

# EVALUACIÓN FORMATIVA DE UN PROGRAMA PARA LA ENSEÑANZA DE LAS MATEMÁTICAS A ALUMNOS DE MAGISTERIO

por Concepción F. ABRAIRA FERNÁNDEZ  
*Universidad de León*

## 1. Introducción

Desde que a finales de la década de los 60 Scriven (1967) acuña el término «formativa» para calificar un tipo de evaluación de programas educativos, se viene observando el incremento del interés por ella desde distintas instancias y enfoques, pero también el de la confusión en cuanto a su significado.

En relación con la evaluación de alumnos, tanto por parte de éstos como de algunos profesores o de la sociedad en general, se identifican y/o confunden los términos evaluación, calificación, medida, examen, incluso acto de examinar. Nadie duda ya que la evaluación de alumnos es un componente más de los *curricula* diseñados para que éstos deseen aprender y lo hagan significativa y satisfactoriamente. En consecuencia, la evaluación de alumnos debe ser concebida como una herramienta de enseñanza en manos del profesor al servicio del aprendizaje de sus alumnos. Así entendida —instrumento de perfeccionamiento y ayuda— sorprende el temor, ansiedad, miedo, etc., que provoca en los alumnos y en su entorno, profesores incluidos. Como dice Rosales (1984), parece que el profesor «Por una parte enseña, demuestra, critica; por otra parte evalúa».

Ahora bien, aun siendo generalmente aceptada en la teoría la potencia de la evaluación formativa como recurso didáctico, en la

práctica apenas se utiliza, tanto menos cuanto más alto es el nivel de la enseñanza. Así pues, ¿qué ocurre? Nuestro punto de vista es: I) la aplicación de la evaluación formativa supone un considerable esfuerzo para los profesores, que además de sus obligaciones como tales, en numerosas ocasiones en el ejercicio de su profesión han de dedicarse a otras tareas distintas de la de ayudar a sus alumnos, II) no hay, o no son fácilmente accesibles para los profesores, datos empíricos suficientes que avalen la eficacia de este tipo de evaluación, III) la resistencia al cambio y la inseguridad ante las innovaciones propician el mantenimiento del papel que tradicionalmente se le ha asignado a la evaluación como acto y al profesor como agente. La primera se considera como instrumento de sanción y control, y el segundo como el juez que dictamina quién aprueba o suspende, quien averigua lo que no saben los alumnos, entendiéndolo que no es de su incumbencia responder a por qué no saben, por qué no aprenden, qué dificultades pueden tener para ello, y todavía menos ser guía y facilitador del aprendizaje. Todavía es infrecuente que se le considere como guía, como organizador de la enseñanza y de las tareas de aprendizaje, y como compañero «experto» en el caso de los estudiantes para maestros, que es el ámbito que nos ocupa.

En definitiva, parece clara la necesidad de diseñar experiencias de aplicación de la evaluación formativa en diversos contextos (diversos tipos de alumnos, nivel educativo, materia, etc.), y efectuar la investigación subsiguiente que confirme o desmienta la valía teórica de la evaluación formativa.

Pues bien, en este artículo recogemos los resultados obtenidos a partir de una investigación evaluativa de corte experimental, la cual puso de manifiesto que *el tipo de evaluación influye en los resultados de los alumnos*, siendo más eficaz la formativa que la tradicional en el contexto de alumnos de matemáticas que en el futuro serán maestros.

## 2. *El estado de la cuestión*

La pregunta ‘¿qué dicen la teoría y la investigación sobre la evaluación en matemáticas?’ está sin responder en el ámbito de alumnos de matemáticas de la Diplomatura en Profesorado de E.G.B., o mejor dicho, nos encontramos con la escasez —o al menos escasa disponibilidad— de investigaciones directamente relacionadas con la evaluación formativa en un contexto relativamente próximo al nuestro. De las que hemos tomado como base (Crooks,

1988; González Galán, 1992; Rodríguez Lajo, 1986) se deduce que la evaluación formativa propicia, en mayor grado que la tradicional, un aprendizaje significativo, gratificante, satisfactorio y duradero, acorde con las actuales teorías de aprendizaje (NCTM, 1990). Esto es debido a que la primera favorece la obtención de mejores resultados que la segunda en las dimensiones cognitiva, afectiva y social de los alumnos.

