

Todo lo que siempre quiso saber sobre la evaluación, pero no se atrevió a preguntar: a propósito de una encuesta

J. De Juan*; R.M. Pérez-Cañaveras**; J.L. Girela*; N. Martínez-Ruiz*; J.L. Soto*; A. Castillejo*; Y. Segovia*; C. Soto*; D. Torrus*; M.F. Vizcaya**; A. Romero*; M.J. Gómez-Torres*; J. Herrero*; A. Martínez*

** Dpto. de Biotecnología. Universidad de Alicante*

*** Dpto. de Enfermería. Universidad de Alicante.*

RESUMEN

La evaluación es un proceso sistemático, continuo e integral destinado a determinar hasta qué punto han sido alcanzados los objetivos educativos. En ella se aúnan dos actividades fundamentales: medir y emitir juicios de valor a partir, de los datos. Sin embargo, a pesar de la claridad de estos conceptos la forma de evaluar a los alumnos todavía sigue plagada de importantes contradicciones entre los profesores, lo que manifiesta unas conductas didácticas alejadas de lo que debiera ser una pedagogía basada en la evidencia. El objetivo de este trabajo es evidenciar las citadas discrepancias, entre los agentes del proceso de enseñanza aprendizaje (alumnos y profesores), a la hora de elaborar, realizar y calificar las pruebas de evaluación (“exámenes”) de los alumnos de varias disciplinas relacionadas con la Biomedicina. La metodología empleada ha consistido en aplicar a los estudiantes un sencillo cuestionario con preguntas de elección múltiple. A partir de los resultados, se pretende reflexionar sobre conjunto de criterios sólidos que permitan realizar las evaluaciones de los alumnos con mayor validez, fiabilidad, objetividad y pertinencia, como recomienda la OMS en su Guía Pedagógica para el Personal de Salud.

Palabras Clave

Evaluación del alumno, tipos de evaluación, métodos y técnicas de evaluación, ideología y evaluación,

1. INTRODUCCIÓN

1.1. La evaluación en el proceso de enseñanza aprendizaje

Para Peterssen (1976) el aprendizaje es la modificación de la conducta de los individuos, debido a una serie de experiencias. Para él la conducta es algo perceptible y observable, lo que permite distinguir entre una conducta inicial y una conducta final (Figura 1) en el proceso de enseñanza-aprendizaje (PEA).

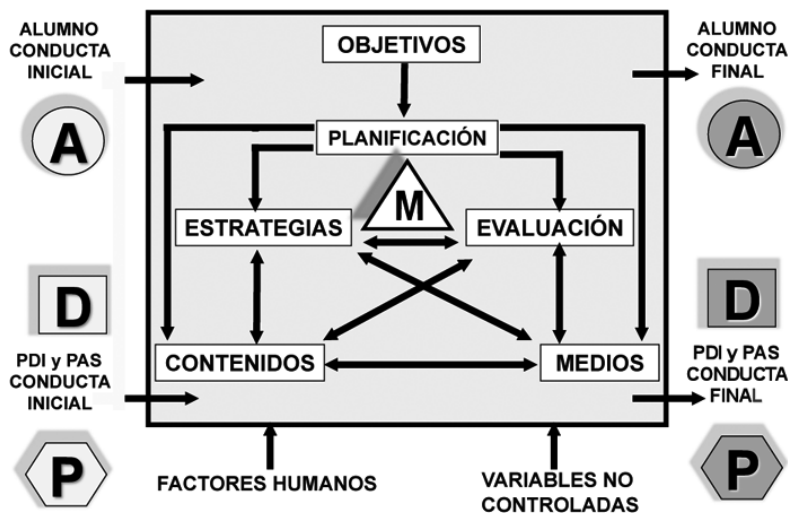


Figura 1: Estructura del PEA. Los círculos (A) representan a los alumnos, los cuadrados al personal docente (D) y los hexágonos al Personal de Administración y Servicios (P). El gris más intenso de los círculos, cuadrados y hexágonos, indican el cambio de conducta. Los objetivos, planificación, contenidos, medios, estrategias o métodos docentes y la evaluación, son los elementos internos básicos del PEA. Los factores humanos y las variables no controladas representan los componentes externos del PEA. Las flechas indican la interacción entre todos los elementos del PEA. El recuadro grande de fondo gris representa el contexto o entorno de aprendizaje (De Juan y Pérez-Cañaveras, 2011).

La participación de los alumnos, del personal docente y del PAS en el PEA, determinan cambios en la conducta de todos ellos, especialmente en los estudiantes. El paso de una conducta a otra se produce cuando el individuo tiene experiencias o interacciones con el ambiente (Skowronek, 1970). Sin embargo “la conducta observable no es idéntica al aprendizaje...”, es entre la conducta inicial y la final cuando se produce el auténtico aprendizaje, o sea, el cambio en las disposiciones de la conducta. Por eso hablamos de proceso de enseñanza/aprendizaje.

