EVALUACIÓN Y MEDICIÓN: COMPARACIÓN DE CUATRO MANERAS DE MEDICIÓN ASISTIDAS POR ORDENADOR. UN ESTUDIO ESTADÍSTICO

ANTONIO ARIZA GARCÍA,
PAULA DAZA NAVARRO
Y JOSÉ TORREBLANCA LÓPEZ

UNIVERSIDAD DE SEVILLA

Este artículo describe un programa informático que genera y corrige pruebas de tipo test. Se constituye como un entorno abierto y flexible que puede ser actualizado y modificado a través de las bases de datos que contienen las preguntas de los diferentes temas. Posee distintos criterios de corrección que son analizados para comprobar su adecuación en la medición de los aprendizajes de un grupo de alumnos.

This article describes software that generates and grades multiple choice tests. It consists of an open and flexible environment that can be updated and modified by means of databases containing discrete item questions for different units. It has different grading standards that are - analysed to check its accuracy measuring the uptake of a group of students.

DESCRIPTORES: Evaluación, Medición, Software educativo, Nuevas tecnologías, Bases de datos, Informática, Corrección asistida por ordenador.

1. INTRODUCCIÓN.

La evaluación es un proceso sistemático, continuo e integral destinado a determinar hasta que punto han sido alcanzados los objetivos educacionales (Fermín, 1971). Metodológicamente, la evaluación resulta de relacionar dos realidades que son: lo observado y lo esperado. Son muchos los procedimientos de evaluación que podemos utilizar, siendo válidos todos aquellos que sean capaces de poner de manifiesto si las actividades del profesor y del alumno llevan al logro de los objetivos propuestos (Torreblanca y otros, 1996).

Como dice De Juan (1996), uno de los aspectos más importantes de la evaluación es la medición. Aunque en muchas ocasiones ambos términos se confunden, la medición es una fase previa que proporciona objetividad a la evaluación. Uno de los sistemas de evaluación utilizado con más frecuencia son las preguntas objetivas de respuestas múltiple, que como ya dijimos en nuestro anterior trabajo, se pueden considerar como un examen escrito de preguntas con características muy concretas (Torreblanca y otros, 1996) entre las que cabe destacar como más interesantes: 1) han de ser muy breves, 2) han de estar enunciadas muy claramente, y 3) deben permitir una calificación cuantitativa rápida

La gran difusión de este tipo de pruebas se debe fundamentalmente a que se hallan libres de muchos de los defectos atribuibles a otro tipo de exámenes (Lafourcade, 1977), como por ejemplo: falta de objetividad, diferencias entre distintos examinadores, criterios cambiantes en una misma prueba, etc. Las principales ventajas de este sistema radican en que se puede abarcar gran parte de los temas tratados en clase, la evaluación es totalmente objetiva y es de fácil corrección. Este tipo de pruebas ha conseguido un alto grado de aceptación de los exámenes, tanto por parte de profesores como de alumnos, debido entre otras razones a su gran fiabilidad y validez, puesta de manifiesto por diferentes autores (Hubbard, 1978; Thyne, 1978 y Guilbert, 1994).

En el caso de grandes grupos la utilización de pruebas objetivas es muy adecuada, y aunque su preparación es compleja, las ventajas anteriormente mencionadas hacen rentable el esfuerzo inicial. La introducción del ordenador ha dotado de un instrumento rápido y preciso con un amplio campo de posibilidades a las pruebas objetivas (Ariza y otros, 1998).

En nuestro anterior trabajo demostramos que la gestión informatizada de bases de datos de preguntas de respuestas múltiples, combinada con la posibilidad de la elección libre por el alumno de la fecha del examen, disminuía drásticamente el número de no presentados, al permitir una

personalización de su proceso de enseñanza-aprendizaje (Torreblanca y otros, 1996).