### 3. Variables de estudio

Pretendimos comparar los resultados obtenidos por los alumnos cuando se les evaluó con el objetivo principal, bien de *mejorar sus resultados*, bien de *asignarles una calificación*. Pensando en una investigación experimental, la variable independiente sería, pues, el *sistema de evaluación*. Dado que éste es uno de los componentes que definen y caracterizan a los programas educativos, al considerar dos tipos de evaluación, aun con idénticos objetivos, contenido, método, actividades y material, estaríamos ante dos programas diferentes, correspondientes a cada una de las situaciones que queríamos comparar. Estos programas constituirían sendos niveles de nuestra variable independiente, estando la investigación destinada a valorar la calidad de los productos de tales programas. En consecuencia, dichos *productos* constituirían nuestra variable dependiente. Ahora bien, al situarnos en un marco de educación personalizada y de formación de maestros, es obvio que no podemos determinar la calidad de los productos únicamente a través de las calificaciones. Sin embargo, la práctica habitual es hacerlo así. Se suelen considerar solamente aquellos productos que pueden medirse fácilmente, y hacerlo a través de pruebas como mucho válidas para registrar conocimientos memorísticos y capacidad de almacenamiento. Este proceder genera un importante reduccionismo ya que no permite tomar en consideración más productos que los *inmediatos e individuales*, olvidando los *mediatos y sociales* que, en realidad, son los que verdaderamente interesan (Orden, 1985).

Por otra parte, en situaciones tan complejas como las de enseñanza/aprendizaje intervienen múltiples variables y de variada naturaleza, de modo que es preciso tener presente la cuestión del *currículum oculto* y del *currículum nulo* (Gimeno Sacristán y Pérez Gómez, 1985). Es sabido que todo *currículum explícito* da lugar a una serie de productos no pretendidos y tal vez no a todos los pretendidos, luego ambos han de ser tenidos en cuenta. Pero en

situaciones tales como la nuestra —simples profesores responsables de la asignatura Matemáticas I (de primer curso de la citada Diplomatura)— es impensable realizar una investigación que abarque *todos* los productos, *todo* lo que ocurre como consecuencia de los programas. Pues bien, en función de:

- I) la incidencia en el aprendizaje de las variables del ámbito afectivo (Gairín, 1987)
- II) la validez del rendimiento previo como predictor del rendimiento futuro (Pérez Juste y García Ramos, 1989)
- III) la elección del enfoque formativo para la evaluación, tanto de programas como de alumnos, y por tanto la ineludible necesidad de prestar atención a los procesos —de aplicación y de aprendizaje respectivamente— decidimos centrarnos en los siguientes productos:
  - \* Comportamiento de los alumnos respecto de la asignatura Matemáticas I a lo largo del curso ( $X_1$ ).
  - \* Aprendizaje de los contenidos de la asignatura Matemáticas I ( $X_2$ ).
  - \* Estado opinático/actitudinal de los alumnos respecto de las matemáticas ( $X_3$ ).

#### 4. Metodología

En un intento de superar la clásica rivalidad entre las metodologías cualitativa y cuantitativa (Cook y Reichardt, 1986) y teniendo en cuenta que las aulas eran nuestro único «laboratorio», en el que hubimos de trabajar con alumnos y en situaciones reales, decidimos adoptar una postura abierta y ecléctica que permitiese compatibilizar, tanto como fuese posible, un enfoque experimental con planteamientos cualitativos. El primero, para constatar si se produce una superioridad de los resultados fruto de la actuación de uno de los programas frente a los del otro; los segundos, para estudiar el proceso de aplicación de éstos y replantearlos en el momento en que se detectasen deficiencias.

##### 4.1. Formulación de hipótesis

Adoptamos la siguiente notación. Sean:

- \* PF el programa que contempla la evaluación en función formativa y PS el programa cuyo sistema de evaluación tiene

sólo función sumativa

- \* GF el grupo que sigue PF y GS el que sigue PS
- \*  $X_{iF}$  y  $X_{iS}$ , respectivamente, el valor que toma la variable  $X_i$  en GF y en GS.

