

ESTUDIO DEL MÉTODO DE EVALUACIÓN EN LA ASIGNATURA DE FISIOLÓGÍA ANIMAL

Pedro A. Núñez Abades

Luis Herrero Rama

M^a Rosario Pásaro Dionisio

Dpto. Fisiología y Biología Animal
Universidad de Sevilla

RESUMEN

Se ha realizado un estudio de las calificaciones obtenidas por los alumnos de la asignatura de Fisiología Animal, en base a diferentes tests de evaluación. Las calificaciones obtenidas por los alumnos fueron extremadamente diferentes ($p < 0,001$) según cada uno de los tres tipos de tests (elección múltiple, problemas y preguntas de desarrollo). Las calificaciones fueron superiores en los exámenes de preguntas de elección múltiple (mediana en el notable). Las preguntas de desarrollo obtuvieron una calificación menor que los otros dos tipos de tests. La conclusión principal es que una evaluación correcta de los conocimientos no deberá basarse en un solo tipo de prueba. Se propone, por tanto, que el peso de la nota final de Fisiología Animal (teoría) debe provenir en un 50% de un examen de preguntas de elección múltiple y en un 50% de preguntas de desarrollo, para conseguir el objetivo de una calificación correcta de los conocimientos.

ABSTRACT

In the present work it has been studied the grades obtained by the students in the course of Animal Physiology, based into three different evaluation tests (multiple election questionnaire, problems, and free writing questions). The obtained grades of the students were statistically different ($p < 0.001$) in each one of the three different tests. The highest grades were obtained in the multiple election questionnaire (average of B). The free writing questions obtained the lowest qualification, respect to the other two types of test. The main conclusion was that a trustworthy evaluation of the knowledge should never be based in a unique type of test. We propose, in this case, that the final grade of Animal Physiology (theory) must be provided by a 50% of multiple election questionnaire and a 50% of free writing questions, to obtain a trustworthy evaluation.

INTRODUCCIÓN

En la actualidad, todos los sistemas e instituciones incorporan la evaluación como uno de los componentes fundamentales de su funcionamiento. Lo que pretenden con ello no es otra cosa que controlar la calidad de sus procesos y productos para atender al reto inexcusable de una permanente mejora. La educación no puede ser en esto una excepción (Castillo Arredondo y Gento Palacios, 1995). Por tanto, es interesante estudiar los métodos de evaluación utilizados en la enseñanza universitaria. Históricamente la evaluación está ligada a la valoración del rendimiento de los alumnos y se realiza a través de pruebas que determinan el saber de los estudiantes. Dichas pruebas tienen lugar en fechas determinadas y se basan, sobre todo, en la comprobación del aprendizaje de contenidos disciplinares (Estebaranz García, 1994). La evaluación de los conocimientos alcanzados por los alumnos permite alcanzar tres metas fundamentales (Estebaranz García, 1994; Fernández Uría, 1979; Jorba y Sanmartí, 1994):

- a. Obtener información sobre la calidad de la educación impartida y el grado de consecución de los objetivos educacionales previamente planteados.

b. Reformar el proceso didáctico cuando se estime conveniente, de cara a mantenerlo, mejorarlo o suprimirlo.

c. Hacer progresar al alumno por las líneas maestras planteadas por el profesor. De ahí, la importancia de diseñar buenas evaluaciones, que centren el trabajo y el estudio de los alumnos en aspectos básicos y relevantes desde el punto de vista cognoscitivo. También el profesor puede adquirir información estadística del alumnado y puede calificarlo a partir de un conocimiento preciso de su nivel de formación y razonamiento.

Tradicionalmente, es el examen escrito el utilizado en la valoración de los conocimientos en las áreas de Ciencias de la Naturaleza y Matemáticas. Entre las distintas pruebas escritas utilizadas cabe citar: a) preguntas de elección múltiple; b) problemas de razonamiento científico y c) preguntas de desarrollo. Por ejemplo, los exámenes de especialización del área de la salud (Biólogo Interno Residente, BIR; Médico Interno Residente, MIR; Farmacéutico Interno Residente, FIR) están constituidos exclusivamente por preguntas de elección múltiple. Los problemas de razonamiento científico y preguntas de desarrollo son los más frecuentemente utilizados en los exámenes de las asignaturas de las licenciaturas de las áreas englobadas en ciencias exactas y naturales.