Previamente (De Juan et al. 1978; De Juan y Pérez 1991; De Juan, 1996, De Juan y Pérez-Cañaveras, 2011) hemos distinguido una serie de elementos internos al PEA, a saber: los *objetivos*, a partir de los cuales se realiza la *planificación* de los *contenidos*, de los *medios*, de las *estrategias* o *métodos docentes* y de la *evaluación*. Todos estos elementos tienen lugar en un *contexto* (Figura 1) o *entorno* (Vizcaya et al, 2004) que también influye, de manera decisiva, en los cambios de conducta de las personas implicadas en el PEA. Igualmente, debemos considerar otros elementos exógenos al PEA que de manera indirecta, también inciden en él, como ocurre con los factores humanos (inteligencia, motivación, salud, situación familiar, etc.) y las variables no controladas (acontecimientos políticos, culturales, catástrofes, etc.)

1.2. El problema de la evaluación

1.2.1. Concepto de evaluación

La evaluación es "un proceso sistemático, continuo e integral destinado a determinar hasta qué punto han sido alcanzados los objetivos educativos" (Fermín, 1971). Son numerosas las obras y los trabajos especializados relacionadas con la evaluación. Un compendio de los más clásicos pueden verse en De Juan (1996).

Uno de los aspectos importantes de la evaluación, y que a veces se confunde con ella misma, es la medición. Medir, es la acción de recoger información y ordenarla, teniendo en cuenta un aspecto cuantitativo numérico. Si bien la medición es una fase previa que proporciona objetividad a la evaluación, ésta sobrepasa la mera medición al introducir *juicios de valor* sobre la información obtenida (De Juan, 1996).

Son muchos los tipos de evaluación utilizados en el proceso de enseñanza/aprendizaje, la mayoría de los cuales se hallan resumidos en el Tabla 1. De todos los "objetos" a evaluar en el PEA, aquí nos centramos en la evaluación de los estudiantes.

Tabla 1. Tipos de evaluación (De Juan, 1996)

Según el objeto evaluado: <ul style="list-style-type: none">• Evaluación del sistema evaluativo considerado globalmente.• Evaluación de la administración.• Evaluación de los docentes.• Evaluación de los discentes.• Evaluación de los procedimientos de enseñanza (estrategias).• Evaluación de las instalaciones.• Evaluación de la evaluación (objeto de este trabajo)
Según la interpretación de los resultados: <ul style="list-style-type: none">• Evaluación normativa.• Evaluación por criterios.
Según el momento del PEA en el que se aplica y su función: <ul style="list-style-type: none">• Evaluación diagnóstica ("pre-test" y "test de pre-requisitos")• Evaluación formativa.• Evaluación sumativa o sancionadora

1.2.2. Tipos de evaluación.

A) Según el momento del Proceso de Enseñanza/Aprendizaje en el que se evalúa:

Atendiendo a este criterio podemos distinguir tres tipos distintos de evaluación (Iñiguez et al., 1979; Guilbert, 1994; De Juan, 1996):

a) Evaluación sumativa:

Esta evaluación (Scriven, 1967; Iñiguez, 1979; Guilbert, 1994 y De Juan, 1996), también conocida como acumulativa, sancionadora o de certificación y se realiza al final de un PEA. Su objetivo es la valoración final de la tarea realizada a lo largo del PEA y su importancia radica en

que manifiesta el grado de consecución de los objetivos educativos que se pretendían alcanzar. Se emplea tradicionalmente para clasificar a los estudiantes y tomar decisiones sobre el paso de éstos al curso siguiente así como para decidir la obtención de un título o diploma.

b) Evaluación diagnóstica:

Se realiza al principio del PEA, es decir, antes de que éste se ponga en marcha, con el fin de determinar si los alumnos dominan una serie de objetivos. Dentro de la evaluación diagnóstica podemos distinguir dos modalidades diferentes: los "pre-test" y los "test de pre-requisitos" (Iñiguez, et al. 1979; De Juan, 1996).

Los *pre-test* son pruebas destinadas a medir los conocimientos que poseen los alumnos sobre la materia que se pretende enseñar. En este tipo de evaluación, los alumnos son sometidos, al principio de un curso o de una unidad de aprendizaje, a un examen cuyos contenidos son análogos a los del examen final. Su interés radica en que proporcionan al docente información acerca de los objetivos pedagógicos del curso que ya son suficientemente conocidos antes de comenzar el mismo, lo que permite obviarlos.

Los *test de pre-requisitos* o test de conocimientos previos son aquellas pruebas tendentes a manifestar si los alumnos poseen las bases mínimas indispensables para abordar con éxito la materia (Núñez, 1977). Su importancia y valor intrínseco, dentro del sistema educativo, está ampliamente reconocida en la literatura clásica (Domingues, 1977; Iñiguez et al., 1979; Lafourcade, 1974; Lafourcade, 1977; Livas, 1977; Núñez, 1977; Sobrado, 1977). La información que suministran es muy útil para el profesorado, el alumnado y las instituciones docentes. A través de ellos, los alumnos obtienen una idea clara de sus conocimientos reales necesarios para emprender con éxito el nuevo PEA. La utilidad del test de pre-requisitos para el profesor radica en que éste le permite planificar el curso. En efecto, en lugar de utilizar para ordenar y agrupar a los estudiantes, utilizando el orden alfabético, el número de matrícula o el número del DNI, el test de pre-requisitos permite hacerlo basándose en su capacidad, lo que convierte a estas pruebas diagnósticas en el único criterio técnico válido para ello (Livas, 1977; Iñiguez et al. 1979). Siendo también muy útil para discernir qué alumnos están capacitados para ingresar en un determinado centro para realizar determinados estudios (Iñiguez et al. 1979).

c) Evaluación formativa:

La evaluación formativa (Scriven, 1967) consiste en el conjunto de pruebas que se realiza durante el PEA, con el fin de determinar los logros que se van alcanzando. Juega un importante papel de control retroactivo para el profesor y para el alumno, indicándole al primero si debe modificar sus estrategias de enseñanza, e incluso sus objetivos pedagógicos y proporcionándole al segundo información sobre los logros alcanzados y orientación para establecer un plan de estudio.