Después de aceptar la igualdad inicial de los grupos —tal como veremos en el apartado dedicado a la evaluación del contexto, consideramos las variables *Conocimientos previos* de matemáticas, *Método de estudio*, *Madurez* para el razonamiento matemático y *Opinión/actitud* respecto de matemáticas— y con las limitaciones inherentes a la carencia de una verdadera situación experimental, la hipótesis de trabajo y la hipótesis nula, respectivamente, serían:

$$H_E: X_{iF} > X_{iS} \quad H_0: X_{iF} = X_{iS}, \quad i=1, 2, 3$$

#### 4.2. Contraste de hipótesis

Son conocidas y generalmente aceptadas las limitaciones del paradigma *input/output* para abarcar todas las variables implicadas en los fenómenos educativos. Por ello, para estudiarlos y comprenderlos es preciso tener en cuenta que los *productos* de los programas son consecuencia de la interacción de múltiples variables de *proceso* que actúan en un *contexto* particular y en unas condiciones de *entrada* características. En consecuencia, es patente la insuficiencia de los diseños experimentales clásicos en los que la variable independiente está perfectamente controlada. Puesto que es patente su adecuación para estudiar *efectos*, decidimos incardinar el cuasiexperimental *de grupo de control no equivalente* (Campbell y Stanley, 1982) en el *modelo CIPP de Evaluación de Programas* (Stufflebeam y Shinkfield, 1997) que contempla los cuatro tipos de variables (entrada, contexto, proceso y producto) que nos interesan. Con la evaluación del *contexto* pudimos conocer la situación inicial de los participantes, con la de *entrada*, efectuar el diseño de los programas, y con las del *proceso* y del *producto* estudiar las posibles diferencias entre los efectos de los programas. Con la evaluación del *proceso*, además, efectuamos la evaluación formativa de éstos en el sentido que Scriven (1967) propugnaba.

Usando de nuevo la notación anterior, con la evaluación del proceso intentamos rechazar  $X_{iF} = X_{iS}$  y validar  $X_{iF} > X_{iS}$ . Con la del producto, rechazar  $X_{2F} = X_{2S}$ ,  $X_{3F} = X_{3S}$ , y validar  $X_{2F} > X_{2S}$  y  $X_{3F} > X_{3S}$ .

#### 4.3. Validez del diseño

Si bien la validez externa es una característica trascendental de ciertas investigaciones, en el caso de la nuestra podemos prescindir

dir de ella porque no es nuestra intención extender las conclusiones más allá de la población de estudio. Esto no impide, sin embargo, manifestar nuestro deseo de que el programa PF fuese aplicado, con las necesarias adaptaciones, en otros contextos.

Sí nos hemos preocupado, como es preceptivo, de la determinación de la validez interna. Pudimos controlar los factores de invalidación que describen Campbell y Stanley (1982), bien gracias al propio diseño cuasiexperimental, bien por la forma de incardinar en él nuestra situación particular.

#### 4.4. Población y muestra

Sobre la idea de Stenhouse (1984) «(...) la situación de campo en la que tiene lugar la acción es única», y queriendo ser rigurosos, la población sólo podría estar formada por los alumnos que cursaban por primera vez la asignatura Matemáticas I antes citada en el curso académico en que se aplicaron los programas.

La muestra estaba constituida por dos de los cuatro grupos en los que se habían distribuido dichos alumnos, según el criterio «orden alfabético». Los grupos GF y GS —muestra para la evaluación del contexto, elegidos al azar entre los cuatro posibles— estaban formados inicial y respectivamente por 91 y 78 alumnos.

La pérdida experimental fue de 39 alumnos en el primero y de 46 en el segundo, no resultando significativa la diferencia. En consecuencia, la muestra para la evaluación del producto estuvo formada por 52 alumnos de GF y por 32 de GS.

El carácter continuo de la evaluación del proceso no permite, obviamente, establecer un tamaño concreto para la muestra. De un día de clase a otro podría variar, y de hecho así ocurrió.

