia y participación, anécdotas y hechos significativos en el desarrollo de las actividades cotidianas, etc.) en *fichas de observación* elaboradas a tal efecto, y que hemos cumplimentado nada más terminar cada clase o cualquier otra actividad «extra» con los alumnos, en todo caso en el plazo más breve posible.

Hemos procurado que las categorías de observación estuvieran exhaustiva y claramente definidas, lo que condujo a que las fichas citadas tuviesen una validez y fiabilidad aceptables (Selltiz, Wrightsman y Cook, 1980).

Trabajamos con un nivel de confianza del 95%, usando para el contraste de

hipótesis los valores de p obtenidos a partir del test *binomial de una cola (superior) para dos muestras independientes*. La hipótesis nula se rechazó en los casos de $p < 0.05$, aceptando en consecuencia la hipótesis de trabajo.

4. RESULTADOS Y CONCLUSIONES

En lo que sigue, nos limitaremos a expresar la frecuencia media en porcentaje, y a reflejar la aceptación o rechazo de la hipótesis correspondiente para cada una de las subvariables. Este hecho, junto con datos de observación —no siempre cuantificables, nos llevan a las conclusiones que enunciaremos. Cuando hablemos de «mayor que», «mejor que», «superior a», etc., se entenderá que la diferencia tiene significación estadística al nivel de confianza preestablecido.

4.1. Abandono de la asignatura

La importante pérdida en el tamaño de la muestra a la que antes hemos aludido nos permitió constatar que el número de alumnos que abandonan la asignatura es significativamente mayor en GS que en GF ($p=0,03$). La conclusión derivada es que *la evaluación formativa parece contribuir a evitar el tradicionalmente elevado número de abandonos en la asignatura Matemáticas I*.

4.2. Asistencia a clases ordinarias

Grupo	Frecuencia Media			Curso	
	1º Trim.	2º Trim.	3º Trim.		
GF	67,3	53,3	43,7	51,6	
GS	34,6	24	20	26,6	
Dif.	Z*	4,27	3,855	3,277	3,313
	p	0,000009	0,00006	0,0005	0,0005

Los valores encontrados para p, en todos los casos inferiores a 0.05, permiten asegurar que asisten a clase de matemáticas diariamente un número de alumnos significativamente mayor en GF que en GS. Además, los alumnos de GF se muestran más dispuestos a recuperar clases perdidas. En consecuencia, *la evaluación formativa favorece la implicación sistemática de los alumnos en el proceso de enseñanza/aprendizaje de las matemáticas*.

Por otra parte, a pesar de que en GS asistían a clase menos alumnos, lo cual favorecería la relación y la comunicación entre ellos y entre ellos y el profesor, pudimos constatar que ambas eran más intensas —dentro y fuera de clase—, y el

ambiente más distendido en GF. Los alumnos de GF trabajan en pequeños grupos asiduamente, mientras que los de GS suelen mostrarse reticentes, trabajando generalmente de forma individualista. La conclusión es que la *evaluación formativa afecta positivamente al grupo —alumnos y profesor— en los ámbitos afectivo y de relaciones humanas.*

4.3. Participación en clase mediante preguntas

Grupo	Frecuencia Media				
	1 ^{er} Trim.	2º Trim.	3 ^{er} Trim.	Curso	
GF	11,1	11,5	12	11,6	
GS	2,5	2,3	2,1	2,3	
Dif.	Z*	2,166	2,318	2,398	2,295
	p	0,015	0,0102	0,0082	0,0107

Tanto en GF como en GS la participación no fue todo lo alta que hubiésemos deseado. Hemos visto que los valores obtenidos para esta variable en GS eran debidos prácticamente a las mismas personas. Sin embargo en GF solían participar todos los asistentes. Como resultado adicional, hemos llegado a la conclusión de que, mientras los alumnos de GS que preguntaban eran de los mejores de la clase, en GF sucedía lo contrario. Aceptando que preguntar refleja duda, y por tanto que ha habido estudio y reflexión previos, mayor interés, mejor aprendizaje, concluimos que *la evaluación formativa favorece el aprendizaje a los alumnos con más carencias en formación básica.*

4.4. Asistencia a tutorías

Grupo	Frecuencia Media				
	1 ^{er} Trim.	2º Trim.	3 ^{er} Trim.	Curso	
GF	21,0	21,0	19,5	20,6	
GS	4,2	1,0	2,0	2,7	
Dif.	Z*	3,22	3,839	3,319	3,53
	p	0,0006	0,00006	0,0004	9,00021

Se aprecia que a lo largo de todo el curso asisten a tutorías más alumnos de GF que de GS. La gran diferencia que se encuentra a favor del primero complementa la

conclusión anterior, en tanto que los alumnos que asisten sistemáticamente a tutorías son aquellos que tratan de resolver las dificultades de aprendizaje en el momento en que se les plantean. Además —igual que ocurre con las preguntas en clase— los alumnos de GF asisten a tutorías con regularidad mientras que los de GS suelen hacerlo en fechas próximas a exámenes. Concluimos que *la evaluación formativa propicia un método de estudio sistemático, ayudando a corregir la práctica frecuente de estudio memorístico unos pocos días antes de los exámenes.*