En esta comunicación, se presenta un estudio de las calificaciones obtenidas por los alumnos de la asignatura de Fisiología Animal, de la Licenciatura de Biología de la Universidad de Sevilla, cuando se enfrentan a diferentes tests de evaluación, a saber: a) preguntas de elección múltiple, b) problemas de razonamiento científico y c) preguntas de desarrollo. Las conclusiones del presente estudio se aplicarán en la mejora del método de evaluación de las asignaturas que imparte el Departamento.

MATERIALES Y MÉTODOS

La asignatura de Fisiología Animal se imparte en el 4º curso de la Licenciatura de Biología (2º ciclo). Es una asignatura anual en la que se realizan distintas actividades prácticas y cuya evaluación de los conocimientos teóricos se realiza mediante tres exámenes parciales y un examen final en Junio. Los exámenes parciales se realizan al final de cada trimestre. Esta asignatura está dividida en bloques temáticos, aunque el nivel de complejidad es creciente durante el curso. Los conocimientos adquiridos en el primer trimestre fundamentan los del segundo trimestre y éstos, a su vez, son necesarios para lograr una visión integradora en el tercer trimestre. Por tanto, la evaluación, en este caso, es un proceso continuo en el que el alumno tiene que integrar los conocimientos adquiridos a lo largo del curso.

Los criterios de evaluación se dieron a conocer a los alumnos, junto con el programa, los objetivos de la asignatura, bibliografía recomendada y otras actividades docentes, al principio del curso. Los criterios de corrección y puntuación de los distintos tests de evaluación utilizados se detallan a continuación. Así mismo, estos criterios se detallan en una hoja de instrucciones que se entrega con el examen a cada alumno. En cada uno de los exámenes de preguntas de elección múltiple se utilizaron 60 preguntas. Cada pregunta consistió en un enunciado al que siguieron cinco posibles respuestas, precedidas de las letras A, B, C, D y E. El alumno tuvo que elegir la respuesta más adecuada y rellenar en la hoja de respuestas el círculo situado bajo la letra correspondiente, entre las que se encuentra a la derecha del número igual a la pregunta.

EJEMPLO: En la hoja de respuestas aparecería así, 18. A B C D E

o o o o o

Si hubiera elegido la respuesta B como la más adecuada para la pregunta número 18. Los enunciados se refieren siempre a las circunstancias y condiciones más generales. Las contestaciones se valoran exclusivamente dependiendo de cómo aparecen en la hoja de respuesta. Todas las preguntas tienen una sola respuesta como más adecuada, aunque varias o todas puedan ser válidas o verdaderas. Se anulan automáticamente las respuestas en las que aparezca más de un círculo relleno. Si el alumno desea rectificar su respuesta se le explica que ha de hacer una X sobre el círculo relleno que anula. Es necesario acertar 35 preguntas para obtener 5 puntos sobre 10. El máximo, 10 puntos, se obtiene con 60 preguntas acertadas. Un ejemplo de pregunta de elección múltiple es el siguiente:

¿Cuál de los factores siguientes estimula la regulación de la secreción de ADH?

- A. la disminución de la presión arterial
- B. el aumento de la ingesta de agua
- C. el aumento de la ingesta de sodio
- D. A y B son ciertas
- E. A y C son ciertas

La respuesta A y C, son ciertas por lo que el alumno debe contestar E como única respuesta válida.

Entre las actividades docentes a realizar por los alumnos a lo largo del curso están las sesiones de discusión de problemas de Fisiología. Dichas sesiones de discusión tienen lugar en grupos reducidos (10 alumnos). Previamente a la sesión de discusión los alumnos disponen de un cuadernillo con unos 60 problemas tipo y su solución. Por tanto, los razonamientos para llevar a cabo la ejecución de los problemas se discuten con anterioridad al examen. En el examen de problemas, el alumno dispone de papel para hacer los cálculos necesarios antes de contestar a lo que se le pregunta. Cada problema consta de varios apartados (al menos cuatro). Cada uno de los apartados se puntúa respecto de la dificultad que conlleva. En el primer y tercer parcial los problemas seleccionados para el examen fueron similares a los problemas tipo discutidos previamente en las sesiones de discusión. En el segundo parcial, por el contrario, se eligió un problema que englobara distintos problemas tipo. Un ejemplo de problema del II parcial es el siguiente:

Un excursionista realiza una ascensión desde una altura al nivel del mar hasta una altura de 1.000 m (presión barométrica = 674,1 mmHg), sin darse cuenta de que tenía fiebre a causa del virus de la gripe (39 °C, P_{H_2O} = 52,4 mmHg). Con el ejercicio de la ascensión, su consumo de oxígeno aumentó a 1.200 ml/min (STPD), siendo el gasto cardíaco 14 l/min. Calcular: a) ¿Qué disminución habrá experimentado la presión parcial de oxígeno alveolar respecto de la que tenía a nivel del mar?, b) ¿Qué disminución habrá experimentado, consecuentemente, la concentración de oxígeno alveolar?, c) ¿Cuál será la diferencia arterio-venosa de la concentración de oxígeno? y d) Podrá satisfacer las demandas de oxígeno en dicha altitud?. Tenga en consideración los datos siguientes: $\beta_g = 51,73 \mu\text{ml/l. mmHg}$, $F_{O_2} = 0,2095$. Razone brevemente las respuestas.

El examen de preguntas de desarrollo constó de tres preguntas. Cada pregunta corta tiene una puntuación máxima de un tercio de la nota global de este tipo de examen. Las instrucciones para las preguntas cortas fueron las siguientes: 1. Conteste de forma ordenada

y concisa pero simple. 2. Escriba de forma fácilmente legible, utilizando como máximo media carilla de folio donde se encuentra el enunciado. 3. De su contestación se valorará exclusivamente lo que esté relacionado directamente con la pregunta. 4. Revise cuidadosamente lo que escriba. Un ejemplo de pregunta de desarrollo es la siguiente: Regulación del vaciamiento gástrico.

El tamaño de la muestra fue de 153 alumnos de cuarto curso de Fisiología Animal de la Licenciatura de Biología. Dicha muestra se corresponde con el 100% de los alumnos presentados a los distintos exámenes y con el 72,8% de los alumnos matriculados en la asignatura durante el curso 1996-97. El estudio estadístico de los datos se realizó mediante el test de ANOVA para grupos relacionados (Colás Bravo, 1993) utilizando el paquete estadístico INSTAT 2.0 para DOS. El análisis estadístico se realizó con las calificaciones de 0 a 10 puntos. Sin embargo, en las figuras 1 y 2 las calificaciones se encuentran agrupadas, por simplificación, en las cuatros categorías que finalmente quedan reflejadas en las actas: de 0 a 4.9 puntos, suspenso; de 5 a 6.9, aprobado; de 7 a 8.9, notable; de 9 a 10, sobresaliente.

RESULTADOS

Las calificaciones obtenidas por los alumnos en los distintos tipos de pruebas se representan en las figuras 1 y 2. En la figura 1A se puede observar que la evaluación de las preguntas de elección múltiple muestra una distribución en forma de curva de Gauss. Esta distribución de las calificaciones fue similar para los tres parciales. Sin embargo, se observa que las calificaciones obtenidas mejoran a lo largo del curso. Al hacer la media de los tres parciales (Fig. 2) se observa que un cincuenta por ciento de los alumnos obtendría notable. Con este tipo de prueba, un 92% de los alumnos presentados al examen superarían la asignatura si el examen estuviese constituido exclusivamente por preguntas de elección múltiple.

Las calificaciones obtenidas en el examen de problemas (Fig. 1B) muestran una distribución distinta en los tres parciales. Un mayor número de alumnos obtuvo sobresaliente en el primer y tercer parcial. Sin embargo, la mayoría de los alumnos suspendió en el segundo parcial. Al

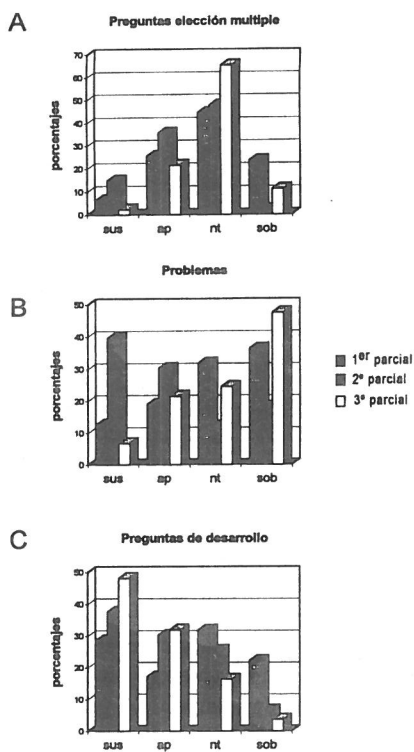


Figura 1: Diagrama de barras que muestra las calificaciones obtenidas por los alumnos con el examen de preguntas de elección múltiple (A), problemas (B) y de desarrollo (C). Las calificaciones se agruparon en: de 0 a 4.9 puntos, suspenso (sus); de 5 a 6.9, aprobado (ap); de 7 a 8.9, notable (nt); de 9 a 10, sobresaliente (sob). Las barras grises muestran las puntuaciones obtenidas en el primer parcial. Las barras negras muestran las puntuaciones del segundo parcial y las barras blancas las del tercer parcial.

hacer la media de los tres parciales, las calificaciones obtenidas por los alumnos se distribuyeron de forma homogénea (Fig. 2).