La gran utilidad de este tipo de pruebas ha permitido el desarrollo de proyectos de investigación como "the Cooperative project on Evaluation of Results of Training" (CERT). En el seno de este proyecto se desarrolló un programa informático que apoya y facilita el uso de pruebas objetivas en la evaluación educativa, posibilitando gestionar y analizar gran cantidad de información recogida a través de pruebas objetivas de opción múltiple, agilizando el proceso de manipulación y cálculos manuales de los formadores para evaluar aprendizajes (De Pablos y otros, 1993). Una de las características principales de este programa es la inclusión de los denominados niveles de confianza que son una estimación que el alumno realiza en base a una escala de 0-5 del grado de confianza o seguridad con que responde a cada una de las cuestiones de la prueba (Ariza y otros, 1998).

Tomando como partida la idea del proyecto CERT hemos desarrollado un programa informático de evaluación asistida por ordenador. En nuestro programa hemos salvado los inconvenientes descritos por Arriaga y otros (1996) que exponen que el software educativo, normalmente por ser entornos cerrados en los que no es posible la modificación, no suscita el interés de los profesores por incorporarlos a la docencia. Nuestro primer objetivo fue verificar la validez del sistema de elaboración de exámenes, que como ya demostramos con anterioridad (Torreblanca y otros, 1996), obtenía resultados en la medición de cursos estadísticamente iguales a los que lograba el profesor gestionando personalmente la misma base de datos.

El programa, denominado **"EVA" (Evaluación Asistida por ordenador mediante pruebas objetivas de opción múltiple)**, permite un entorno totalmente abierto capaz de:

1. Almacenar preguntas de diferentes asignaturas en distintas bases de datos con su respuesta verdadera y sus distractores. Admite 3, 4 o 5 respuestas y la opción de verdadero o falso.
2. Clasificar las preguntas por temas.
3. Visualizar, modificar y borrar las preguntas almacenadas.
4. Generar pruebas de forma totalmente aleatoria en las que la posición de la respuesta verdadera y de los distractores es seleccionada de la misma manera. De forma que si la misma pregunta formara parte de dos pruebas diferentes sus opciones no tendrían por que ocupar la misma posición.
5. Visualizar en pantalla, pregunta a pregunta, la prueba generada indicando distractores y respuesta correcta.
6. Realizar la prueba en el ordenador.
7. Imprimir la prueba generada para que pueda ser realizada por la clase.
8. Imprimir la plantilla de respuestas correctas, que es generada y almacenada simultáneamente con cada prueba.
9. Corregir de forma automática con cualquiera de los cuatro criterios de evaluación de los que dispone el programa.
10. Gestionar los grupos de clase realizando: listados con los resultados de las pruebas, visualizando, añadiendo, modificando, borrando, buscando... alumnos. También incluye cálculo de porcentaje de la calificación y realiza copias de seguridad de los ficheros de asignaturas, alumnos y pruebas.

A pesar de estar en periodo de desarrollo, su funcionamiento es satisfactorio, al descargar de la tarea de la preparación y corrección de los exámenes. Además no se constituye como un entorno cerrado sino que estas bases de datos se pueden ir actualizando y aumentando continuamente, con lo que se incrementa la objetividad del sistema. El programa nos permite fijar, de manera automática, los niveles de dificultad de las diferentes preguntas, y de esta manera poder fijar la dificultad de exámenes posteriores.

Una de las cuestiones esenciales a la hora de calificar una prueba objetiva de respuesta múltiple es la eliminación de los efectos del azar en el cálculo numérico de la puntuación bruta de la prueba. La eliminación del efecto del azar se realiza empleando distintos criterios que contemplen una corrección de la puntuación final en función de la posibilidad de acierto de forma aleatoria de los distintos ítem.

En función de las distintas decisiones que adoptemos para corregir la influencia del acierto aleatorio en la cumplimentación de las pruebas objetivas, obtendremos distintas fórmulas

matemáticas que nos permitirán calcular la puntuación bruta de dichas pruebas. Este conjunto de decisiones y de fórmulas matemáticas es lo que denominamos criterios de evaluación o corrección.