B) Según la interpretación de los resultados:

Atendiendo a este criterio podemos distinguir (De Juan, 1996; Glaser, 1963 y Livas, 1977) dos tipos de evaluación claramente definidos: evaluación por normas y evaluación por criterios.

a) En la *evaluación por normas* lo que se pretende medir son los conocimientos que posee un determinado alumno respecto a los conocimientos de los demás. Las pruebas diseñadas para evaluar según este método, constan de muchas preguntas cuyo grado de dificultad es muy variable, explorando lo básico y lo accesorio de los objetivos. Consecuencia de ello, es que los rendimientos de los alumnos se distribuyen con arreglo a una curva normal, valorándose las diferencias individuales más sutiles, lo que permite clasificarlos según diferentes grados de aptitud (sobresaliente, notable, etc.). La toma de decisiones se ve muy influida por el rendimiento medio del grupo, por lo que el criterio de evaluación es relativo y modificable "a posteriori".

b) En la *evaluación por criterios* lo que se pretende medir son los conocimientos del alumno respecto a lo que se considera que debe saber. En el diseño de las pruebas se calcula el número de preguntas necesario para medir el nivel suficiente de logro en los objetivos educativos, por lo que la dificultad de las preguntas viene determinada por el *Nivel Aceptable de Rendimiento* o NAR (Nedelsky, 1954). La distribución de los alumnos es binomial, en **apto** y **no apto**. La toma de decisiones se basa, por tanto, en el dominio razonable de los objetivos educativos y el criterio de evaluación que es estipulado "a priori" y no es influenciado después. En esta evaluación se admite una variabilidad en los aciertos de un 10 a un 20 % por debajo de lo esperado. Para algunos autores (Turnbull, 1989) este debe ser el método ideal, al menos en algunos centros, como por ejemplo en una Facultad de Medicina cuyo objetivo es formar buenos profesionales y no clasificarlos en orden a su rendimiento académico en base a conocimientos superfluos.

1.2.3. Pruebas de evaluación.

A) Exámenes escritos objetivos ("test"):

La gran difusión adquirida por este tipo de pruebas, y más concretamente por las preguntas de elección múltiple (PEM), se debe fundamentalmente a que se hallan libres de muchos de los defectos atribuibles a otros tipos de exámenes (Lafourcade, 1977). El uso, de PEM ha conseguido un alto grado de aceptación por parte de profesores y alumnos. Esta aceptación se debe, entre otras razones, a su gran fiabilidad y validez, puesta de manifiesto por numerosos autores (De Juan, 1996).

Además de las críticas que se pueden formular a todas las pruebas de "papel y lápiz", las llamadas "pruebas objetivas" poseen dos importantes inconvenientes: su incapacidad para medir objetivos del pensamiento creador y la dificultad de su construcción. Sin embargo, muchas de las críticas que se hacen a las pruebas objetivas solo son atribuibles a aquellas construidas de modo deficiente (Beard, 1974) o bien debidas a meras opiniones y no basadas en experiencias llevadas a cabo imparcialmente (Lafourcade, 1977). Muchos de los detractores de este tipo de pruebas olvidan que el examen objetivo constituye fundamentalmente la fase cuantitativa del proceso global de

evaluación y que, como instrumento de medida, es mucho más válido y fiable que un examen oral o escrito. Respecto a las características técnicas, modos de construcción, etc., de este tipo de pruebas, puede consultarse en Gilbert, 1994, Lafourcade, 1974 y Lafourcade, 1977.

B) Análisis crítico de los exámenes "tradicionales".

Muchas son las críticas que el tradicional sistema de exámenes ha recibido en todos los tiempos. Una de las quejas más frecuentes, es la consideración de que los exámenes no suelen tener en cuenta lo que es importante en el aprendizaje, así como su falta de fiabilidad, lo que se debería, entre otros, a los siguientes factores (De Juan, 1996): a) Falta de objetividad del examinador, cuyos criterios pueden cambiar, por ejemplo con el número de exámenes corregidos, b) Diferencias, en las calificaciones, de un examinador a otro y c) Distinta respuesta del examinador debidas a factores accidentales (la caligrafía, la redacción, etc.). Estos y otros muchos factores (fatiga, errores del profesor, etc.) han sido las causas de las abundantes críticas surgidas en torno al examen tradicional.