Las calificaciones de las preguntas de desarrollo presentó una distribución decreciente en el segundo y tercer parcial (Fig. 1C), y al hacer la media de los tres parciales (Fig. 2), con un porcentaje mayor de alumnos en el grupo de suspensos (37%) y menor en el de sobresalientes (10%). Sin embargo, los resultados del primer parcial muestran un menor número de suspensos.

Por último, al comparar la media de las calificaciones obtenidas por cada alumno, mediante el estudio estadístico, éstas fueron extremadamente diferentes ($p < 0,001$) tras cada uno de los tres tipos de tests (Figura 2).

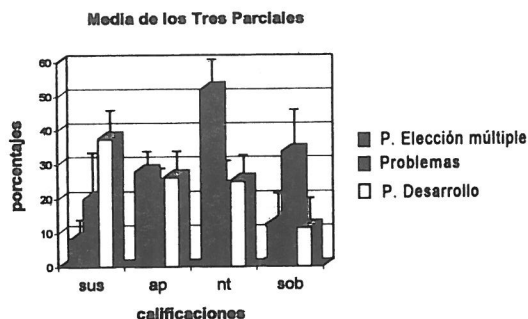


Figura 2: Diagrama de barras que muestra las calificaciones medias (media de los tres parciales), con su correspondiente error estándar, obtenidas por los alumnos en el examen de preguntas de elección múltiple (barras grises), problemas (barras negras) y de desarrollo (barras blancas).

DISCUSIÓN

En la presente comunicación se ha estudiado las distintas formas de evaluación de los conocimientos en la asignatura de Fisiología Animal de 4º curso de la Licenciatura de Biología. El propósito ha sido el diseño de pruebas objetivas como instrumento adecuado para la evaluación de dichos conocimientos.

Las preguntas de elección múltiple se elaboraron sobre la base de preguntas de enunciado breve, al que suceden una o más alternativas de respuesta (Estebaranz García, 1994). Las preguntas de elección múltiple presentan la ventaja de la objetividad en la corrección. Un test obtiene la misma puntuación por parte de diferentes correctores o en momentos distintos (Fernández Uría, 1979). En el diseño de preguntas de elección múltiple lo más importante es definir el campo y seleccionar las cuestiones más relevantes de la materia que se somete a prueba, de forma que sean una muestra de las partes más representativas del conocimiento o destrezas específicas que deben aprenderse (Madaus y Kellagham, 1992). En este sentido, el grado de representatividad y dificultad de las preguntas se ha ensayado repetidamente a lo largo de los distintos cursos académicos, concretamente desde el curso 1978 hasta el 1997. En la elección de las preguntas que entran en el examen se pondera una mayoría de preguntas de mediana o poca dificultad y alta representatividad (80%), un 15% de preguntas de dificultad alta y 5% de preguntas con un grado muy alto de dificultad. Para Madaus y Kellagham (1992), los tests de elección múltiple siguen ocupando una posición central dentro de las actividades de evaluación. Los problemas más importantes que se plantean en el examen de preguntas de elección múltiple se relacionan con la validez; es decir, si cumple la función para la que se ha realizado y la fiabilidad, o consistencia de los resultados (Estebaranz García, 1994). Asimismo, presenta el inconveniente de la dificultad de medición de objetivos de nivel superior (capacidad de razonamiento y relación de los conocimientos) (Fernández Uría, 1979). El

resultado fundamental del presente estudio es que con cada tipo de evaluación el alumno obtiene calificaciones estadísticamente diferentes. El menor número de alumnos suspensos (calificación <5) se obtuvo con los tests de preguntas de elección múltiple. Con estos tests, un 50% de los alumnos obtuvieron una calificación de notable. Este éxito en la puntuación de los alumnos puede deberse a que tienen a su disposición preguntas de elección múltiple de otros años, con la consiguiente ventaja que supone el conocer el tipo de pregunta que pueden presentarse en el examen. Otra posibilidad es que el alumno encuentra gran facilidad cuando tiene que elegir la respuesta correcta. Dicha tarea no es fruto de su capacidad constructiva. Además, hay que tener en cuenta la validez de estos tests para medir capacidad de razonamiento y relación de conocimientos. Por último, la mejoría observada en las calificaciones a lo largo de los tres parciales, probablemente refleja el proceso de adquisición de conocimientos a lo largo del curso.