Para la calificación el programa contempla cuatro criterios de corrección posibles, que pueden ser seleccionados en cualquier momento, permitiendo así la comparación de las calificaciones obtenidas por el alumno en una misma prueba.

Basándonos en esta cualidad hemos corregido un número determinado de exámenes usando los cuatro criterios posibles. A los resultados le hemos aplicado diversos test estadísticos y hemos realizado la comparación de los datos para tratar de establecer cuál es el más adecuado para:

a. discernir el grado de capacitación del estudiante.
b. aportar información sobre las diferencias generalizadas que permitan una modificación de las estrategias docentes del profesor.

2. MATERIAL Y MÉTODO

El sistema de evaluación asistida por ordenador mediante pruebas objetivas de opción múltiple ("EVA") se basa en el conocido programa de base de datos **dbase**. Como ya indicamos, dispone de cuatro formas distintas de medición dependiendo del criterio seleccionado. La descripción y formulación de los criterios es la siguiente:

2.1. PRIMER CRITERIO DE MEDICIÓN (Criterio Clásico)

En este criterio de corrección, que denominaremos **clásico**, se tienen en cuenta los aciertos, los errores y además, se ofrece la posibilidad de no responder a alguna pregunta, no interviniendo éstas, en el cálculo de la puntuación bruta. El efecto del azar se corrige utilizando la probabilidad de acierto de forma aleatoria en función del número de opciones o distractores de cada ítem, según empleemos unos u otros, dotaremos a la prueba de una mayor rigurosidad. Es decir, la puntuación bruta, asignando un punto por respuesta acertada, se obtiene restando al número de aciertos el número de errores multiplicado por la probabilidad de elegir un distractor como respuesta (número de distractores = opciones -1). La fórmula matemática (Rodríguez, 1980) para conseguir la puntuación bruta queda de la siguiente forma:

$$Puntuación = Aciertos - \frac{Errores}{Opciones - 1}$$

Esta calificación se puede suavizar en cierta medida si las omisiones, respuesta en blanco, se consideran un distractor más, lo que equipararía el número de distractores más la respuesta en blanco con el número total de opciones en cada ítem. Así, si el número de opciones a cada íteme es de **"n"**, el número de distractores será **"n-1"** y si la respuesta en blanco se considera un distractor más, el número de estos aumenta a **"n"** coincidiendo con el número de opciones, con lo cual, la probabilidad de elegir uno de ellos será de **"1/n"** en el primer caso y de **"1/(n-1)"** en el segundo, esto quiere decir que el factor **[Errores / n]** es menor que el factor **[Errores / (n-1)]** con lo cual la puntuación bruta aumenta. La fórmula matemática adopta la siguiente estructura:

$$Puntuación = Aciertos - \frac{Errores}{Opciones}$$

A igualdad de aciertos y errores la fórmula (1.2) da una mayor puntuación bruta que la fórmula (1.1). El paso de una a otra fórmula se obtiene introduciendo un nivel de rigurosidad **"r"** que podrá tomar los valores de **"1"** para el primer caso, fórmula (1.1), y de **"0"** para el segundo caso, fórmula (1.2). Con lo cual la puntuación bruta la obtendremos unificando en una (1.3) ambas fórmulas:

$$Puntuación = Aciertos - \frac{Errores}{Opciones - r}$$

En nuestro sistema de enseñanza, es costumbre generalizada otorgar la calificación final sobre un máximo de 10 puntos pero, en algunos casos, la prueba puede ser un porcentaje de la calificación final. Ambos casos son contemplados por el programa que emplea, de forma automática, un factor de conversión **(k)** entre la puntuación bruta y la calificación máxima que elija el profesor que realice la prueba. El programa permite que dicha nota máxima tome valores comprendidos entre 1 y 100.

La fórmula matemática que nos proporciona la calificación final vendrá dada por el producto de la puntuación bruta por el factor de conversión K:

$$Calificación = Puntuación \cdot k = [Aciertos - \frac{Errores}{Opciones - r}] \cdot k$$

Siendo el valor de k:

$$k = \frac{N_{máxima}}{N_{ítemes}}$$

donde:

N máxima = Nota máxima elegida por el profesor.
N ítemes = Número de ítem de la prueba.

