EVALUACIÓN DE OPCIÓN MÚLTIPLE VS. EVALUACIÓN TRADICIONAL. UN ESTUDIO DE CASO EN INGENIERÍA

OSCAR M. GONZÁLEZ CUEVAS*

Resumen

Se presenta un estudio comparativo entre la evaluación tradicional y la múltiple, aplicado a estudiantes de Ingeniería, con el objetivo de analizar los problemas vinculados al diseño de cada tipo de evaluación y las dificultades que se presentan al tratar de medir aprendizajes de conceptos abstractos. Se concluye que el diseño de las pruebas de evaluación requieren de mayor interés de los profesores, considerando diferentes niveles taxonómicos para medir habilidades de resolución de problemas y aplicación de conocimientos; al mismo tiempo se considera que un tipo de evaluación sea superior a otro.

Palabras clave: evaluación institucional, enseñanza-aprendizaje, ingeniería.

* Universidad Autónoma Metropolitana-Azcapotzalco. División de Ciencias Básicas e Ingeniería
Correo-e: omgc@correo.azc.uam.mx
Se agradecen los comentarios y revisiones de los profesores Dr. Eduardo de la Garza, Dr. Miguel Ángel Gutiérrez Andrade y Silvia González Brambila, así como la ayuda técnica de Julio Pineda.
Recepción: 10 de enero de 2002.
Aceptación: 22 de abril de 2002.

Abstract

A comparative study between traditional and multiple choice evaluation applied to engineering students is presented. The key purpose is to analyze problems linked to the design of each type of evaluation tool and to the evaluation of abstract concepts under each modality. The paper concludes that evaluation test design calls for greater attention from faculty members due to the diversity of taxonomic levels applicable to the measurement of problem solving skills and knowledge application. At the same time, it is assumed that one type of evaluation is better than the other.

Key words: institutional evaluation, teaching-learning, engineering.

Antecedentes

El autor considera que la evaluación del aprendizaje logrado por los alumnos en un curso normal de un semestre, o de un trimestre en el caso de la institución en que labora, es un asunto no resuelto por completo. Quizá cada profesor ha desarrollado su propio método, en el que le da cierto peso a la realización de tareas, a la participación del alumno en clase, a los exámenes parciales y al final, a la ejecución de un proyecto o de un ensayo, etc. En el ámbito institucional también se han puesto en práctica diversas modalidades como los exámenes departamentales (el mismo examen para todos los grupos de una misma asignatura), el diseño del examen y su calificación por profesores distintos a los que imparten el curso, o la libertad absoluta a cada maestro. El tipo de asignatura también influye en el diseño del examen: en algunas se acostumbran exámenes orales, en otras, escritos, o una combinación. El tema de la equidad en la evaluación surge con frecuencia, especialmente cuando hay varios profesores de la misma asignatura que miden con "distintas varas" el aprendizaje de sus respectivos alumnos. Quienes no son profesionales de la educación, como es el caso del autor, suelen quedarse con dudas fundadas sobre sus estrategias de evaluación.

Por otra parte, la competencia entre cada vez mayores números de aspirantes por ingresar a instituciones de educación superior, la competencia entre las instituciones por atraer a los mejores alumnos y el interés de la sociedad por conocer el grado de preparación de los egresados, han propiciado la elaboración y la aplicación de exámenes a grandes grupos de aspirantes o de egresados. Los fines de estas evaluaciones son diversos: el derecho a ingresar a cierto nivel educativo, la selección entre un número de aspirantes mayor al cupo de la institución o del sistema, la certificación profesional, etc. Las características de estas evaluaciones son distintas de las que se aplican en un salón de clase, ya que el número de sustentantes es mucho mayor y se demanda de ellas una objetividad total. Desde luego que este tipo de evaluaciones también tiene sus limitaciones, más serias, según algunos estudiosos del tema, que las del párrafo anterior. Quizá la más usual entre estas evaluaciones sea la llamada de opción múltiple, en la que se ofrecen al sustentante varias respuestas a una pregunta para que señale la correcta.

El objetivo de este trabajo es presentar los resultados de un estudio comparativo realizado por el autor entre una evaluación tradicional y una evaluación de opción múltiple. Se aplicaron a los alumnos de dos cursos de análisis estructural, de la carrera de ingeniería civil, exámenes que consistían de dos partes: una evaluación tradicional y otra de opción múltiple. Se pretendió, no sólo comparar las calificaciones obtenidas en cada tipo de evaluación, sino analizar los problemas vinculados al diseño de cada tipo de evaluación y las dificultades que se presentan al tratar de medir aprendizajes de conceptos importantes de naturaleza abstracta.

En una revisión bibliográfica se encontraron pocas referencias vinculadas al tema, específicamente para las carreras de ingeniería. Leuba (1986a y 1986b) hace una interesante reflexión sobre los objetivos de evaluar el aprendizaje de los alumnos, presenta recomendaciones y ejemplos de preguntas para ser calificadas en computadora, de las cuales forman parte las de opción múltiple, y concluye que este tipo de evaluaciones sí permite comprobar que los alumnos hayan logrado una comprensión básica de la disciplina y una habilidad adecuada para la resolución de problemas. Citando a otro autor, Robert E. Ebel (1979), se adhiere a la opinión de que "... se adapta (este tipo de evaluaciones) a la medición de los resultados educativos más importantes: conocimiento, comprensión y buen juicio; a la habilidad de resolver problemas, de recomendar acciones apropiadas y hacer predicciones. Casi cualquier comprensión o habilidad que pueda medirse con otras formas de evaluación ... puede medirse también con exámenes de opción múltiple". Leuba rechaza que las preguntas de opción

múltiple sólo puedan asociarse a memorizaciones sencillas y sostiene que pueden diseñarse preguntas que midan actividades intelectuales de orden superior, como resolución de problemas, creatividad y capacidad de síntesis.

Kessler (1988) realizó un experimento parecido al presentado en este artículo. Aplicó a un grupo de alumnos un examen para ser calificado en computadora, consistente en una mezcla de preguntas de opción múltiple y del tipo falso/verdadero, y comparó las calificaciones con el promedio que habían obtenido los alumnos en los semestres anteriores y con el promedio obtenido en las otras asignaturas del semestre en que se realizó el experimento (octavo semestre de ingeniería química). En un segundo experimento, Kessler aplicó seis exámenes a un grupo de alumnos que alternaban exámenes tradicionales y exámenes para ser calificados en computadora. De estos últimos, dos fueron del tipo falso/verdadero y uno, de opción múltiple. Los resultados de Kessler se presentan más adelante y se comparan con los obtenidos por el autor.