Teniendo en cuenta las ventajas e inconvenientes de los exámenes basados exclusivamente en preguntas de elección múltiple (Fernández Uría, 1979), el examen se complementa con problemas y preguntas de desarrollo. La inclusión de este tipo de problemas en los exámenes tiene la ventaja de que se toma en consideración el proceso mental y las vías adoptadas por el alumno para encontrar la respuesta (Fernández Uría, 1979). Además, el alumno tiene que memorizar fórmulas. Sin embargo, poseen un grave inconveniente: son pruebas cerradas, con una sola respuesta válida. Tras la evaluación, se observó que las calificaciones que obtuvieron los alumnos en el examen de problemas dependía de la similitud que los problemas guardaban con los del cuadernillo previo o con los expuestos en clase. Así, en el primer y tercer parcial, las notas fueron significativamente más altas ($p < 0.01$) que en el segundo parcial.

En las pruebas de desarrollo se deja un amplio margen a la iniciativa del alumno, siendo posible diferentes enfoques por su parte. La ventaja es que una prueba de ensayo bien diseñada es un instrumento de primera magnitud para la medida y evaluación de la capacidad de síntesis, dado que se exige la integración de conocimientos previos en el diseño y redacción de algo nuevo, fruto de la actividad intelectual y de la capacidad constructiva del alumno (Fernández Uría, 1979). Sin embargo, con frecuencia se critica esta forma de evaluación por su escasa objetividad. Así, profesores distintos, o el mismo profesor en momentos diferentes, otorgarían calificaciones distintas a una misma prueba realizada por un alumno concreto. Para que estos puedan ser corregidos objetivamente deben explicitarse los criterios y puntuación de los mismos antes de su realización (Estebaranz García, 1994). Al margen de esta última consideración (ya comentada en material y métodos), se observó que las calificaciones de los alumnos presentaron un orden decreciente, con un porcentaje máximo de alumnos en el grupo de suspensos. Dicha distribución fue similar para cada uno de los parciales. Este resultado muestra claramente que el alumno encuentra mayor dificultad a la hora de realizar una tarea no mediatizada fruto de su propia capacidad constructiva.

Como ya se ha comentado en la introducción, las conclusiones del presente estudio se aplicarán en la mejora del método de evaluación de las asignaturas del Departamento. En la actualidad, son numerosos los métodos utilizados para medir la calidad de la educación (recopilados en Castillo Arredondo y Gento Palacios, 1995). Muchos de estos métodos de evaluación de programas docentes están basados en el modelo conductista-eficiencista. Uno de ellos, el de la evaluación de programas a través de la consecución de objetivos (de Tyler), se fundamenta en la comparación interna entre resultados logrados y objetivos planteados (Stufflebeam y Shinfield, 1987). El grado de éxito de un programa se fija, por tanto, en la

determinación del logro de unos objetivos que fueron claramente establecidos (lo que implica todo un planteamiento de la programación curricular centrado en los contenidos a adquirir) y en la estimación de la consecución de dichos objetivos. En estos programas, el método prevé la elaboración de pruebas de valoración o tests que trate de medir el logro de los objetivos previstos (Castillo Arredondo y Gento Palacios, 1995). Nuestros resultados demuestran que la propia consecución de los resultados, entendida como que el alumno apruebe la asignatura, depende enormemente del método de evaluación seguido. Un alumno tras estudiar la asignatura obtendría una calificación de notable si los exámenes fueran exclusivamente de preguntas de elección múltiple y de suspenso si el examen fuera solo de preguntas de desarrollo. De ahí que nuestro estudio redunde en la crítica al Método de Consecución de Objetivos; debido a la inflexibilidad con la que actúa en busca del logro de los objetivos. Este método no atiende a otros aspectos de la puesta en acción del programa y que pueden tener consecuencias verdaderamente importantes para el mismo (Stecher y Davis, 1990; Popham, 1981). En definitiva, los modelos que pretendan estudiar la calidad de la enseñanza no pueden basarse exclusivamente en el número de alumnos que aprueban la asignatura. Por contra, criterios más globales que tengan en cuenta los antecedentes (condiciones previas del programa), el proceso de interacción entre los implicados en el programa y los resultados parecen ser los más adecuados para la evaluación de la calidad de un programa docente. Entre ellos destacaríamos los de Stufflebeam (Stufflebeam y Shinfield, 1987) y el de la figura de Stake (Stake, 1967).