1.3. Propósito

Evidentemente, el título de este trabajo es una formula retórica utilizada para agudizar la atención sobre el proceso evaluador. Realmente, este es un ensayo piloto cuyo objetivo fundamental es crear las bases de un estudio más amplio y profundo a cerca de las opiniones que los alumnos tienen de las evaluaciones a las que se ven sometidos durante sus estudios. Se trata de buscar las características más relevantes que nos permitan profundizar en una evaluación basada en la evidencia y no en una tradición acrítica de conductas estereotipadas.

2. METODOLOGÍA

2.1. Diseño de la investigación:

Con el fin de conocer las opiniones que los estudiantes de asignaturas del área de conocimiento de Biología Celular, tienen sobre los métodos y pruebas de evaluación, elaboramos un cuestionario con 20 PEM sobre los siguientes aspectos de la Evaluación: (1) Valoración de los diferentes tipos de exámenes, (2) Modo de realizar los exámenes y (3) Opinión sobre la evaluación sancionadora.

2.2. Sujetos de estudio y características de la muestra:

La muestra fue de 128 alumnos de Biología de la Facultad de Ciencias. Alumnos del primer curso del Grado de Biología y de quinto curso de la Licenciatura de Biología. La selección se realizó mediante muestreo no probabilístico accidental al aplicarles el cuestionarios en el aula, en

horario de clase, dándoles las instrucciones pertinentes. Todos los entrevistados dieron su consentimiento para ser incluidos en el estudio,

2.3. Instrumento de estudio:

Se trata de un cuestionario, creado por nosotros, formado por 20 PEM, con cinco alternativas, a elegir una (Tabla 2). Las respuestas de las PEM, aunque ordenadas, no estaban polarizadas como en las escalas de Likert.

2.4. Recogida de datos y análisis del material:

Una vez cumplimentados los cuestionarios, se obtuvieron y analizaron los datos. Para tener una idea global del peso de las variables empleadas (sexo, grupo de alumnos y las puntuaciones dadas a las 20 preguntas del cuestionario) realizamos un Análisis Factorial de Componentes Principales con rotación varimax.

3. RESULTADOS

Dado su carácter exploratorio, presentamos los resultados de forma meramente descriptiva, ajustados a las preguntas de la encuesta (Tabla 2).

3.1. Valoración por su bondad, eficacia y fiabilidad de los diferentes tipos de “exámenes”:

3.1.1. Valoración de las pruebas de tipo “Ensayo” con Preguntas de Redacción Abierta Largas (PRAL) y Cortas (PRAC):

Como se refleja en el Tabla 2 (Pregunta 1) las pruebas con PRAL son valoradas como “regulares”, por casi un tercio de los estudiantes y como “buenas” por casi otro tercio de ellos. Es decir, algo más del 60% las consideran entre regulares y buenas y casi un 30% las valoran como muy poco importantes o nulas, siendo un número exiguo de ellos los que piensan que son “muy buenas”. En general el número de las alumnas que las valoran como “buenas” es mayor (42%) que el de los alumnos que tienden a considerarlas como “regulares” en un 38%. La mayor preferencia de las mujeres por las PRAL, aunque discreta, se corresponde con el mayor dominio del lenguaje por parte de ellas.

Respecto a las PRAC (Tabla 2, Pregunta 2), más de la mitad de los encuestados (54%) las valoran como “buenas” aunque el número de las mujeres que así lo hacen, es bastante mayor (63%) que el de los hombres (40%). Probablemente por las mismas razones aludidas en el párrafo anterior.

3.1.2. Valoración de las pruebas de tipo “Test” con Preguntas de Elección Múltiple (PEM):

En el Tabla 2 (Pregunta 3) observamos como la gran mayoría de los encuestados (72%) valoran los “exámenes” con PEM como “buenos” (37%) o “muy buenos” (35%). En este tipo de exámenes el

número de hombres que los valoran como “buenos” (44%) o “muy buenos” (42%) es del 86% frente a un 63% de las mujeres (32%, buenos y 31%, muy buenos). Solo alrededor de una cuarta parte de los alumnos las valoran como “nulas” (1,7%), “muy poco” (6%) o “regulares” (19%). En resumen el porcentaje de alumnos que las valora por encima de las PRAL y PRAC es mayoritario.

3.1.3. Valoración de las pruebas de tipo oral:

El número de alumnos que consideran las pruebas orales (Tabla 2, Pregunta 4) como “buenas” o “muy buenas” es pequeño (14%), siendo algo mayor en las mujeres (11%) que en los hombres (6%). De nuevo observamos una tendencia a valorar más el uso del lenguaje por parte de las mujeres. Sin embargo, el uso de la expresión oral, adquiere mucha mayor relevancia cuando es utilizado por los estudiantes para presentar los trabajos realizados a lo largo del curso (Tabla 2, Pregunta 8). En efecto, casi la mitad de los estudiantes (49%) la valoran como “buena” o “muy buena”, alcanzando el 55% en las mujeres frente al 40% de los hombres. Nuevamente observamos la preferencia de la comunicación oral en la mujer.