El entorno

Con el fin de que esta experiencia pueda serle útil a profesores de diversas instituciones y carreras, se presenta en este apartado el entorno de la universidad en que se trabajó y del curso de que se trata. El profesor y los alumnos pertenecen a la Universidad Autónoma Metropolitana-Azcapotzalco. El curso forma parte del plan de estudios de ingeniería civil, dentro del cual existen tres áreas de concentración o especialización al final de la carrera: estructuras, construcción y mixta. Análisis estructural es un curso obligatorio para las tres áreas de concentración. La duración de la carrera es de 12 trimestres y a este curso se inscriben los alumnos del décimo trimestre en adelante. Como el plan de estudios es flexible, pueden elegir en qué momento cursarlo, y es frecuente que lo dejen para el final de la carrera. Esto se debe a que se considera un curso difícil, que les demanda mucha dedicación. La clase se imparte cinco días a la semana, con duración de hora y media, y para seguir el ritmo los alumnos deben dedicar entre ocho y diez horas de estudio personal a la semana. Además del autor, hay otros tres profesores que imparten el curso, y el índice de aprobación, cualquiera que sea el profesor, normalmente es menor al 50%, en ocasiones mucho menor. Esta situación se presenta también en la mayoría de las escuelas de ingeniería civil del país.

El curso está considerado como uno de los más importantes en la carrera de ingeniería civil. Además de que es necesario para aprender la práctica del diseño estructural, una de las ramas más antiguas y consolidadas de la ingeniería civil, adiestra a los alumnos en el manejo de conceptos abstractos, requiere el empleo de herramientas matemáticas de cierta complejidad y desarrolla en los alumnos habilidades intelectuales propias de la ingeniería. Se incluye en todos los planes de estudio del país, aunque el nombre puede cambiar en algunos de ellos. En forma muy resumida, lo que se enseña son métodos para determinar las acciones o fuerzas internas que se presentan en estructuras o elementos estructurales, como vigas, columnas, losas o muros, cuando actúan cargas sobre ellos. Hacer esta determinación es esencial para el diseño y construcción de estructuras, de ahí la importancia del curso.

La experiencia que se presenta tuvo lugar en el año de 2000, en dos cursos consecutivos impartidos por el autor en los trimestres de primavera y otoño. Los grupos a los que se aplicaron las evaluaciones son típicos de los que se inscriben al curso, en cuanto a número de alumnos, edad, composición por sexo, etcétera. Ya que los alumnos se inscriben cerca del final de su carrera, como ya se ha señalado, aproximadamente la mitad trabaja ya en actividades vinculadas a la ingeniería. Este hecho es importante de remarcar porque muchos no están dedicados exclusivamente al estudio y encuentran dificultades para cumplir con las tareas asignadas durante el curso.

Todos los alumnos habían tenido la experien-

cia de someterse a una evaluación de opción múltiple, ya que la Universidad tiene como requisito de ingreso ser seleccionado mediante examen de admisión, que consiste en una evaluación de este tipo. Sin embargo, no habían tenido ninguna otra durante su estancia en la misma. El tipo de evaluación más usual en las carreras de ingeniería es similar al llamado evaluación tradicional en este trabajo, cuyas características se presentan en el siguiente apartado.

Evaluación tradicional en los cursos de análisis estructural

La evaluación de los cursos, aprobada en uno de los cuerpos colegiados de la Universidad, consiste normalmente en dos exámenes parciales y un examen final, o global como se denominan en la UAM. A los parciales se les asigna un peso de entre 40 y 60%, según el profesor, y al global el porcentaje restante. Algunos profesores también le asignan un cierto peso, relativamente menor, al cumplimiento en la realización de tareas y a la participación en clase; en este caso, se reduce el peso de los exámenes.

La práctica usual es incluir tres o cuatro problemas en cada examen. Cada uno consiste en la resolución de alguna estructura sencilla, como una viga continua o un marco pequeño, con diversas condiciones de carga. Se permite que los alumnos preparen un "acordeón" o formulario para auxiliarse en el examen, o bien, que consulten tablas o gráficas del texto. Un primer problema que se presenta en los exámenes tradicionales es el del tiempo necesario para su resolución. Cada problema requiere de una a dos horas, así que por sencillo que sea el examen se necesitan tres o cuatro horas de trabajo con una demanda de alta concentración. Aún así, tres o cuatro problemas no son suficientes para cubrir todo el programa de la asignatura, por lo que no se puede saber si el alumno domina todo el contenido. Tampoco se puede alargar más la duración del examen, pues los alumnos están cursando otras asignaturas, cuando se aplican los exámenes parciales, o están presentando los exámenes de otras asignaturas en la semana de exámenes finales.

Durante la resolución de los problemas, los alumnos deben llevar a cabo una serie de etapas, cada una de las cuales depende del resultado obtenido en la anterior. Aquí surge otra dificultad: si yerran en una de las primeras etapas, todo el problema resulta equivocado, y no es fácil detectar si el yerro es por falta de dominio del tema, por una equivocación numérica, o por cansancio después de varias horas de trabajo intelectual intenso. En estas circunstancias, resulta difícil separar lo que el alumno domina de lo que desconoce o no comprendió adecuadamente. Para asignar la calificación, suele considerarse todo el procedimiento y no sólo la respuesta final (se asignan puntos parciales), pero aun así resulta complicado determinar si el alumno realmente conoce el procedimiento y, sobre todo, si entiende por qué se aplica dicho procedimiento. Existen ejemplos de exámenes de este tipo, en el campo de la ingeniería, en los que se establecen de antemano "estrategias de calificación" para cada pregunta, con lo cual se reduce la subjetividad en la calificación (National Council of Examiners for Engineering and Surveying, 1996). Pero no es usual emplear estas técnicas en exámenes rutinarios.

Otro problema en las evaluaciones tradicionales de este curso, muy importante desde el punto de vista del autor, es el siguiente. Durante el curso, el alumno debe comprender conceptos básicos y luego aplicar estos conceptos a la resolución de problemas. Esta segunda fase puede llegar a hacerse en forma mecánica o rutinaria, inclusive sin haber comprendido bien los conceptos. El alumno estudia un problema similar ya resuelto, memoriza los pasos que se siguieron y luego los aplica al nuevo problema sin entender el fondo de la solución. Aplica una especie de algoritmo, sin meditar en el significado del proceso de solución. Puede entonces suceder que el alumno obtenga una buena calificación en

el examen aunque no tenga conceptos sólidos. Desde luego que esta situación no es deseable en lo absoluto, pues de lo que se trata es de dominar conceptos que puedan aplicarse a problemas novedosos y no de darle al alumno un entrenamiento mecánico.