#### 4.5. Instrumentos de recogida de datos y de medida

Además de utilizar documentos disponibles (actas de calificaciones, expedientes de alumnos, etc.) construimos, para cada fase de evaluación los instrumentos específicos siguientes:

1. Para la evaluación del contexto: i) *Prueba de conocimientos matemáticos* básicos para la comprensión de la asignatura, ii) *Prueba de madurez* para el razonamiento matemático en el nivel universitario que nos ocupa, iii) *Cuestionario de opinión/actitud* respecto de matemáticas antes de aplicar los programas (versión A).

2. Para la evaluación del proceso: i) *Fichas de observación* para cada una de las dimensiones contemplada en la variable de proceso, pudiendo así conocer  $X_{1F}$  y  $X_{1S}$ , ii) *Pruebas formativas* (una cada dos unidades de aprendizaje).

3. Para la evaluación del producto: i) *Pruebas sumativas*, para determinar  $X_{2F}$  y  $X_{2S}$ , ii) *Cuestionario de opinión/actitud* respecto de matemáticas, para hallar  $X_{3F}$  y  $X_{3S}$  después de haber actuado los programas (versión B).

Tanto la *validez* (de contenido) como la *fiabilidad* (consistencia interna) de los instrumentos fueron debidamente determinadas, resultando en todos los casos aceptables (Abraira y González, 1992a; 1992b; Abraira y cols., 1993a; 1993b)

#### 4.6. *Análisis de datos*

Los datos que obtuvimos pueden clasificarse en dos grandes categorías: cualitativos y cuantitativos, medidos, respectivamente, en escala nominal u ordinal. Para tratar los primeros utilizamos análisis de tablas de contingencia, y la prueba  $\chi^2$  junto con la significación de este coeficiente para el contraste de hipótesis. Cuando en alguna tabla aparecían celdillas con frecuencia esperada menor que 5, procedimos a calcular el coeficiente  $\lambda$  para poder interpretar la significación de  $\chi^2$ .

Para las variables de proceso cuantitativas, o aquéllas que con la justificación debida (Pérez Juste, 1984) tratamos como tales, usamos la prueba *binomial de una cola (superior) para dos muestras independientes*.

Para los datos cuantitativos correspondientes a las variables de producto, usamos análisis de *varianza*. Para las subvariables en las que existían diferencias iniciales entre los grupos GF y GS (aunque en ningún caso eran estadísticamente significativas), realizamos análisis de *covarianza* complementarios para rechazar la hipótesis nula.

Trabajamos con  $\alpha=0,05$ . En consecuencia, rechazamos la hipótesis nula cuando el valor de p, de la significación de F, o de la significación de  $\chi^2$ , según el caso, resultó menor que dicho valor.

### 5. *Resultados y conclusiones*

Siendo nuestra intención última rechazar la hipótesis nula con un nivel de confianza del 95%, y puesto que pudimos hacerlo en todos los casos, no consideramos de interés reproducir los valores particulares obtenidos para las variables ni aquellos coeficientes utilizados. Sólo relacionaremos las conclusiones que obtuvimos al interpretar la validación de las hipótesis, ayudándonos para ello en ciertos datos de observación no cuantificables.

### 5.1. *Evaluación de entrada*

La consecuencia fundamental fue la elaboración del programa (Abraira y González, 1992a; 1992b; Bloom y cols., 1975; Pérez Juste y García Ramos, 1989) que incluía la vertiente formativa de la evaluación de alumnos. Al revisar la literatura sobre el tema — en busca de posibles modelos probados— pudimos constatar la confusión que existe en nuestro país en cuanto a los significados de los términos evaluación o evaluación formativa. La identificación de evaluación y calificación está fuertemente arraigada, y la evaluación formativa se identifica en un gran número de trabajos que hemos consultado, fundamentalmente, con pruebas frecuentes, evaluación continua (aunque tenga carácter sumativo) o retroalimentación informativa o correctiva. Es infrecuente, sobre todo en los niveles universitarios, el seguimiento permanente del trabajo dentro de la clase o el análisis de tareas de los alumnos.