El **N ítemes** de la prueba coincidirá con la puntuación máxima posible ya que estamos asignando un punto por ítem.

SEGUNDO CRITERIO DE MEDICIÓN. (Porcentaje Puro)

En este criterio el efecto de acierto por azar se corrige restando al número de aciertos el número de preguntas que pueden ser acertadas de forma aleatoria. Así, en una prueba objetiva con cuarenta ítem y cuatro opciones de respuesta por ítem tendremos que el número de ítemes que pueden ser acertados de forma aleatoria es de 40/4 = 10. Es decir, el número de ítemes de la prueba multiplicado por la probabilidad de acertar cada ítem. Seguimos asignando un punto a cada ítem:

$$Puntuación = Aciertos - \frac{N_{ítemes}}{Opciones}$$

La máxima puntuación bruta posible se deberá corresponder, no al número de ítem de la prueba sino al mayor número de ítem que puedan ser acertados, al número de ítem de la prueba menos el número de ítem multiplicado por la probabilidad de acertar cada uno. En el ejemplo anterior de 40 ítem y cuatro respuestas por ítem la mayor puntuación posible (Puntuación máxima) será:

$$P_{máxima} = N_{ítemes} - \frac{N_{ítemes}}{Opciones} = 40 - \frac{40}{4} = 40 - 10 = 30$$

Si introducimos el nivel de rigurosidad **"r"** con valores 0 y 1, igual que en el criterio anterior y considerando a la respuesta en blanco como una opción más, para el caso de menor rigurosidad

tendremos que la puntuación pasaría a ser:

$$Puntuación = Aciertos - \frac{N_{Ítemes}}{Opciones + 1}$$

Que es ligeramente superior a la obtenida con la fórmula (2.1). Combinando ambas ecuaciones nos quedará:

$$Puntuación = Aciertos - \frac{N_{Ítemes}}{Opciones + (1 - r)}$$

En estas condiciones el factor **"K"** de conversión a la nota máxima elegida por el profesor quedará de la siguiente forma:

$$K = \frac{N_{máxima}}{N_{Ítemes} - \frac{N_{Ítemes}}{Opciones}}$$

Y la calificación final quedará:

$$Calificación = Puntuación \cdot k$$

Debemos hacer notar que el factor **"r"** no se introduce en el factor de conversión K ya que el aumento producido al aplicarlo en el cálculo de la puntuación bruta sería parcialmente compensado si se aplica también al cálculo de la puntuación máxima.

TERCER CRITERIO DE MEDICIÓN. (Niveles de dificultad)

En este criterio, el factor aleatorio lo eliminamos de igual forma que en el primer criterio pero teniendo en cuenta que cada pregunta va a tener un nivel de dificultad que puede variar desde **"1"**, máxima dificultad, hasta **"5"**, mínima dificultad, y que por tanto, a diferencia de los dos anteriores criterios, cada pregunta, acertada o fallada, tiene un valor distinto en función de dicho nivel de dificultad. El valor que se le ha asignado a cada pregunta lo podemos observar en la tabla siguiente:

	NIVELES DE DIFICULTAD				
	Nivel 1	**Nivel 2**	**Nivel 3**	**Nivel 4**	**Nivel 5**
Respuesta Correcta	1	0.9	0.8	0.6	0.4
Respuesta Errónea	0.4	0.6	0.8	0.9	1

Es decir una pregunta acertada en el **nivel 1** tiene un valor de **1** acierto, mientras que si la pregunta es fallada en el mismo nivel su valor es de **0.4** errores.