CONCLUSIONES

Las conclusiones de este estudio son las siguientes:

1) Con cada tipo de evaluación se obtienen calificaciones de los conocimientos diferentes. Por tanto, las pruebas objetivas para una correcta evaluación deben englobar al menos dos tipos de preguntas. La evaluación no debería basarse nunca en un solo tipo de prueba, puesto que ninguna de ellas permite evaluar los conocimientos adquiridos en su totalidad. Las calificaciones obtenidas por un mismo alumno fueron muy dispares con preguntas de elección múltiple y con preguntas de desarrollo.

2) Cuando el objetivo es obtener información sobre la calidad de la enseñanza impartida es arriesgado sacar conclusiones utilizando los resultados de un sólo tipo de evaluación; en concreto, los tests con preguntas de desarrollo. En estos tests se evalúa la capacidad constructiva del alumno y no solo sus conocimientos (Estebanz García, 1994).

En una asignatura como la Fisiología Animal donde el fin último es que el alumno comprenda y relacione las variables fisiológicas, no conviene dar un peso excesivo a los exámenes de preguntas de elección múltiple. Por ello, no sería recomendable que el examen teórico se basara exclusivamente en este tipo de preguntas. Tampoco es oportuno unos exámenes basados solamente en preguntas de desarrollo, ya que se infravalorarían los conocimientos teóricos que el alumno tiene de la asignatura. Por último, la inclusión de problemas en los exámenes solo sería aconsejable si hay una preparación previa con sesiones de problemas. Por todo ello, se propone que para una mejor medida de los conocimientos teóricos de los alumnos de Fisiología Animal es necesario realizar un examen en el que un 50% corresponde a preguntas de elección múltiple y otro 50% a preguntas de preguntas de desarrollo.

REFERENCIAS

- CASTILLO ARREDONDO, S. y GENTO PALACIOS, S. (1995). *Modelos de evaluación de programas educativos*, en A. MEDINA RIVILLA y L.M. VILLAR ANGULO (Eds). Evaluación de programas educativos, centros y profesores. Madrid, Editorial Universitas, S.A., 25-72.
- COLÁS BRAVO, M.P. (1993). *Análisis e interpretación de la información recogida*, en M.P. COLÁS BRAVO y M.A. REBOLLO CATALÁN (Eds). Evaluación de programas: una guía práctica. Sevilla, Editorial Kronos.
- ESTEBARANZ GARCÍA, A. (1994). *Evaluación de programas curriculares y procesos de enseñanza-aprendizaje*, en L.M. VILLAR ANGULO (ed.). Manual de entrenamiento: Evaluación de Procesos y Actividades Educativas. Barcelona, Editorial PPU, 245-279.
- FERNÁNDEZ URÍA, E. (1979). *Estructura y didáctica de las Ciencias*. Madrid, Servicio de Publicaciones del Ministerio de Educación.
- JORBA, J. y SANMARTÍ, N. (1994). *Enseñar, aprender y evaluar: un proceso de regulación continua. Propuestas didácticas para la áreas de ciencias de la naturaleza y matemáticas*. Barcelona, Centro de publicaciones. Secretaría general técnica del Ministerio de Educación y Cultura.
- MADAUS, G.F. y KELLAGHAM, Th. (1992). *Curriculum Evaluation and Assessment*, en Ph. W. JACKSON (ed.). Handbook of Research on Curriculum. New York, MacMillan Publishing Company, 119-154
- POPHAM, W.J., (1981). *Modern Educational Measurement*. Englewood Cliffs (NJ), Prentice Hall.
- STAKE, R.E. (1967). The countenance of education evaluation. *Teachers College Record*, 68: 523-44.
- STECHEER, B.M. y DAVIS M.A., (1990). *How to focus on evaluation*. Newbury Park (Calif), Sage.
- STUFFLEBEAM, D.L. y SHINFIELD, A.J. (1987). *Evaluación sistemática*. Madrid: Paidós-MEC.

AGRADECIMIENTOS

Los autores desean expresar su agradecimiento al Prof. Dr. L.M. Villar Angulo por su asesoramiento.