En diversas ocasiones se encuentran programas educativos en los que se explicita únicamente el contenido. Los objetivos no se recogen o son demasiado generales para que sean entendidos por los alumnos. De la evaluación, como mucho, se dicen las fechas para unas pocas pruebas escritas.

A pesar de la aceptada insuficiencia del rendimiento para valorar la calidad de los productos de los programas educativos en el ámbito cognitivo, en muchos casos se siguen utilizando para ello únicamente las calificaciones obtenidas en la asignatura.

Por todo lo anterior, apoyamos la idea de la necesidad urgente (Webb, 1993) de elaborar una teoría de la educación matemática en general y de la evaluación en matemáticas en particular.

### 5.2. *Evaluación del contexto*

En esta fase de nuestro trabajo, a pesar de que hemos tenido en cuenta ciertas variables de grupo (tamaño, composición por especialidades, proporción de repetidores, horario y profesorado) nos centramos especialmente en las variables de alumno. De estas últimas, y en el ámbito afectivo, en la *Opinión/actitud* respecto de matemáticas, formada por las dimensiones *Importancia* para la formación personal, profesional y como materia de EGB, *Interés*, *Agrado*, *Utilidad*, *Actitud global* y *Autovaloración* (tabla 1). En el ámbito cognitivo consideramos *Conocimientos previos* —integrada por las subvariables *Conocimiento numérico*, *Destreza operativa*, y *Resolución de ecuaciones y de problemas de ecuaciones*—, *Madurez* para el razonamiento matemático —por *Razonamiento espacial*, *Razona-*

*miento lógico y Comprensión rigurosa de enunciados— y Método de estudio.*

Los resultados obtenidos nos llevaron a aceptar la igualdad de los grupos en términos estadísticos —a pesar de no estar formados por alumnos asignados aleatoriamente— en el caso de todas las variables analizadas. El hecho de haber aplicado ambos programas en las mismas condiciones, la diferencia entre los resultados de los grupos se debería principalmente al uso de la evaluación formativa en GF. En consecuencia, las hipótesis podrían formularse tal como lo hemos hecho.

Ahora bien, queriendo ser rigurosos, no podemos dejar de se-

ñar que en diversas ocasiones —por ejemplo, tablas 1, 2 y 5— aparecen valores de  $F$  menores que 1 que obligan a dudar del cumplimiento de las condiciones subyacentes en el AVAR. Así pues, en tanto no se lleve a cabo una revisión de la verificación de tales condiciones, y la realización de las transformaciones oportunas en las variables cuando proceda, es preciso aceptar para las conclusiones el grado de provisionalidad consecuente.

En todo caso, pudimos constatar que los alumnos que inician los estudios de Magisterio se sitúan en una posición media-baja en relación con las variables estudiadas, llegando a ser preocupante en *Conocimientos previos*, *Madurez* y *Método de estudio* (tabla 2).

### 5.3. *Evaluación del proceso*

Esta fase de evaluación constituyó la fase formativa de nuestra investigación evaluativa. Los programas fueron reconducidos siempre que detectamos en ellos alguna deficiencia corregible, tanto de diseño, como de aplicación. Los cambios se hicieron idénticamente en ambos grupos.

En cuanto a las conclusiones, pese a la valía teórica y a la confianza depositada en PF, constatamos que —al igual que PS— no consiguió evitar, en el grado deseado, el tradicional elevado número de abandonos y la desconexión de los alumnos de las actividades cotidianas. Tampoco logramos que los alumnos realizaran un trabajo tan participativo y sistemático como hubiésemos deseado.

De todas maneras el comportamiento respecto de la asignatura

(en las dimensiones *asistencia a clase, participación en clase, asistencia a tutorías, realización de exámenes y revisión de exámenes para conocer fallos y/o para intentar subir nota*) resultó significativamente superior en GF que en GS (tabla 3). En consecuencia, validamos la subhipótesis *la evaluación formativa es más eficaz que la tradicional en la mejora del comportamiento de los alumnos respecto de la asignatura Matemáticas I*. Además, encontramos que los alumnos evaluados formativamente se implicaron más —y me-

— en el proceso de enseñanza/aprendizaje, y se responsabilizaron en mayor grado, que aquéllos para los que la evaluación fue usada sólo con objeto de asignarles la calificación.