El cálculo de la puntuación correspondiente al número de aciertos y de errores se realizará por medio de la expresiones:

$$Puntuación_{Aciertos} = \sum_i Puntuación_{NivelAciertoi}$$

$$----0--0----$$

$$Puntuación_{Errores} = \sum_i Puntuación_{Nivelerroresi}$$

Es decir la puntuación correspondiente a los aciertos se obtiene sumando el valor correspondiente al nivel de cada acierto y la de los errores sumando el valor correspondiente al nivel de cada error. Por tanto la puntuación bruta se obtiene restando a la puntuación correspondiente a los aciertos la puntuación correspondiente a los errores multiplicada por la probabilidad de elegir un distractor como respuesta (número de distractores = opciones -1).

La fórmula matemática para calcular la puntuación bruta queda de la siguiente forma:

$$Puntuación_{Bruta} = Puntuación_{Aciertos} - \frac{Puntuación_{Errores}}{Opciones - 1}$$

También se introduce, al igual que en los criterios anteriores, un nivel de rigurosidad **"r"** que hace que la calificación sea más o menos rigurosa.

La expresión matemática que nos proporciona la calificación final vendrá dada por el producto de la puntuación bruta por el factor de conversión k:

$$Calificación = Puntuación_{Bruta} \bullet k =$$

$$-----$$

$$= [Puntuación_{Aciertos} - \frac{Puntuación_{Errores}}{Opciones - r}] \bullet k$$

Siendo el valor de K:

$$K = \frac{N_{máxima}}{P_{máxima}}$$

donde:

Nmáxima = Nota máxima elegida por el profesor.
Pmáxima = Puntuación máxima de la prueba.

CUARTO CRITERIO DE MEDICIÓN. (Niveles de Confianza CERT)

En este criterio, la corrección de los efectos del azar son dejados al alumno estableciendo los llamados niveles de confianza, puntuación de 0 a 5 que el alumno le asigna a cada una de sus respuestas, indicando con ello, la confianza que tiene en responder correctamente las preguntas.

El **"0"** indica que el alumno no tiene ninguna confianza en responder correctamente a la pregunta es decir, que está contestando completamente al azar. El **"5"**, por el contrario, que tiene plena confianza en la respuesta y las otras puntuaciones indican una confianza intermedia entre ambas.

El programa **"CERT"** ofrece inicialmente un baremo para la corrección (Alonso, 1992) conforme a la teoría de decisiones y a las probabilidades subjetivas. En función del nivel de confianza elegido al contestar la pregunta el valor de esta, correcta o erróneamente contestada será uno u otro como podemos observar en la tabla adjunta.

	NIVELES DE CONFIANZA					
	0	**1**	**2**	**3**	**4**	**5**
Respuesta Correcta	13	16	17	18	19	20
Respuesta Errónea	4	3	2	0	-6	-20

Debido a la diferente puntuación obtenida en la respuesta, tanto negativa como positiva, dependiendo del nivel de confianza elegido es de vital importancia que el alumno tenga interés en decir la verdad, es decir expresar su confianza sin deformarla.

El proyecto **"CERT"** para calcular la nota del alumno utiliza la fórmula siguiente:

$$Puntuación_{Cert} = \frac{Puntuación_{Bruta} \bullet 20}{Rigurosidad \bullet Número_{items}}$$

Donde:

$$Puntuación_{Bruta} = Puntuación_{aciertos} + Puntuación_{Errores}$$

La puntuación final la realiza sobre 20,(factor correctivo 20/Númeroitem) para adaptarla a la puntuación tradicional española sobre 10 tendremos que sustituir el factor correctivo (20 / Númeroitem) por (10 / Númeroitem).

Si en vez de realizar esta sustitución introducimos la variable **Nota máxima** que nos pedirá el programa a la hora de elegir el criterio de evaluación, el factor de corrección de la nota queda de forma variable y sirve para cualquier nota que elija el profesor. Con lo cual, el programa **"EVA"** calcula la nota con la siguiente fórmula matemática:

$$Puntuación_{Cert} = \frac{Puntuación_{Bruta} \bullet Nota_{máxima}}{Rigurosidad \bullet Número_{items}}$$

En esta caso la rigurosidad, es un factor que varía de 16 a 20 y lo que hace es disminuir o aumentar la puntuación necesaria para obtener el aprobado.