4.5. Realización de pruebas formativas

	1º Trimestre				2º Trimestre			3º Trimestre		Curso
	1ª	2ª	3ª	Med	1ª	2ª	Med	1ª	Med	
Frec. Med.	52,7	47,2	36,2	45,3	34	34	34	24,1	24,1	34,5

Estas pruebas —tres en el primer trimestre, dos en el segundo y una en el tercero— formaban parte de nuestra aplicación particular de la evaluación formativa, por lo que sólo se aplicaron en GF. Tomando como referencia la participación de alumnos de otros grupos y/o cursos en actividades «extra», consideramos aceptable el resultado hallado. Dado que la realización de pruebas formativas supone interés y preocupación por el control del aprendizaje, llegamos a que *la evaluación formativa propicia una autoevaluación responsable por parte de los alumnos.* La disminución en el tercer trimestre probablemente fue debida a la inminencia de los últimos exámenes, parciales y finales, en otras asignaturas.

4.6. Asistencia a sesiones de recuperación

	1º Trimestre				2º Trimestre			3º Trimestre		Curso
	1ª	2ª	3ª	Med	1ª	2ª	Med	1ª	Med	
Frec. Med.	52,7	47,2	36,3	34	34	34	34	24,2	24,2	34,5

Estas sesiones se celebraban después de cada prueba formativa, de modo que los resultados no vienen sino a corroborar la última conclusión. La tendencia decreciente —igual que en la variable anterior— la interpretamos no como pérdida de interés por parte de los alumnos, sino como consecuencia del abandono considerado en 4.1. De hecho, se dio el caso de alumnos que habiendo decidido no examinarse continuaron realizando pruebas formativas y participando en las actividades de clase.

4.7. Realización de exámenes

Grupo	Frecuencia Media						Final	
	1 ^{er} Trim.		2 ^o Trim.		3 ^{er} Trim.			
	P	A:P	P	A:P	P	A:P	P	A:P
GF	62,6	0,3684	48,3	0,5682	59,3	0,5370	35,2	0,75
GS	69,2	0,1666	38,4	0,2267	34,6	0,2963	33,3	0,5
Dif. Z*	-0,9	2,928	1,292	3,948	3,208	3,155	0,25	3,365
p	0,8159	0,0017	0,0985	0,0004	0,0007	0,0008	0,4013	0,0004

No hay diferencia entre el número de presentados a las pruebas sumativas —los exámenes clásicos—, pero sí la hay, y esto es lo importante, en la relación aprobados/presentados. Además, la presencia de exámenes —de matemáticas o de cualquier otra asignatura— conlleva una ausencia a las clases masiva en GS y mínima en GF. Esto viene a complementar la conclusión de 4.4, en el sentido de que *la evaluación formativa potencia una autoevaluación más realista* por parte de los alumnos.

4.8. Revisión de exámenes

Grupo	Frecuencia Media						Final	
	1 ^{er} Trim.		2 ^o Trim.		3 ^{er} Trim.			
	C	S	C	S	C	S	C	S
GF	16,5	5,5	13,2	4,4	13,2	3,3	14,3	3,3
GS	3,8	14,1	3,8	12,8	3,8	12,8	3,8	11,5
Dif. Z*	2,653	-1,906	2,128	-1,739	2,128	-2,315	-2,309	-2,079
p	0,004	0,0281	0,0166	0,0409	0,0166	0,0102	0,104	0,0188

Utilizando el test binomial de una cola (inferior) para dos muestras, encontramos que el número de alumnos de GF que revisan los exámenes para conocer fallos es superior al de GS, mientras que cuando la intención es subir nota, ocurre lo contrario. En todo caso, se confirma que *la evaluación formativa potencia el afán por conocer errores y carencias de conocimiento y/o de método.*

5. CONSIDERACIONES FINALES

Aunque no sea aceptado por toda la comunidad universitaria ni por todos los responsables de la política educativa, somos partidarios de que la investigación, especialmente la investigación educativa, sea una función de los profesores tan importante como la docencia, a tener especialmente en cuenta en el marco de la formación de maestros. De acuerdo con Stenhouse (1985):

«[...] la investigación educativa será responsable de su relevancia en la práctica, y esta práctica sólo puede ser validada por los que la llevan a cabo. [...] en todo lo esencialmente práctico como la educación, es necesario que la investigación y la capacitación docente que proporcionemos apoyen esa indagación llevada a cabo por el profesor. [...] el conocimiento educativo existe en, y se verifica o falsifica en, su práctica.»

Pero la investigación necesita tiempo, del que no siempre disponemos los profesores: en multitud de casos hemos de atender a un elevado número de alumnos, durante un buen número de horas, para no siempre obtener los resultados deseados. En consecuencia, sólo si estamos suficientemente motivados nos embarcaremos en un trabajo tan laborioso y de tanta responsabilidad como es —y debe ser— el que se desarrolla en las aulas de formación de los pilares de nuestro sistema educativo.