3.1.4. Valoración de las actividades prácticas:

El número de alumnos que las consideran “buenas” o “muy buenas” no alcanza a la mitad (45%), salvo en el caso de las mujeres que sube discretamente hasta un 51%, mientras que en los hombres baja hasta un 38%. El resto, dispersan su opinión entre una valoración “nula”, “muy poca” o “regular” (Tabla 2, Pregunta 5). Al parecer éste tipo de pruebas no despiertan mucho entusiasmo entre los alumnos, probablemente dada la diversidad de formas existentes. Resultados parecidos se obtienen cuando se les pregunta por la evaluación de las manipulaciones en las tareas de laboratorio (Tabla 2, Pregunta 10). Para casi la mitad de los alumnos (49%), este tipo de evaluación es “bueno” o “muy bueno”, siendo más alta en los varones (53%) que en las mujeres (45%). ¿Podría indicar esto una mayor preferencia de los hombres por las manipulaciones técnicas?

3.1.5. Valoración de trabajos escritos:

El portafolios se ha convertido en un instrumento más para evaluar a los estudiantes. Cuando se les pregunta por su importancia (Tabla 2, Pregunta 6) vemos que algo más de la mitad de los alumnos (52%) lo consideran como un método “bueno” o “muy bueno” Esta valoración es mayor en las alumnas (58%) que en los alumnos (42%), sugiriendo una mayor valoración del lenguaje escrito, por las mujeres.

Lo mismo sucede con la realización de trabajos escritos sobre temas de la materia de estudio (Tabla 2, Pregunta 7). En éste caso el 66% de los estudiantes los valoran como “buenos” o “muy buenos”, siendo mayor el porcentaje en las mujeres (59%) que en los hombres (42%). No creemos exagerar si decimos que existe una clara tendencia a valorar más positivamente las evaluaciones basadas en el lenguaje escrito u oral, por parte de las mujeres respecto de los hombre.

3.1.5. Valoración de la realización de problemas:

Los estudiantes valoran muy positivamente (71%) como “bueno” o “muy bueno” la realización de problemas como método de evaluación. Esto es así tanto para las mujeres (76%) como para los hombres (61%).

3.2. Valoración de las formas de realizar de pruebas de evaluación:

3.2.1. Duración máxima de un examen:

En la pregunta 11 hemos recabado información sobre cual es la máxima duración que debería tener un examen. Para casi un 90% de los estudiantes, una prueba debería tener entre 2 y 3 horas como máximo. Valores que alcanzan el 93% en las mujeres y del 83% en los hombres.

3.2.2. Número de exámenes:

Para el 61% de los estudiantes (Tabla 2, Pregunta12) el número de evaluaciones a realizar, serian 2 ó 3. El porcentaje de mujeres que apoyan esta opción es del 58% frente al 76% de los varones. No obstante un 34% de los estudiantes admitirían hasta 4 y 5 exámenes por asignatura. Parece evidente que el tiempo empleado para evaluar a los alumnos, expresado en horas o en número de pruebas a realizar, está bien calibrado por los alumnos.

3.2.3. Sobre los documentos usados en las evaluaciones y la ética en la evaluación:

Sobre si deben llevarse o no el documento (“examen”) utilizado para su evaluación, una vez concluida la prueba (Tabla 2, Pregunta 13), la inmensa mayoría de los alumnos (79%) consideran que si (el 78% de las mujeres y el 80% en los hombres).

Sobre si aceptarían firmar un documento deontológico o de honor, comprometiéndose a no copiar en los exámenes, ni a plagiar en los trabajos (Tabla 2, Pregunta 14), menos de la mitad de los alumnos (42%) contestan afirmativamente a dicho compromiso: un 42% de las mujeres y un 45% de los hombres. En concordancia con lo anterior (Tabla 2, Pregunta 16), se les solicitó a los alumnos que indicaran si los estudiantes copian en los exámenes. Para un 49% de ellos, en los exámenes se copia “regular”. Este porcentaje asciende al 57% en opinión de las mujeres y desciende al 30% según la opinión de los hombres. Sin embargo, para una cuarta parte (25%) de los encuestados (26% de las mujeres y 23% de los hombres) los estudiantes copian mucho. De estos datos se puede inferir que el nivel deontológico, relacionado con la evaluación, se encuentra a un nivel preocupante.

3.2.4. Sobre la homogeneidad de los exámenes y de su evaluación:

Una misma asignatura puede tener varios grupos de alumnos, diferentes profesores y diferentes pruebas o exámenes para cada grupo. Qué opinan los alumnos frente a la siguiente pregunta (Tabla 2, Pregunta 15): ¿todos los alumnos deben ser evaluados con la misma prueba o examen?. Para el 58% de los encuestados (57% las mujeres y 60% los hombres) todos los alumnos deben ser evaluados con la misma prueba. El resto de los alumnos se debaten entre los que dicen que no (18%), los que tienen dudas (6%), los que tienen dudas (18%) y la indiferencia (1%).

3.2.5. Sobre como obtener la calificación:

Son escasos los alumnos (2,54%) que fían su calificación al arbitrio exclusivo del profesor (Tabla 2, Pregunta 17). Para menos de la mitad de los alumnos (41%, las mujeres y 36%, los varones), la evaluación debe ser sumativa, es decir discriminándolos según sus notas. Los que creen que debe ser por criterios (Apto/No apto) se reducen a tan solo un 9% (10%, las mujeres y 8%, los hombres). Sin embargo, entre un quinto y un cuarto de ellos (24%, las mujeres y 17%, los hombres) preferirían una combinación de ambas formas de evaluación, mientras que para un 28% (22%, las mujeres y 36%, los hombres) se deberían aplicar ambos tipos de evaluación y el criterio del profesor.