Es cierto que las evaluaciones tradicionales en el curso que nos ocupa han sido usadas durante muchos años, y se aceptan ampliamente como un método correcto para medir el aprendizaje de los alumnos. Sin embargo, por las razones expuestas, el autor considera que tienen limitaciones que conviene analizar para mejorar sobre bases objetivas nuestros métodos de evaluación. Este análisis puede llevarse a cabo a partir de la Taxonomía para la Resolución de Problemas, presentada por Plants (1980) y comentada con amplitud en el excelente libro de Wankat y Oreovicz (1993). Según estos autores, existen cinco niveles taxonómicos para la resolución de problemas:

1. Rutinas. Operaciones o algoritmos que pueden realizarse sin necesidad de tomar decisiones, como resolver una ecuación de segundo grado o una integral. En la taxonomía de Bloom sobre dominios del conocimiento, correspondería al nivel de aplicaciones.
2. Diagnosis. Selección de la rutina correcta o de la manera correcta de usar una rutina, como calcular los esfuerzos en una viga. Se traslapa con los niveles de aplicación y análisis de la taxonomía de Bloom.
3. Estrategia. Es la selección de rutinas y de su orden de aplicación cuando se requieren varias para resolver un problema. Corresponde a los niveles de análisis y evaluación de la taxonomía de Bloom.
4. Interpretación. Significa reducir un problema de la vida real a otro que pueda resolverse con las herramientas disponibles. Es indispensable para la resolución de problemas reales. Este nivel y el siguiente no suelen vincularse a alguno de la taxonomía de Bloom.
5. Generación. Es la generación de rutinas nuevas para el usuario. Involucra a la creatividad en el sentido de que las nuevas rutinas pueden no ser obvias para el usuario.

En la experiencia del autor, en los exámenes tradicionales de análisis estructural (y de muchas asignaturas de la carrera) se utiliza principalmente el primer nivel de esta taxonomía, y a veces el segundo. Diseñar exámenes tradicionales que involucren a los otros niveles de la taxonomía y que puedan resolverse con las limitaciones de tiempo, y otras de tipo práctico, es realmente complicado. Por eso le parece interesante estudiar la posibilidad de usar los exámenes de opción múltiple, que en su experiencia y en la de autores como Leuba (1986) y Kessler (1988), tienen un buen potencial de incursionar en los niveles más avanzados de la taxonomía.

Evaluaciones de opción múltiple

En este apartado se presenta una descripción breve de las evaluaciones de opción múltiple, para los lectores no familiarizados con esta técnica, y la experiencia que ha tenido el autor durante su participación en la elaboración del Examen General de Egreso de la Licenciatura (EGEL) del área de ingeniería civil en el Centro Nacional de Evaluación para la Educación Superior (CENEVAL). Esta participación y los estudios comentados en el párrafo anterior fueron los que lo motivaron para explorar la posibilidad y la conveniencia de aplicar evaluaciones de opción múltiple en cursos normales de la carrera de ingeniería civil.

La técnica, en términos generales, consiste en plantear una pregunta o problema, denominado *reactivo*, que consta de un *enunciado* y una serie de respuestas, llamadas *opciones*. Entre estas respuestas hay una correcta, llamada *solución*, y otras incorrectas, conocidas como *distractores*. El número de respuestas opcionales en cada reactivo depende de la probabilidad que se acepte de que el alumno conteste bien al azar. En los exámenes del CENEVAL se plantean cuatro opciones en

cada reactivo, incluyendo la correcta. Leuba y Kessler usan cinco opciones por reactivo.

En el caso de los exámenes de ingeniería del CENEVAL, existen dos tipos de reactivos: de *conocimiento* y de *aplicación*. Los primeros pretenden evaluar si el sustentante tiene una adecuada comprensión de los conceptos y principios de las ciencias básicas y de la ingeniería, y los segundos, si posee la habilidad para aplicar dichos conceptos y principios a la solución de problemas de ingeniería. Los reactivos de conocimiento se deben poder contestar en un tiempo promedio de dos minutos, y los de aplicación, de cinco minutos. Estos tiempos, relativamente cortos, permiten que el examen de ingeniería civil, por ejemplo, en su conjunto, se pueda resolver en un tiempo aproximado de 12 horas, en varias sesiones, cubriendo el amplio espectro de temas que constituyen la carrera.

El CENEVAL ha publicado una "Guía para la elaboración de reactivos" en la que se presentan recomendaciones y reglas para elaborar buenos reactivos. Esta guía se distribuye entre numerosos profesores que atienden una convocatoria de la institución y envían reactivos que son revisados, y en su caso aprobados, por un comité académico. El autor ha pertenecido durante varios años a este comité.

Después de cada aplicación de un EGEL, los reactivos son revisados por el comité académico para verificar si cumplen dos condiciones que debe tener un buen reactivo. La primera es su *grado de dificultad*. Los reactivos no deben ser ni muy fáciles ni muy difíciles. El grado de dificultad se mide por el porcentaje de alumnos, *p*, que responden correctamente al reactivo. Ya que la probabilidad de responder bien al azar un reactivo es de 25%, el mínimo de dificultad debe ser por lo menos este valor. Leuba (1986a: 91) señala que el valor de *p* para exámenes estandarizados, como el del CENEVAL, suele ser cercano a 50%, pero que en exámenes de una asignatura debe esperarse un valor mayor. Cabe hacer notar que él usa algunos reactivos del tipo falso/verdadero que tienen mayor probabilidad de responderse bien al azar. En el Manual Técnico del CENEVAL (Vidal, *et al*., 2000) se establece que un reactivo debe tener un grado de dificultad entre 20 y 80% para se aceptable y que el intervalo óptimo debe estar entre 27 y 73%.

La segunda condición es su *poder de discriminación.* Un buen reactivo debe permitir diferenciar entre un sustentante con buena preparación y otro que no la tenga. Esto se analiza verificando que el reactivo sea respondido correctamente por una mayoría de los sustentantes pertenecientes al grupo superior (los que obtienen mejor calificación) en el conjunto de reactivos, e incorrectamente por la mayoría de los sustentantes pertenecientes al grupo inferior en todo el examen. El poder de discriminación, *PD*, se define como la diferencia entre el porcentaje de alumnos del grupo superior que responde correctamente *(PGS)* y el porcentaje de alumnos del grupo inferior que responde correctamente (*PGI*). De tal manera que

$$PD = (PGS - PGI)$$

Esta ecuación indica que un reactivo con un alto poder de discriminación es respondido por un alto porcentaje de alumnos del grupo superior y un pequeño porcentaje de alumnos del grupo inferior. La línea de separación entre los grupos superior e inferior suele establecerse por la mediana de la muestra, aunque algunos autores definen al grupo superior como a aquellos alumnos ubicados en el 25% mayor, y al inferior, en el 25% menor. Leuba (1986b: 184), con base en su experiencia, recomienda valores de *PD* del orden de 0.30, usando los cuartiles superior e inferior.