El programa educativo, parte de cuyos efectos hemos ofrecido en este trabajo, no dio un fruto de tanta calidad como hubiésemos deseado. Este hecho, que desde nuestra vertiente de profesores podría llegar a desanimarnos, nos estimula en la vertiente investigadora: contribuye a incrementar, tanto nuestra motivación para proseguir con la investigación en el aula como el deseo de que otros miembros de la comunidad educativa en cualquier nivel colaboren en la aceptación o rechazo de nuestra firme creencia de que **la evaluación formativa en matemáticas es un factor de eficacia docente**. Ciertamente, la eficacia de nuestro programa PF (evaluación formativa) resultó ser superior, con significación estadística, a la de PS (evaluación sumativa), pero no en un grado tan elevado como para poder aceptar definitivamente su relevancia pedagógica. Así las cosas, ahora cobra sentido la función formativa de la evaluación de programas: debemos replantear el diseño de PF para mejorarlo y volver al principio tantas veces como sea necesario. Es preciso aplicar el nuevo programa (preferiblemente durante más tiempo y también en otros niveles) y evaluarlo, hasta llegar al punto —por cierto pocas veces alcanzable— en que podamos estar suficientemente satisfechos con los resultados, por supuesto, sin perder de vista nuestras limitaciones, las del medio en que nos desenvolvemos y las de los recursos de que podemos disponer.

6. BIBLIOGRAFÍA

- ABRAIRA, C. F. y GONZÁLEZ RODRÍGUEZ, M. F. (1992): *La evaluación formativa en Matemáticas como recurso didáctico: una propuesta de programa*. Congreso Internacional: Las didácticas específicas en la formación del profesorado, Santiago de Compostela, julio, (paper).
- AUSUBEL, D. P., NOVAK, J. D. y HANESIAN, H. (1987): *Psicología cognoscitiva. Un punto de vista educativo*. México: Trillas.
- BLOOM, B. S., HASTINGS, J. T. y MADAUS, G. F. (1975): *Evaluación del aprendizaje*. Troquel: Buenos Aires.
- CROOKS, T. (1988): The Impact of Classroom Practices on Students. *Review of Educational Research*, 58, 438-481.
- GARCÍA HOZ, V. (1986): *Educación Personalizada*. Madrid: Rialp.
- MINISTERIO DE EDUCACIÓN Y CIENCIA (1989): *Libro blanco para la Reforma del Sistema Educativo*. Madrid: Autor.
- NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS (1990): *Curriculum and Evaluation Standards for School Mathematics*. Virginia: Autor. (Existe traducción efectuada por la SAEM Thales: Sevilla).
- NEWMAN, R. S. (1990): Children's Help-Seeking in the Classroom: The Role of Motivational Factors and Attitudes. *Journal of Educational Psychology*, 82, 71-80.
- NISS M. (Ed.) (1993): *Cases of Assessment in Mathematics Education. An ICMI Study*. Roskilde (Dinamarca): Kluwer Academic Publishers.
- NOVAK, J. D. y GOWIN, D. B. (1988): *Aprendiendo a aprender*. Barcelona: Martínez Roca.
- PÉREZ GÓMEZ, A. (1989): Paradigmas contemporáneos de la investigación didáctica. En J. Gimeno Sacristán y A. Pérez Gómez, A. (Eds.). *La enseñanza: su teoría y su práctica* (pp. 95-138). Madrid: Akal.
- PÉREZ JUSTE, R. y GARCÍA RAMOS, J. M. (1989): *Diagnóstico, Evaluación y Toma de decisiones*. Madrid: Rialp.
- RICO, L. (1993): Mathematics Assessment in the Spanish Educational System. En M. Niss (Ed.). *Cases of Assessment in Mathematics Education. An ICMI Study* (pp. 9-20). Roskilde (Dinamarca): Kluwer Academic Publishers.
- RODRÍGUEZ DIÉGUEZ, J. L. (1980): Aportaciones de la investigación experimental a la formación de profesores. *Revista Española de Pedagogía*, 147, 37-58.
- ROSALES LÓPEZ, C. (1984): *Criterios para una evaluación formativa*. Madrid: Narcea.
- SCRIVEN, M. (1967): The Methodology of Evaluation. En R. Stake (Ed.). *Perspectives on Curriculum Evaluation* (pp. 39-63). Chicago: Rand McNally.
- SELLTIZ, C., WRIGHTSMAN, L. S. y COOK, S. W. (1980): *Métodos de Investigación en las relaciones sociales*. Madrid: Rialp.
- STENHOUSE, L. (1985): El profesor como tema de investigación y desarrollo. *Revista de Educación*, 277, 43-53.
- STUFFLEBEAM, D. L. y SHINKFIELD, A. J. (1987): *Evaluación sistemática. Guía teórica y práctica*. Madrid: Paidós/MEC.
- WEBB, N. L. (1993): Visualizing a Theory of the Assessment of Students's Knowledge of Mathematics. En M. Niss (Ed.). *Cases of Assessment in Mathematics Education. An ICMI Study* (pp. 253-263) Roskilde (Dinamarca): Kluwer Academic Publishers.
- WEISS, C. H. (1990): *Investigación evaluativa. Métodos para determinar la eficiencia de los programas de acción*. México: Trilla.