Para alcanzar la máxima calificación final, es decir, sobresaliente o matricula de honor (Tabla 2, Pregunta19), más de un tercio (38%) de los estudiantes (36%, las mujeres y 42%, los hombres) creen que se debe alcanzar con las diferentes puntuaciones obtenidas en las actividades programadas. El número de alumnos que consideran que se debe realizar un examen especial con esa finalidad, es prácticamente nulo (0%, las mujeres y 2%, los varones). Sin embargo, una cuarta parte de ellos piensan que las máximas calificaciones deben ser obtenidas durante el curso más un examen específico (31% las mujeres y 15% los hombres). Finalmente, casi un tercio de ellos creen que mediante las calificaciones obtenidas durante el curso, un examen específico y el criterio del profesor (29% las mujeres y 32% los hombres). Resumiendo podemos decir que la mayoría de los alumnos confían en el uso de criterios mixtos para evaluar su rendimiento.

Como se puede deducir de estos resultados los alumnos todavía siguen muy mentalizados con la evaluación sumativa, siendo escasa la consideración de una evaluación por criterios.

3.2.6. Sobre la convocatoria de exámenes de recuperación

Respecto a la posibilidad del profesor para convocar exámenes de recuperación de las materias no superadas, en los exámenes programados por el centro, casi la mitad de los alumnos consideran que debería ser potestad privativa del profesor. Sin embargo, no llegan a un tercio los que consideran que solo se deben realizar los exámenes programados. Entre ambas posturas, una cuarta parte de los estudiantes, estarían de acuerdo en que se realizasen pero, bien en situaciones excepcionales, bien autorizados por el Decano o en ambas casos. Resumiendo, observamos dos bloques claramente definidos, uno liberal y permisivo (46%) y otro más conservador (52%).

3.3. Valoración de la evaluación sancionadora o sumativa por parte de los alumnos:

En la última pregunta del cuestionario tratamos de obtener cierta luz acerca de cómo debe realizarse la evaluación sumativa. Para algo más de un tercio de los estudiantes (33% las mujeres y 39% los hombres) esta evaluación es lo menos importante del PEA. Para casi una cuarta parte (30% las mujeres y 12% los hombres) debería ser una actividad colaborativa o colegiada. Sin embargo,

otra cuarta parte de los encuestados (23% las mujeres y 30% los hombres) piensan que debería ser responsabilidad única del profesor de la materia. Son muy escasos los encuestados que consideran que debería ser realizado por otros profesores ajenos al curso. Finalmente, no llega a una décima parte (10% las mujeres y 14% los hombres) los que piensan que debería ser llevada a cabo por otros profesores y de forma colegiada.

3.3. Resultados del análisis factorial exploratorio:

Tras un análisis factorial exploratorio, las variables utilizadas se agruparon en tres factores. El primer factor recoge la pregunta 17 (¿Cómo se debe valorar el rendimiento del alumno?) y las pruebas de evaluación (Preguntas 1, 2, 5, 6, 7, 8, 9 y 10), excepto los exámenes con PEM (Pregunta 3) y los orales (Pregunta 4,) ambos agrupados en el factor 3. El segundo factor incluye las preguntas 12, 13 y 14, relacionadas con la forma de aplicar las pruebas y el grupo al que pertenece el alumno. Finalmente el factor 3 recoge además de las preguntas 3 y 4 ya comentadas, también recoge las preguntas 11 (duración de las pruebas) y 15 (homogeneidad de los exámenes).

4. CONCLUSIONES

4.1. Aunque la opinión que los alumnos tienen sobre los tipos de pruebas que habitualmente utilizamos en el aula es muy variable, existen algunas tendencias que merecen ser resaltadas:

(1) Existe un mayor predicamento, entre los alumnos, de las denominadas pruebas objetivas frente a los exámenes escritos tradicionales con preguntas del tipo PRAL) o PRAC. Existe una clara tendencia de género en sus preferencias: las mujeres muestran una mayor preferencia por los exámenes de redacción abierta, lo contrario ocurre con las PEM.

(2) Frente a las pruebas de “papel y lápiz” (PRAL, PRAC y PEM), las pruebas de carácter oral tienen menor predilección por parte de los estudiantes. Sin embargo, su aceptación se hace mayor, especialmente en las mujeres, cuando la exposición oral es utilizada para la presentación de sus trabajos.

4.2. Valoración de los trabajos escritos: En el ámbito de los trabajos escritos, dos llaman la atención dada su preferencia por los alumnos: los trabajos dirigidos y el portafolios. Ambos gozan de gran predicamento entre los alumnos, especialmente entre las mujeres, probablemente debido a su mayor y mejor manejo del lenguaje.

4.3. Valoración de la resolución de problemas y de las habilidades prácticas: tanto las mujeres como los hombre, valoran muy positivamente la resolución de problemas como método de evaluación.

Sin embargo, la evaluación de las prácticas y de las manipulaciones de laboratorio no gozan de la misma predilección.

4.4. Sobre la aplicación de los exámenes:

(1) Casi la totalidad de los estudiantes, consideran que los exámenes no deberían durar mas de 2 y 3 horas.