Tristán (1996) ha hecho notar que el grado de dificultad y el poder de discriminación no son independientes. Un reactivo muy fácil o muy difícil no discrimina, porque lo resuelven casi todos o casi nadie. Esta idea se ha incorporado al Manual Técnico del CENEVAL, en el que se establece que el poder de discriminación, en vez de tener un valor fijo, satisfaga una norma

discriminativa (*ND*) que depende del grado de dificultad. Los valores propuestos son:

$$ND = 0.3\,GD;\ si\ 0 \leq GD\ 76.92\%$$
$$ND = 100 - GD;\ si\ 76.92 \leq GD100\%$$

Para verificar si un reactivo cumple con la norma, se define la relación discriminativa, que es el cociente del poder de discriminación y la norma discriminativa:

$$RD = \frac{PD}{ND}$$

Si *RD* es mayor que 1, el poder de discriminación es mayor que la norma y el reactivo se acepta automáticamente. Si es menor, se recomienda analizarlo en cuanto a contenido y redacción y aceptarlo o rechazarlo según el resultado de este análisis.

En la experiencia del autor de este artículo, la elaboración de reactivos para exámenes de opción múltiple, al menos en el campo de la ingeniería, es realmente difícil. Desde luego que se requiere un dominio completo del tema por parte de quienes los elaboran, pero además hay otras condiciones. En primer término, el concepto y la idea que se tratan de evaluar deben estar perfectamente definidas y ser muy puntuales, ya que de otra manera no se puede elaborar un reactivo factible de ser contestado en el tiempo disponible. Esto contrasta con los exámenes que aquí se han llamado tradicionales, en los que se plantea un problema completo a resolver y en el que aparecen multitud de conceptos e ideas a lo largo de la resolución, generalmente vinculados entre sí.

Otra condición, que se presenta frecuentemente en reactivos de aplicación, es la de diseñar un reactivo que permita detectar si el sustentante conoce un determinado método, ya que puede darse el caso de que resuelva el reactivo por un método diferente que de todas maneras le permita obtener la respuesta correcta. O bien, que compruebe únicamente las opciones ofrecidas para encontrar la correcta, lo que se llama una *solución hacia atrás*. Ya que no es posible verificar por cuál método el sustentante resolvió un problema determinado, el reactivo tiene que diseñarse de tal manera que la resolución esté vinculada con el método cuyo conocimiento se quiere detectar. Esto no es fácil, pero en la experiencia del autor sí es posible, aunque requiere más imaginación por parte de quien diseña el reactivo. Los ejemplos presentados por Leuba y Keller corroboran esta idea.

Los reactivos de aplicación requieren normalmente que se efectúen algunas operaciones aritméticas. Éstas deben ser muy sencillas para poder responder el reactivo en el tiempo asignado, que como ya se mencionó, es relativamente corto. También deben ser sencillas para minimizar la posibilidad de un error numérico, cuya detección no sería posible durante el proceso de calificación, ya que sólo se revisan los resultados. Por lo tanto, hay que diseñar reactivos que no impliquen muchas operaciones numéricas. En ingeniería no siempre es fácil lograr esta condición.

También resulta difícil diseñar reactivos que se mantengan en los márgenes recomendados de grado de dificultad y poder de discriminación o relación discriminativa. Los miembros del comité académico del CENEVAL revisaron detenidamente cada reactivo, antes de su aplicación, para verificar la respuesta correcta y la idoneidad de los distractores, la redacción, la presentación, los dibujos, y desde luego, que el contenido fuese apropiado y de acuerdo con los temarios vigentes en las escuelas de ingeniería del país. Sin embargo, al volverlos a revisar después de su aplicación, se encontraba, con frecuencia, que habían resultado demasiado fáciles, o demasiado difíciles, o que no discriminaban. Entonces se volvían a revisar cuidadosamente, y si se encontraban causas justificadas para considerarlos confusos o se veía algún error de redacción o de dibujo, se corregían para volverlos a aplicar en su versión modificada. Aún así, algunos seguían sin cumplir los requisitos de dificultad y discriminación, y entonces eran eliminados.

Y la parte más difícil, en la experiencia del autor, es el diseño de los distractores o respuestas equivocadas. Estos distractores deben ser

respuestas que no parezcan absurdas, ya que el sustentante las desecharía automáticamente. Deben ser respuestas posibles para quien no domina el tema y requerir el análisis de todas las posibilidades antes de elegir la respuesta correcta.

Al igual que los exámenes tradicionales, las evaluaciones de opción múltiple tienen desventajas y ventajas. Sus desventajas de tipo general han sido señaladas con énfasis por sus detractores, quienes llegan a descalificar totalmente este tipo de evaluación[1], aunque debe aclararse que muchas críticas se refieren a la utilización de los resultados de los exámenes y no a la técnica. En el caso de la evaluación de la asignatura de análisis estructural, una de sus desventajas particulares es la imposibilidad de indagar si el sustentante domina un método completo, es decir, si conoce todas las etapas de resolución de un problema y su vinculación. Los reactivos se refieren a etapas específicas, pero es importante que el alumno sepa cómo vincularlas para llegar a la solución de un problema determinado. Quizá sea posible diseñar reactivos que permitan hacer esta indagación, pero el autor no los ha visto entre los que son propuestos normalmente para los exámenes del CENEVAL. Los ejemplos presentados por Leuba (1986a) corresponden a una asignatura de estática, tema similar al análisis estructural, y parece ser que sí es posible superar este inconveniente.

Su principal ventaja, a juicio del autor, es que permite averiguar, de manera muy clara, si el alumno domina los conceptos y no únicamente la mecanización de un procedimiento; se recordará que ésta es la principal desventaja del examen tradicional. Es decir, permite indagar si el alumno está en etapas avanzadas de la taxonomía de resolución de problemas comentada anteriormente. Otra ventaja importante es que reduce al mínimo la influencia de errores numéricos cometidos por el sustentante en su calificación final, errores muy frecuentes en los exámenes tradicionales. Y una ventaja más es que reduce considerablemente el tiempo necesario para calificar los exámenes, aunque el tiempo de preparación de los reactivos es mucho mayor que el de elaboración de exámenes tradicionales; a largo plazo resulta ventajoso el examen de opción múltiple, ya que un reactivo puede ser aplicado varias veces, sobre todo si se deja transcurrir algún tiempo antes de volverlo a aplicar; esta práctica es común en este tipo de exámenes. Esta última ventaja permite que el profesor pueda realizar un mayor número de exámenes a lo largo del curso, lo cual es recomendado enfáticamente por especialistas en el tema (Wankat y Oreovicz, 1993: 214-216); o bien, facilitan la aplicación de exámenes por profesores distintos a los que imparten el curso, lo cual tiene ventajas pedagógicas. Por las razones anteriores, y a sabiendas de que las evaluaciones de opción múltiple se desarrollaron para aplicar un examen a números muy grandes de sustentantes, el autor consideró interesante analizar su aplicabilidad en las evaluaciones de los cursos de análisis estructural.