Al complementar lo anterior con datos no cuantificables obtenidos de la observación cotidiana, también llegamos a que la evaluación formativa incide más que la tradicional en la mejora del proceso educativo, especialmente en las vertientes afectiva y de relaciones humanas.

#### 5.4. *Evaluación del producto*

Esta fase de evaluación constituyó la fase sumativa de nuestra investigación. Los resultados obtenidos en ella nos llevaron a validar las subhipótesis relativas a la superioridad de PF frente a PS en las variables de producto consideradas. Es decir, *la evaluación formativa es más eficaz que la tradicional para mejorar el aprendizaje de los contenidos de la asignatura Matemáticas I* (tablas 4a y 4b), y *la opinión/actitud respecto de Matemáticas* (tabla 5).

En resumen, utilizando la evaluación del contexto como referencia inicial para juzgar los productos de los programas, consideramos que estamos en condiciones de afirmar, con un grado de seguridad razonable, que *el tipo de evaluación influye en los resultados de los alumnos*, obteniéndolos mejores los que son evaluados con propósitos formativos. En otras palabras, *la evaluación formativa es un factor de eficacia docente*.

### 6. *Consideraciones finales*

Creemos que las páginas anteriores justifican debidamente la idea según la cual *la evaluación que pretenda formar parte de un programa que quiera llamarse educativo ha de contemplar la vertiente formativa de aquélla*. Creemos que la evaluación verdaderamente pedagógica podría prescindir de la función sumativa (aunque como responsables de asignatura no podamos eludirla), pero nunca de la formativa ni, como primera fase de ésta, de la diagnóstica.

Es cierto que la aplicación de la evaluación formativa requiere tiempo y supone un esfuerzo adicional para los docentes —el realizado por nosotros no es sino un ejemplo más—. Pero los destinatarios de la evaluación que hemos efectuado eran futuros maestros y por tanto futuros usuarios de ella. Siendo conocido el papel que los enseñantes representamos como modelo de lo que hay que hacer o evitar —más para nuestros alumnos—, creímos que no deberíamos escatimar esfuerzos en darles muestras reales de posibles vías de mejora del aprendizaje, y de la motivación y el deseo de aprender, especialmente necesarias en el campo de las matemáticas. En definitiva, aunque la investigación evaluativa que hemos descrito no

nos llevó a resultados tan buenos como hubiésemos deseado, es nuestra intención continuar usando la evaluación formativa, eliminando las deficiencias detectadas en PF, mejorando lo que ha «funcionado» adecuadamente, y evaluando los nuevos resultados. Además, una pequeña mejora es preferible al retroceso, luego si hay ganancia, por escasa que sea, vale la pena continuar con dicho intento de mejora, porque además en los programas de acción, especialmente en los educativos, no existe la perfección absoluta, de modo que siempre son susceptibles de ser mejorados (Weiss, 1990).

**Dirección de la autora:** Concepción F. Abraira Fernández, Departamento de Matemáticas, Edificio de la E. U. de Graduados Sociales, Universidad de León. Campus de Vegazana, s/n 24071 León.