La puntuación máxima posible, se calcula sobre la hipótesis de que todas las preguntas son respondidas correctamente con el máximo nivel de confianza. Por tanto, para obtener una puntuación media habría que dividir por 20, que es la máxima puntuación posible por ítem, si dividimos por 16, 17, 18 o 19 en vez de por 20 estamos aumentado la puntuación media, que es la que nos va a proporcionar la calificación final sobre la nota máxima al multiplicarla por el factor de corrección**(Notamáxima/Númeroitem)** .

Cuando empleamos un factor de rigurosidad menor que 20 un alumno puede obtener una calificación final mayor que la **Notamáxima**. Esta "anomalía"también es corregida por el programa **"EVA"**que califica con la Nota máxima cuando esta es superada. También elimina la puntuación negativa sustituyendola por **"0"**. Esta sustitución se realiza en todos los criterios.

Debido a las diferencias de los 4 criterios de corrección utilizados, para obtener un grupo de exámenes que pudieran ser corregidos con todos ellos en similares circunstancias, hemos tenido que hacer algunas concesiones como las siguientes:

1. Para salvar las dificultades que plantea la aplicación del 4º criterio de corrección, como material sólo se utilizaron aquellos exámenes (40 preguntas de 4 respuestas cada una) de alumnos de la asignatura de Biología Celular y General del primer curso de la Especialidad de Maestro de Educación Primaria correspondientes al curso académico 1998/99 que habían rellenado voluntariamente el nivel de confianza correspondiente de cada pregunta al conocer que no les afectaría para el cálculo de la nota del ejercicio.
2. Para cubrir la posibilidad de las respuestas en blanco en el primer criterio y la conveniencia de contestar todas las preguntas en el 2º, se han considerado como no respondidas en aquel todas las preguntas con nivel de confianza 0.

Pensamos que las posibles distorsiones que pudieran producir estas acotaciones son pequeñas en comparación con la ventaja que supone tener un grupo homogéneo a estudiar.

RESULTADOS y DISCUSIÓN.

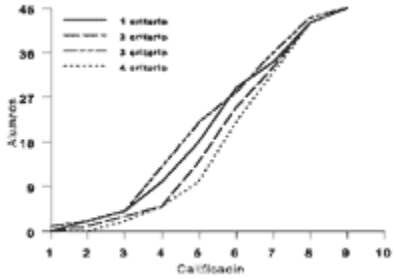
Los exámenes de los 45 alumnos que habían cumplido los requisitos anteriormente mencionados, fueron corregidos con los cuatro criterios explicados anteriormente, dando como resultado las notas medias que aparecen en la Tabla I.

Tabla I.

Criterio	1°	2°	3°	4°
Media	5.46	5.91	5.22	5.96
Error Estándar	0.27	0.24	0.27	0.21
N	45	45	45	45

La aplicación del test T de Student de comparación de medias demostró que todos los criterios medían de igual manera al conjunto.

Por tanto se podría concluir que todos los criterios utilizados son igualmente válidos si de lo que se trata es de medir al conjunto. No obstante uno de los objetivos de evaluar es clasificar a los individuos. Por esta razón creemos que el criterio que produzca una mayor dispersión del grupo es mejor que el que concentre las notas del conjunto. Según esta perspectiva el criterio más eficaz sería el 3º, pues las calificaciones de los 45 exámenes corregidos se distribuyen más ampliamente que en el resto de los criterios y más uniformemente (ver gráfico).



No obstante creemos que a la hora de evaluar no sólo hay que fijarse en el grupo sino que hay que tratar de acercarse a cada alumno para que sea el propio individuo el que nos muestre su grado de capacitación. Para cubrir este objetivo y poder discernir qué criterio es más efectivo, hemos utilizado el test T de Student de datos apareados. Como se aprecia en la Tabla II existen 2 bloques formados por los criterios 1º y 3º, que estadísticamente son iguales entre sí y por los criterios 2º y 4º a los que les ocurre igual, siendo los componentes de ambos bloques distintos a los que componen el otro.