La investigación

La investigación consistió en aplicar a los alumnos de dos cursos de análisis estructural, impartidos por el autor, exámenes que consistían de dos partes, la primera era un examen tradicional y la segunda, uno de opción múltiple. La hipótesis de investigación era que los alumnos que obtuviesen mejores calificaciones en un tipo de examen las obtendrían también en el otro tipo. En cada curso se aplicaron dos exámenes parciales y uno final o global, por lo que fueron seis exámenes en total. En todos ellos se utilizó un examen de dos partes como se mencionó. El primer curso

[1] Por ejemplo, se ha criticado que se utilice como mecanismo de selección para el ingreso a bachillerato o a licenciatura porque no permite predecir con un grado de confiabilidad adecuado el comportamiento académico de los sustentantes.

se impartió entre los meses de mayo y julio de 2000, y el segundo, entre septiembre y diciembre del mismo año.

En la Tabla 1 se presentan las principales características de los grupos y de los exámenes. El número de sustentantes es típico en los cursos de análisis estructural en la UAM; quizá el primer curso tenga un número mayor de alumnos que el promedio regular. En el momento de aplicar el primer parcial algunos alumnos, del orden del 10%, han desertado ya del curso. Los sustentantes en el primer parcial son los que quedan inscritos en definitiva. Puede verse que el número de alumnos disminuye todavía más a lo largo del trimestre.

La parte de opción múltiple fue la primera que se aplicó en todos los casos. Se dio un tiempo de 90 minutos y, en el primer parcial, se aplicaron 15 preguntas, estimando que las preguntas podrían ser respondidas en seis minutos cada una, tiempo un poco mayor que el de los exámenes del CENEVAL. Una pregunta de este primer parcial tuvo que ser eliminada en el momento de calificar porque tenía un error en un dibujo que imposibilitaba obtener la respuesta correcta. Por eso en la Tabla 1 hay una columna de preguntas propuestas y otra de preguntas consideradas. Un error similar ocurrió en el primer parcial del segundo curso.

Después de la aplicación de la parte de opción múltiple, se permitió a los alumnos un descanso de 15 minutos y se aplicó la parte de preguntas abiertas, o sea un examen tradicional. Para esta segunda parte se destinaron también 90 minutos, y en todos los casos se plantearon dos problemas para resolver, tratando de que no fuesen muy largos, pero similares a los de los exámenes tradicionales.

En la primera aplicación del examen de opción múltiple se observó que sólo un alumno había entregado su examen antes de agotarse el periodo de 90 minutos. Además, en entrevistas con los alumnos, ellos opinaron que el tiempo les había parecido escaso. Por estas razones, en las siguientes aplicaciones se redujo a 12 el número de preguntas de opción múltiple. Mientras mayor sea el número de reactivos, menor es la probabilidad de obtener una buena calificación contestando todas las preguntas al azar. Con 12 preguntas y cuatro opciones en cada pregunta, la probabilidad de obtener una calificación de 5 o mayor, en una escala de 1 a 10, es de 1.4%[2].

Una vez realizada la primera aplicación, se revisaron los reactivos de la misma manera que se hacía con los reactivos del CENEVAL, es decir, se verificó su grado de dificultad, su poder de discriminación y su relación de discriminación.

En los exámenes de opción múltiple del segundo curso se utilizaron algunos reactivos del primer curso. Desde luego se eliminaron los que resultaron inaceptables en la revisión, pero algunos se volvieron a usar tomando en cuenta que los alumnos no guardaron copia de los exámenes y ni siquiera supieron cuál era la respuesta correcta. Era entonces muy difícil que los alumnos de un curso pudiesen orientar a los del siguiente curso sobre las preguntas del examen. En cambio, los problemas de la parte tradicional sí fueron completamente diferentes, ya que en este caso los alumnos suelen llevarse una copia de los problemas, tres a lo sumo, para estudiar después del examen y para preparase en siguientes pruebas.

Resultados de la investigación

Después de la primera aplicación del examen, se sostuvo una plática con los alumnos del grupo para conocer sus impresiones sobre el tipo de examen, antes de que conociesen sus calificaciones. Todos se mostraron un tanto

[2] La probabilidad de tener k aciertos en n reactivos, siendo p la probabilidad de acertar en un reactivo, es:

$$P(k)=\binom{n}{k} \quad p^k(1-p)^{n-k}$$

. Véase, por ejemplo, Volk, 1958: 22-25.

sorprendidos, aunque se les había advertido en términos generales cómo sería el examen[3]. Durante la plática muchos de ellos admitieron que se habían preparado para resolver ejercicios en forma mecánica, estudiando ejemplos ya resueltos y memorizando los pasos o etapas para la solución. Las preguntas de la parte de opción múltiple, que ellos llamaban "de teoría", resultaron inesperadas. Se les preguntó qué tipo de examen preferirían en el futuro. La gran mayoría, 42 de los 45 sustentantes, respondieron que una combinación de preguntas de opción múltiple y problemas abiertos les parecía más conveniente. El argumento que prevaleció fue que en los exámenes tradicionales, si aparecía un problema sobre un tema que no dominaban, perdían de entrada una buena parte de la calificación. En cambio, en el de opción múltiple, como aparecían más temas, la falta de conocimiento o dominio de uno de ellos no repercutía tan desfavorablemente en la calificación final.

En las tablas 2 a 7 se muestran los resultados de los exámenes de opción múltiple, para cada uno de los exámenes aplicados. Para cada reactivo se indica cuántas personas lo contestaron correctamente, de donde se calculó el grado de dificultad, y los valores del poder de discriminación, de la norma discriminativa y de la relación discriminativa, de acuerdo con los criterios presentados en la Sección 4. Los reactivos señalados con un asterisco son los que se consideraron inaceptables; puede verse que su relación discriminativa es mucho menor que 1. Algunos que tenían relación discriminativa menor que 1, pero no menor que 0.67, se consideraron aceptables después de analizar su contenido y su grado de dificultad. En la siguiente sección se comentan estos resultados.

Se mencionó al inicio de este trabajo que el objetivo principal de la investigación era verificar si los alumnos que obtenían buena calificación en el examen de opción múltiple también la obtenían en el examen de problemas abiertos. El autor pensaba que así sería. Algunos colegas con los que se comentó la investigación que se pretendía llevar a cabo opinaban igual. Leuba (1986a: 91) "acepta como un acto de fe que un alumno que puede obtener una B en un examen tradicional también la puede obtener en un examen calificado por computadora".