(2) La inmensa mayoría de los estudiantes consideran que una vez realizado el examen, tienen derecho a llevar consigo el documento con las cuestiones planteadas.

(3) Parece un hecho comprobado que los estudiantes, en una proporción importante, copian durante la realización de los exámenes, sin embargo, menos de la mitad de ellos aceptarían firmar un documento deontológico comprometiéndose a no copiar en los exámenes, ni plagiar en los trabajos.

(4) Algo más de la mitad de los alumnos consideran que los exámenes deberían ser similares para los distintos grupos de una signatura, deberían ser evaluaciones sumativas y no por criterios y la calificación final debería proceder de la combinación de diferentes métodos de evaluación.

(5) En cuanto a la posibilidad de recuperar partes de la materia no superadas, los alumnos se alinean en dos bloques: los liberales o permisivos que no tienen problemas en que se realicen dichas recuperaciones y los conservadores que o no los permiten o ponen condiciones.

4.5. Sobre evaluación sancionadora habitual:

La mayoría de las pruebas de evaluación que se realizan en la Universidad tienen una finalidad sancionadora o sumativa, sin embargo, una tercera parte de los estudiantes la consideran como el elemento menos importante del PEA. En cuanto a la forma de llevarla a cabo, las opiniones de los alumnos se dividen entre los que consideran que debería ser una actividad colegiada y los que piensan que debería ser responsabilidad única del profesor de la materia.

4.6. Tanto los datos descriptivos del cuestionario como los resultados del análisis factorial sugieren una importante coherencia del cuestionario el cual, con algunas modificaciones, podría ser utilizado a gran escala para profundizar en el estudio de la evaluación de los alumnos.

REFERENCIAS BIBLIOGRÁFICAS

- Beard R. (1974). *Pedagogía y didáctica de la enseñanza universitaria*. Vilassar de Mar: Ed. Oikos Tau.
- De Juan, J. (1996). *Introducción a la Enseñanza Universitaria*. Madrid: Dykinson.
- De Juan, J., & Pérez-Cañaveras, R. M. (1991). Sistematización de los tipos de estrategias didácticas empleadas en la enseñanza universitaria. *Segundas Jornadas de Didáctica Universitaria* (pp 245-254). Madrid: Consejo de Universidades. Secretaría General.

- De Juan Herrero, J., & Pérez-Cañaveras, R. M. (2011). Ese oscuro objeto de la docencia. En María Cecilia Gómez Lucas, José Daniel Álvarez Teruel (coord.). *El trabajo colaborativo como indicador de calidad del Espacio Europeo de Educación Superior* (pp. 105-124). Alcoy: Marfil-Universidad de Alicante.
- De Juan, J, Ribera, D. y Cortejoso, A. (1978). *Objetivos, planificación y evaluación en la enseñanza de la Histología*. Valladolid: ICE de la Universidad de Valladolid.
- Domingues, Z. (1977). *Módulos para medir y evaluar en educación*. Madrid: Ed. Narcea.
- Fermin, M. (1971). *La evaluación, los exámenes y las calificaciones*. Buenos Aires: Ed. Kapelusz.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. *American psychologist*, 18, 519-521.
- Guilbert, J.J. (1994). *Guía Pedagógica para el personal de salud*. Valladolid: Organización Mundial de la Salud (OMS)-ICE de la Universidad de Valladolid.
- Iñiguez, C.; De Juan, J. y Carreres, J. (1979). *Estudio sobre Evaluación diagnóstica al comienzo de las carreras de Medicina y Enfermería*. Valladolid: ICE de la Universidad de Valladolid.
- Lafourcade, P.D. (1977). *Evaluación de los aprendizajes*. Madrid: Ed. Cincel.
- Lafourcade, P.D. (1974). *Planeamiento, conducción y evaluación en la enseñanza superior*. Buenos Aires: Ed. Kapelusz.
- Livas, I. (1977). *Análisis e interpretación de los resultados de la evaluación educativa*. México: Ed. Trillas.
- Nedelsky, L. (1954). Absoluting grading standards for objective test. *Educ Psychol Meas.* 14, 3-19.
- Núñez, L. (1977). Enseñanza programada. En: La formación del profesorado. Madrid: Educación abierta/Santillana.
- Peterssen, W.H. (1976). *La enseñanza por objetivos*. Madrid: Santillana.
- Scriven, M. (1967). The methodology of evaluation, En R.E. Stake (editor), *Perspectives of curriculum evaluation*, American Educational Research Association Monographs Series on Evaluation. Chicago: Rand McNally.
- Skowronek, H. (1970). *Lernen und Lernfähigkeit*, Munich.
- Sobrado, L.M. (1977). Evaluación de la enseñanza y del aprendizaje. En *La formación del Profesorado; nuevas contribuciones*. Madrid: Santillana.
- Vizcaya-Moreno, M. F., De Juan-Herrero, J., & Pérez-Cañaveras, R. M. (2004). El clima social: valoración del entorno de aprendizaje clínico desde la perspectiva de los estudiantes de enfermería. En *Premios Nacionales de Investigación Educativa 2004* (pp. 293-310). Madrid: Ministerio de Educación y Ciencia. (Colección Investigación Nº170). Centro de Investigación y Documentación Educativa (CIDE).