*Fecha de recepción de la versión definitiva de este artículo:* 20.VII.1995

#### BIBLIOGRAFÍA

- ABRAIRA, C.F. y GONZÁLEZ RODRÍGUEZ, M.F. (1992a) La evaluación formativa en Matemáticas como recurso didáctico: una propuesta de programa, *Programa y resúmenes del Congreso Internacional «Las didácticas específicas en la formación del profesorado»* (Santiago de Compostela, Universidad).
- ABRAIRA, C.F. y GONZÁLEZ RODRÍGUEZ, M.F. (1992b) Dificultades en el proceso de enseñanza-aprendizaje de las matemáticas: una propuesta de material de ayuda, pp. 209-213, en EUROPEAN PHYSICAL SOCIETY *History of the Physical-Mathematical Sciences and the Teaching of Sciences* (Madrid, Universidad Complutense).
- ABRAIRA, C.F.; ÁLVAREZ ESTEBAN, R.; GONZÁLEZ RODRÍGUEZ, M.F. y MURES, M<sup>a</sup> J. (1993a) *La evaluación formativa como factor de eficacia docente: fiabilidad de una prueba de conocimientos básicos* (VI JAEM, Badajoz) (paper).
- ABRAIRA, C.F.; ÁLVAREZ ESTEBAN, R.; GONZÁLEZ RODRÍGUEZ, M.F. y MURES, M<sup>a</sup> J. (1993,b) Elaboración y determinación de características técnicas de un cuestionario de opinión/actitud acerca de matemáticas, *Enseñanza de las Ciencias*, número extra (IV Congreso), 1993, pp. 301-302.
- BLOOM, B.S.; HASTINGS, J.T. y MADAUS, G.F. (1975) *Evaluación del aprendizaje* (Buenos Aires, Troquel).
- CAMPBELL, D. y STANLEY, J. (1979) *Diseños experimentales y cuasiexperimentales en la investigación social* (Buenos Aires, Amorrortu Editores).
- COOK, T.D. y REICHDART, CH.S. (1986) *Métodos cualitativos en investigación evaluativa* (Madrid, Morata).
- CROOKS, T. (1988) The Impact of Classroom Practices on Students, *Review of Educational Research*, 58:4, pp. 438-481.
- ESCUADERO, T. (1985) Las actitudes en la enseñanza de las Ciencias: un panorama

ma complejo, *Revista de Educación*, 278, pp. 5-25.

GAIRÍN, J. (1987) *Las actitudes en Educación. Un estudio sobre educación matemática* (Barcelona, PPU).

GIMENO, J. y PÉREZ GÓMEZ, A. (1985) *La enseñanza: su teoría y su práctica* (Madrid, Akal/Universitaria).

GONZÁLEZ GALÁN, M<sup>a</sup> A. (1992) *Predictores del rendimiento académico en la UNED* (Madrid, UNED).

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS (1990) *Curriculum and Evaluation Standards for School Mathematics* (Virginia: Autor). (Existe traducción efectuada por la SAEM, Sevilla, Thales).

NISS, M. (ed.) (1993) *Cases of Assessment in Mathematics Education. An ICMI Study* (Roskilde, Dinamarca, Kluwer Academic Publishers).

ORDEN, A. de la (1985) Hacia una conceptualización del producto educativo, *Revista de Investigación Educativa*, 3:6, pp. 271-283.

PÉREZ JUSTE, R. (1984) *Pedagogía experimental. Adaptación* (Madrid, UNED).

PÉREZ JUSTE, R. y GARCÍA RAMOS, J.M. (1989) *Diagnóstico, Evaluación y Toma de decisiones* (Madrid, Rialp).

RODRÍGUEZ LAJO, M. (1986) Incidencia de la evaluación en el rendimiento de los Universitarios en Estadística Aplicada, Tesis Doctoral (Barcelona, Facultad de Filosofía y Ciencias de la Educación. Universidad de Barcelona).

ROSALES LÓPEZ, C. (1984) *Criterios para una evaluación formativa* (Madrid, Narcea).

SCRIVEN, M. (1967) The Methodology of Evaluation, en STAKE, R. (ed.) *Perspectives on Curriculum Evaluation* (Chicago, Rand McNally).

STENHOUSE, L. (1984) *Investigación y desarrollo del curriculum* (Madrid, Morata).

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 NISS, M. (ed.) *Investigations into Assessment in Mathematics Education. An ICMI Study* (Roskilde, Dinamarca, Kluwer Academic Publishers).

WEISS, C.H. (1990) *Investigación evaluativa. Métodos para determinar la eficacia de los programas de acción* (México, Trillas).

SUMMARY: FORMATIVE EVALUATION OF AN EDUCATIONAL PROGRAM: A CASE STUDY WITH FUTURE PRIMARY SCHOOL TEACHERS IN THE SUBJECT AREA OF MATHEMATICS.

The purpose of this paper is to publish the results of a formative syllabus research carried out among future primary school teachers in the subject area of mathematics. We conclude that the formative assessment of pupils is an effective part of the teaching process.

KEY WORDS: Attitude. Didactics of Mathematics. Formative Assessment of pupils and syllabi. Teacher Training. Achievement.