Con el objeto de comparar gráficamente las calificaciones obtenidas en cada tipo de examen, se elaboraron las figuras 1 a 6, una para cada examen. Los puntos en los ejes de las abscisas representan a cada alumno que presentó el examen. Para cada uno de ellos, se indica en el eje de las ordenadas la calificación que obtuvo en el examen de opción múltiple con todos los reactivos, línea llena, la que obtuvo si se eliminan los reactivos que resultaron inaceptables (los señalados con asterisco en las tablas 2 a 7), línea de punto y raya, y la que obtuvo en el examen tradicional, línea punteada. Los resultados se presentan por orden decreciente de calificaciones en el examen de opción múltiple con todos los reactivos. Si los alumnos que obtienen buena calificación en una parte del examen la obtuviesen también en la otra parte, las líneas quedarían ubicadas en la gráfica cerca una de otra, o por lo menos seguirían la misma tendencia. Las líneas correspondientes a todos los reactivos y a los reactivos aceptables sí quedan cercanas entre sí, pero la correspondiente a los exámenes tradicionales se separa sensiblemente de las otras dos. Más adelante se analiza el grado de correlación entre las calificaciones de los exámenes de opción múltiple y las calificaciones de los tradicionales.

Para calificar los exámenes tradicionales, el autor siguió un procedimiento más detallado que el que había usado normalmente. Dividió las

[3] También se les dijo a los alumnos que por tratarse de un experimento y para no perjudicarlos en su calificación final, la parte del examen en que obtuviesen calificación más alta se tomaría como el 70% de la calificación total, y la otra parte sería el 30%. Se les recomendó igualmente no dejar ninguna pregunta sin respuesta, ya que aun respondiendo al azar tenían una probabilidad de 25% de acertar.

preguntas en una serie de subtemas y calificó cada uno de ellos con un punto, medio punto o cero. Si al calificar cada subtema, había un error derivado de otro error en un subtema anterior, pero el procedimiento era correcto, otorgaba medio punto; si el procedimiento era incorrecto, otorgaba cero. Trató, de esta manera de reducir la subjetividad. Con el fin de comparar las calificaciones obtenidas con este sistema de asignación de puntos y las obtenidas en cursos anteriores, el autor calculó las medias de tres exámenes tradicionales aplicados en 1998. Esta medias fueron 4.30, 5.70 y 5.25, respectivamente. Si se cotejan con las de los exámenes tradicionales de la Tabla 8, se puede ver que son mayores, especialmente en el trimestre 00-P.

Discusión de los resultados

Un primer resultado que debe resaltarse es que las calificaciones obtenidas por los alumnos, en general, son muy bajas, en ambos tipos de examen. En la Tabla 8 se muestran las calificaciones promedio en los exámenes de opción múltiple y tradicionales, considerando todos los reactivos o eliminando los inaceptables en los de opción múltiple. También se puede ver en las tablas 2 a 7 que los valores de p son muy bajos. Aunque ni Leuba ni Kessler presentan en sus artículos una evaluación sistemática de los valores de p que obtuvieron, sus ejemplos y sus comentarios indican valores mucho mayores.

Las razones de que hayan resultado calificaciones muy bajas pueden ser varias. Una de ellas es desde luego la dificultad de la asignatura, ya comentada. En general, las calificaciones resultan bajas, pero en los exámenes tradicionales se oculta parcialmente este hecho al asignarlas en forma comparativa entre los alumnos. Calificaciones iguales o cercanas a las máximas se otorgan a alumnos que resultan los mejores del grupo, aunque sus exámenes tengan fallas. O se normalizan las calificaciones otorgando 5 sobre 10 al alumno que queda en la media, aunque ésta sea muy baja. El asignar en forma subjetiva puntos parciales a problemas que no están resueltos totalmente bien, incrementa normalmente las calificaciones obtenidas. Desde luego que la manera de calificar los exámenes de opción múltiple no deja lugar a ninguna subjetividad: la respuesta es correcta o incorrecta. En el primer caso el alumno obtiene un punto y en el segundo, ninguno. Y en los exámenes tradicionales se trató de reducir a un mínimo la subjetividad, según se comentó en la sección anterior. Esto explica en parte que las calificaciones obtenidas en este experimento resulten bajas en comparación con las de cursos anteriores.

Otra posible causa de que las calificaciones hayan resultado bajas es la escasa experiencia en la formulación de exámenes de opción múltiple. La dificultad de las preguntas se puede ir ajustando con base en los resultados que se vayan obteniendo hasta lograr valores de p que en promedio sean mayores de 0.5 o 0.6, como las obtenidas en los experimentos de Leuba y Kessler. Debe señalarse, sin embargo, que estos autores usaron preguntas del tipo verdadero/falso en sus exámenes, y en estas preguntas la probabilidad de acertar si se contesta al azar es muy alta, del 50%.

Una observación más sobre las calificaciones obtenidas es el poco impacto que tuvo la eliminación de los reactivos inaceptables. Esto se puede ver comparando las columnas de la Tabla 8 que corresponden a los exámenes de opción múltiple con todos los reactivos y sin los reactivos inaceptables, o las líneas llena y de punto y raya en las figuras 1 a 6.

Con el fin de determinar si existían diferencias significativas estadísticamente entre las calificaciones promedio de opción múltiple y tradicionales, se aplicó a las calificaciones de cada examen una prueba t bimuestral[4]. Los resultados se muestran en las tablas 9 y 10, la primera se refiere a las calificaciones correspondientes a todos los reactivos en los exámenes de opción múltiple, y la segunda, a las que se obtuvieron sin considerar los reactivos inaceptables. Los

valores de la probabilidad *p* son pequeños en los dos parciales del trimestre 00-P y, especialmente, en el segundo examen parcial del trimestre 00-O, en la Tabla 10, por lo que no se acepta la hipótesis de que las medias sean iguales. En los otros casos de esta tabla son demasiado grandes y se acepta entonces la hipótesis de que las medias son iguales. Por lo tanto, en estos últimos exámenes no hay diferencias significativas entre las calificaciones promedio tradicionales y de opción múltiple. Llama la atención, desde luego, la gran diferencia en las calificaciones promedio del segundo parcial del trimestre 00-O: 6.00 en el examen tradicional contra 3.24 en el de opción múltiple con reactivos inaceptables eliminados. El autor no ha encontrado una explicación razonable a esta situación.

No es posible, por lo tanto, concluir con los resultados de esta investigación que los alumnos obtengan mejor calificación en un tipo de examen que en el otro. En algunos casos fueron diferentes y en otros, no lo fueron.