Tabla 2: Cuestionario para evaluar los usos y costumbres en la realización de exámenes

Preguntas	Respuestas				
Valora, por su bondad, eficacia y fiabilidad, los diferentes tipos de “exámenes” o evaluaciones que realizas	(A) Nula	(B) Muy poca	(C) Regular	(D) Buena	(E) Muy buena
1. Examen de tipo “Ensayo”: Preguntas de Redacción Abierta Largas (PRAL).	(14.40%)	(14.40%)	(32.20%)	(32.20%)	(6.78%)
2. Examen de tipo “Ensayo”: Preguntas de Redacción Abierta Cortas (PRAC)	(7.62%)	(7.62%)	(19.49%)	(54.23%)	(11.02%)
3. Examen de tipo “Test”: Preguntas de Elección Múltiple (PEMs)	(1.70%)	(5.93%)	(19.50%)	(37.29%)	(35.59%)
4. Realización de exámenes orales	(32.20%)	(26.27%)	(27.12%)	(9.32%)	(5.08%)
5. Realización de exámenes de prácticas	(13.56%)	(15.25%)	(25.42%)	(41.52%)	(4.24%)
6. Realización y evaluación del Portafolios de prácticas	(10.17%)	(11.86%)	(25.42%)	(38.98%)	(13.56%)
7. Realización de trabajos relacionados con los contenidos de la materia	(3.39%)	(4.24%)	(26.27%)	(51.70%)	(14.41%)
8. Presentación oral de trabajos relacionados con los contenidos	(10.17%)	(9.32%)	(31.36%)	39.83%	9.32%
9. Resolución de casos y/o problemas	(0.00%)	(5.09%)	(23.73%)	(42.37%)	(28.81%)
10. Examen de manipulaciones y técnicas (Clínicas, laboratorio, etc.)	(12.82%)	(10.26%)	28.21%	27.35%	21.37%
Respecto al modo de realizar los “exámenes” valora, según tú grado de acuerdo, cada una de las siguientes sentencias	(A)	(B)	(C)	(D)	(E)
11. Cualquier prueba de evaluación debería durar, como mucho	1h (5.09%)	2h (41.53%)	3h (47.46%)	4h (5.09%)	5h (0.85%)
12. En la evaluación continua de una materia de 6 créditos ECTS, se debería realizar el siguiente número de “exámenes”	1 (3.42%)	2 (27.35%)	3 (34.19%)	4 (22.22%)	5 (12.82%)
13. ¿Una vez realizado el examen, consideras que el alumno puede llevarse el documento con las preguntas?	No (4.27%)	Tengo Dudas (1.71%)	Me es Indiferente (5.13%)	Depende (10.26%)	Si (78.63%)
14. ¿Consideras oportuno que el alumno firme un compromiso deontológico o de honor de no copiar en el examen o de plagiarismo en sus trabajos.	No (15.25%)	Tengo Dudas (6.78%)	Me es Indiferente (27.12%)	Depende (7.63%)	Si (43.22%)
15. Cuando una asignatura o materia tiene varios grupos y varios profesores ¿todos los alumnos deben ser evaluados con la misma prueba (“examen”)?	No (17.80%)	Tengo Dudas (5.93%)	Me es Indiferente (0.85%)	Depende (17.80%)	Si (57.63%)
16. Según tu experiencia, los alumnos copian	Nada (0.86%)	Muy poco (10.26%)	Poco (15.39%)	Regular (48.72%)	Mucho (27.79%)
17. Como se debería, en tu opinión, valorar o calificar el rendimiento del alumno:					
(A) A discreción del profesor	(2.54%)				
(B) Como Apto o No apto, o sea, evaluación por criterios	(9.32%)				
(C) Con calificación numéricas o evaluación por normas (0 a 10)	(38.98%)				
(D) Combinando B y C	(21.19%)				
(E) Combinando A, B y C	(27.97%)				
18. Respecto a los exámenes de recuperación, consideras que					
(A) Solo deben realizarse los exámenes programados	(27.19%)				
(B) Solo deben realizarse en situaciones especiales	(10.53%)				
(C) Deben realizarse con la autorización del Decano	(4.39%)				
(D) Deben realizarse según B y C	(11.40%)				
(E) Deben realizarse a discreción del profesor	46.491%				
19. La máxima calificación final del alumno (sobresaliente o matrícula)					
(A) Derivar de las puntuaciones obtenidas en las actividades programadas	(38.46%)				
(B) Se conseguirá a partir exámenes específicos para subir nota					
(C) Se otorgará según el criterio discrecional del profesor	(0.86%)				
(D) Se conseguirán según A y B	(5.98%)				
(E) Se conseguirán según A, B y C	(24.79%)				
(E) Se conseguirán según A, B y C	(29.92%)				
Consideraciones de carácter “ideológico”					
20. Respecto a la evaluación sancionadora, consideras que debería					
(A) ser lo menos importante en el proceso de aprendizaje ((35.40%)				
(B) ser una actividad ajena al profesor	(4.425%)				
(C) realizarse de manera colegiada	(23.01%)				
(D) realizarse según B y C	(11.50%)				
(E) ser realizada por cada profesor de forma discrecional	(25.66%)				