Tabla II

Criterio	1º	2º	3º
2º	★		
3º	iguales	★	
4º	★	iguales	★

Test t de Student de comparación de medias, para P 0.05

Este emparejamiento es lógico, por ejemplo, entre el criterio 1º y 3º hay relación pues los niveles de dificultad los obtiene el ordenador atendiendo al número de individuos que responden adecuadamente a una determinada pregunta. Hay pues una relación entre el grado de dificultad de una pregunta y el número de respuestas incorrectas obtenidas para la misma dentro del grupo.

La diferencia entre estos dos criterios reside en que el 1º mide atendiendo exclusivamente a la respuesta del individuo en el examen, por su parte en el 3º esto se ve modificado por el nivel del grupo en cuestión. Es decir el individuo es calificado teniendo en cuenta el grupo en el que se encuentra: se tienen en cuenta las circunstancias.

Los criterios 2º y 4º podrían considerarse semejantes por las siguientes razones: en el 2º no se tienen en cuenta las respuestas erróneas y aunque en el 4º si lo son, sólo representan un descenso en la calificación en 2 de los 10 casos posibles, por lo que las semejanzas en el tratamiento de la medición podrían considerarse elevados. Esta semejanza es mayor si tenemos en cuenta que es necesaria la respuesta veraz de los niveles de confianza, ya que un sesgo hacia el nivel 3 supondría una mayor aproximación de ambos criterios puesto que la respuesta errónea no tendría ningún efecto sobre la calificación. Estas dificultades en la aplicación del criterio 4º (necesidad de responder con total veracidad a los niveles de confianza y la posibilidad del sesgo en la respuesta para obtener los mejores resultados) aconsejan modificar los valores que el proyecto CERT utiliza en este criterio. En nuestro caso ya estamos realizando estudios previos para tratar de evitar estas dificultades.

Por todo lo expuesto anteriormente, aunque los 4 criterios son válidos creemos que el tercero, por ampliar la dispersión de la medición del grupo y por calificar al alumno en relación con el grupo en el que se ha producido su proceso de enseñanza-aprendizaje, es el más adecuado.

Referencias bibliográficas.

ALONSO, C. Y OTROS. (1992). Principios comunes para la evaluación de los resultados cognitivos de la formación. Comisión de las Comunidades Europea. **Programa Eurotecnet.**

ARIZA, A. Y OTROS. (1998). CERT: Un modelo matemático y tecnológico de evaluación. **Pixel-Bit. Revista de Medios y Educación, 11**. 51-56

ARRIAGA, J. Y OTROS. (1996). Sistemas de autor orientados a un fin educativo específico. **Pixel-Bit. Revista de Medios y Educación, 6**. 5-13

DE PABLOS Y OTROS. (1993). La evaluación del alumno en la universidad: El proyecto CERT. **Revista de Enseñanza Universitaria, 6.** 49-71.

DE JUAN HERRERO, J. (1996). **Introducción a la Enseñanza Universitaria. Didáctica para la**

Formación del Profesorado. Madrid, Dykinson.

FERMÍN, M. (1971). **La evaluación, los exámenes y las calificaciones**. Buenos Aires, Kapelusz.

GUILBERT, J.J.(1994). **Guía Pedagógica para el personal de salud.** Organización Mundial de la Salud (OMS) ICE de la Universidad de Valladolid.

HUBBARD, J.P. (1978). **Measuring medical education. The test and the experience of the National Board of Medical Examiners.** Philadelphia, Lea and Febiger.

LAFOURCADE, P.D. (1977). **Evaluación de los aprendizajes**. Madrid, Cincel.

RODRÍGUEZ, J.L. (1980). **Didáctica General**. Madrid, Cincel-Kapelusz.

THYNE, J.M.. (1978). **Principios y técnicas de examen.** Salamanca, Anaya.

TORREBLANCA, J.; SANCHO, M. y ARIZA, A.(1996) La utilización de bases de datos como herramientas de evaluación. **Pixel-Bit. Revista de Medios y Educación. 7.**73-82.