Ya se comentó, con referencia a las figuras 1 a 6, que no se ve claramente que los alumnos que obtienen buena calificación en los exámenes tradicionales la obtengan también en el de opción múltiple. Para afinar este análisis, se calculó el coeficiente de correlación entre las calificaciones de ambos tipos de exámenes, para cada uno de los seis exámenes aplicados[5]. Los resultados se muestran también en las tablas 9 y 10, y se puede ver que varían sensiblemente entre los diversos exámenes. Los coeficientes del global del trimestre 00-P, y del primer parcial y del global del trimestre 00-O son razonablemente buenos. Las probabilidades de haberlos obtenido al azar para la Tabla 10, sin que haya correlación, resultaron de 1, 5 y 7% respectivamente, tomando en cuenta el número de sustentantes[6]. Estas probabilidades son razonablemente pequeñas. Las figuras correspondientes también indican una mejor correlación que las de las otras figuras de la 1 a la 6. En cambio, el coeficiente de correlación del segundo parcial del trimestre 00-O resultó negativo. El impacto en los coeficientes de correlación de eliminar los reactivos inaceptables es asimismo pequeño. Y también se advierte que los trimestres en los que se obtuvo una buena correlación entre los dos tipos de examen no son los mismos que aquellos en que hubo diferencias significativas en las calificaciones promedio.

En la investigación de Kessler se pueden ver coeficientes de correlación de un examen de opción múltiple con dos exámenes tradicionales (Kessler, 1988: 707). Fueron de 0.039 y de 0.232. O sea, muy bajos, similares a los que resultaron bajos en esta investigación. Ante estos resultados, Kessler comenta que "... no queda claro cuál método (calificado por computadora o tradicional) proporciona una mejor evaluación del cumplimiento de los objetivos del curso por parte de los alumnos".

El autor considera que la mala correlación en algunos casos se debe en buena medida a que en los dos tipos de examen se están evaluando diferentes aspectos del aprendizaje. En los exámenes tradicionales, como se ha mencionado, se evalúa principalmente el aprendizaje de rutinas de cálculo, básicamente el primer nivel de la taxonomía de resolución de problemas. Mientras que en los exámenes de opción múltiple, al menos en la intención de esta investigación, se pretendía

[4] La prueba *t* se utiliza para determinar si las medias de dos poblaciones se pueden considerar iguales, si se conocen las medias de dos muestras y sus varianzas. Consiste en plantear la hipótesis de que las medias de las poblaciones son iguales, $m_1 = m_2$, y calcular el parámetro *t* que depende del tamaño de las muestras, de sus medias y de sus varianzas. Ya calculado *t*, se determina la probabilidad, *p*, de que las medias de las poblaciones sean iguales. Si esta probabilidad es pequeña, se rechaza la hipótesis, o sea, las medias de las poblaciones son diferentes, aunque hay una probabilidad de cometer un error.
[5] Un coeficiente de 1 indica una correlación directa perfecta y un coeficiente de –1 indica una correlación inversa perfecta.
[6] El cálculo se realizó como se indica, por ejemplo, en Volk (1958).

evaluar la comprensión lograda por el alumno, la cual le permitiría avanzar en niveles superiores de la taxonomía. La costumbre adquirida por los alumnos al resolver problemas tradicionales ha sido la causa de que muchos de ellos le presten poca atención a la comprensión de aspectos básicos del curso. Podría pensarse que algunos alumnos aprenden mejor las rutinas y otros los aspectos básicos de la teoría. Sería necesario, desde luego, comprobar esta hipótesis.

Conclusiones

En opinión del autor y con base en la experiencia lograda en esta investigación, los exámenes de opción múltiple constituyen una herramienta útil e interesante para evaluar el aprendizaje de los alumnos en cursos de ingeniería, sobre todo cuando se quieren evaluar en tiempos razonables niveles superiores de la taxonomía de resolución de problemas.

Los exámenes de opción múltiple presentan ventajas sobre los exámenes tradicionales, en cuanto a la posibilidad de evaluar una parte mayor del programa de la asignatura, de destinar menos tiempo del total disponible a los exámenes y de hacer un mayor número de exámenes. También tienen ventajas cuando se examina a grupos grandes o a varios grupos de una misma asignatura.

Diseñar buenos exámenes de opción múltiple no es sencillo. Se requiere tiempo y experiencia. Pero la inversión inicial de tiempo se compensa ampliamente por la mayor velocidad de calificación.

La hipótesis inicial de la investigación de que los alumnos que obtenían buena calificación en el examen tradicional también la obtenían en el de opción múltiple no pudo ser comprobada fehacientemente. En tres exámenes se obtuvieron coeficientes de correlación razonablemente buenos entre los exámenes tradicionales y los de opción múltiple, y en los otros tres, los coeficientes fueron muy bajos e inclusive uno fue negativo. En cuanto a las calificaciones promedio, en algunos casos resultaron estadísticamente semejantes, y en otros resultaron diferentes.

El autor no puede afirmar que un tipo de examen sea superior al otro. Pero sí considera que el tema del diseño de exámenes para evaluar el aprendizaje logrado por los alumnos merece más atención por parte de los profesores. Es importante que cualquiera que sea el tipo de examen se diseñen preguntas o problemas que evalúen los distintos niveles taxonómicos comentados en este trabajo. Quizá con exámenes mejor diseñados que los usados por el autor en esta investigación mejoren los coeficientes de correlación entre ambos tipos de examen. Finalmente, el autor insiste en que hay mucho campo de trabajo para perfeccionar los sistemas de evaluación del conocimiento y de las habilidades para la resolución de problemas de los alumnos de ingeniería y para mejorar, como resultado de estas evaluaciones, los sistemas de enseñanza-aprendizaje.

Referencias

EBEL, Robert L. (1979). *Essentials of Educational Measurement,* 3er edit., Prentice Hall, 116 pp.

KESSLER, David P. (1988). "Machine-Scored versus Grader-Scored Quizzes-An Experiment". *Engineering Education.* Vol. 78, no. 7, April, pp. 705-709.

LEUBA, Richard J (1986a). "Machine-Scored Testing, Part I: Purposes, Principles, and Practices". *Engineering Education* .Vol. 77, No. 2, November, pp. 89-95.

LEUBA, Richard J. (1986b). "Machine-Scored Testing, Part II: Creativity and Analysis". *Engineering Education* .Vol. 77, No. 3, December, pp. 181-186.

NATIONAL COUNCIL of EXAMINERS for ENGINEERING and SURVEYING (1996). "Guide for Writers and Reviewers in Civil Engineering".

PLANTS, H. L., Dean, R. K., Sears, J. T., and Venable, W. S. (1980). "A taxonomy of problem-solving activities and its implications for teaching." In Lubkin, J. L. (Ed.), *The Teaching of Elementary Problem Solving in Engineering and Related Fields,* ASEE, Washington, DC, 21-34.

TRISTÁN LÓPEZ, Agustín (1995). "Modelo para el análisis de reactivos objetivos por computadora", en *Memorias Foro Nacional de Evaluación Educativa*, México, CENEVAL.

VIDAL, Rafael; Leyva, Yolanda; Tristán, Agustín, y Martínez Rizo, Felipe (2000). *Manual técnico*, CENEVAL, 64 pp.

VOLK, William (1958). *Applied Statistics for Engineers.* McGraw-Hill, 354 pp.

WANKAT, Philip C. and Oreovicz, Frank S. (1993). *Teaching Engineering*, New York, McGraw-Hill, 370 p.

Tabla 1
Características de los exámenes aplicados

		Sustentantes	Preguntas opción múltiple aplicadas	Preguntas opción múltiple consideradas	Preguntas abiertas
1er Curso	1er Parcial	45	15	14	2
	2° Parcial	44	12	12	2
	Global	37	12	12	2
2° Curso	1er Parcial	24	12	11	2
	2° Parcial	17	12	12	2
	Global	17	12	12	2

Tabla 2
Resultados del examen de opción múltiple.Trimestre 00-P. Primer Parcial*

# Reactivo	Aciertos	Personas	AGS	AGI	p	P(GS)	P(GI)	PD	ND	RD
1	13	45	10	3	28.89	22.22	6.67	15.56	8.67	1.79
2	22	45	13	9	48.89	28.89	20.00	8.89	14.67	0.61
3	26	45	14	12	57.78	31.11	26.67	4.44	17.33	0.26 *
4	13	44	7	6	29.55	15.91	13.64	2.27	8.86	0.26 *
5	9	41	8	1	21.95	19.51	2.44	17.07	6.59	2.59
7	37	45	22	15	82.22	48.89	33.33	15.56	17.78	0.88
8	17	43	14	3	39.53	32.56	6.98	25.58	11.86	2.16
9	38	45	22	16	84.44	48.89	35.56	13.33	15.56	0.86
10	21	44	12	9	47.73	27.27	20.45	6.82	14.32	0.48
11	8	42	5	3	19.05	11.90	7.14	4.76	5.71	0.83
12	9	42	5	4	21.43	11.90	9.52	2.38	6.43	0.37 *
13	20	44	14	6	45.45	31.82	13.64	18.18	13.64	1.33
14	15	44	8	7	34.09	18.18	15.91	2.27	10.23	0.22 *
15	16	42	12	4	38.10	28.57	9.52	19.05	11.43	1.67
	Total Promedio =				**43.21**	**27.17**	**16.04**			

* El reactivo 6 se eliminó por contener algún error.

Tabla 3
Resultados del examen de opción múltiple. Trimestre 00-P. Segundo Parcial

# Reactivo	Aciertos	Personas	AGS	AGI	p	P(GS)	P(GI)	PD	ND	RD
1	29	44	16	13	65.91	36.36	29.55	6.82	19.77	0.34 *
2	14	44	12	2	31.82	27.27	4.55	22.73	9.55	2.38
3	5	44	5	0	11.36	11.36	0.00	11.36	3.41	3.33
4	9	43	6	3	20.93	13.95	6.98	6.98	6.28	1.11
5	9	43	6	3	20.93	13.95	6.98	6.98	6.28	1.11
6	18	44	10	8	40.91	22.73	18.18	4.55	12.27	0.37 *
7	6	44	5	1	13.64	11.36	2.27	9.09	4.09	2.22
8	15	44	11	4	34.09	25.00	9.09	15.91	10.23	1.56
9	6	44	4	2	13.64	9.09	4.55	4.55	4.09	1.11
10	22	44	13	9	50.00	29.55	20.45	9.09	15.00	0.61
11	8	44	6	2	18.18	13.64	4.55	9.09	5.45	1.67
12	17	44	12	5	38.64	27.27	11.36	15.91	11.59	1.37
	Total Promedio =				**30.04**	**20.15**	**9.89**			

Tabla 4
Resultados del examen de opción múltiple. Trimestre 00-P. Global

# Reactivo	Aciertos	Personas	AGS	AGI	p	P(GS)	P(GI)	PD	ND	RD
1	26	37	16	10	70.27	43.24	27.03	16.22	21.08	0.77
2	10	37	6	4	27.03	16.22	10.81	5.41	8.11	0.67
3	24	37	16	8	64.86	43.24	21.62	21.62	19.46	1.11
4	8	37	5	3	21.62	13.51	8.11	5.41	6.49	0.83
5	13	37	10	3	35.14	27.03	8.11	18.92	10.54	1.79
6	22	37	16	6	59.46	43.24	16.22	27.03	17.84	1.52
7	8	37	5	3	21.62	13.51	8.11	5.41	6.49	0.83
8	12	37	8	4	32.43	21.62	10.81	10.81	9.73	1.11
9	17	37	12	5	45.95	32.43	13.51	18.92	13.78	1.37
10	15	37	11	4	40.54	29.73	10.81	18.92	12.16	1.56
11	1	37	1	0	2.70	2.70	0.00	2.70	0.81	3.33
12	2	37	1	1	5.41	2.70	2.70	0.00	1.62	0.00 *
	Total Promedio =				**35.59**	**24.10**	**11.49**			

Tabla 5
Resultados del examen de opción múltiple.Trimestre 00-O. Primer Parcial*

# Reactivo	Aciertos	Personas	AGS	AGI	p	P(GS)	P(GI)	PD	ND	RD
1	7	23	6	1	30.43	26.09	4.35	21.74	9.13	2.38
2	13	24	9	4	54.17	37.50	16.67	20.83	16.25	1.28
3	15	24	11	4	62.50	45.83	16.67	29.17	18.75	1.56
4	10	24	8	2	41.67	33.33	8.33	25.00	12.50	2.00
5	15	24	7	8	62.50	29.17	33.33	-4.17	18.75	-0.22 *
7	7	23	4	3	30.43	17.39	13.04	4.35	9.13	0.48 *
8	8	24	6	2	33.33	25.00	8.33	16.67	10.00	1.67
9	8	24	4	4	33.33	16.67	16.67	0.00	10.00	0.00 *
10	3	23	3	0	13.04	13.04	0.00	13.04	3.91	3.33
11	8	24	7	1	33.33	29.17	4.17	25.00	10.00	2.50
12	7	24	5	2	29.17	20.83	8.33	12.50	8.75	1.43
	Total Promedio =				**38.70**	**26.82**	**11.88**			

* El reactivo 6 se eliminó por contener algún error.

Tabla 6
Resultados del examen de opción múltiple. Trimestre 00-O. Segundo Parcial

# Reactivo	Aciertos	Personas	AGS	AGI	p	P(GS)	P(GI)	PD	ND	RD
1	4	17	3	1	23.53	17.65	5.88	11.76	7.06	1.67
2	6	17	4	2	35.29	23.53	11.76	11.76	10.59	1.11
3	6	17	4	2	35.29	23.53	11.76	11.76	10.59	1.11
4	8	17	4	4	47.06	23.53	23.53	0.00	14.12	0.00 *
5	4	17	4	0	23.53	23.53	0.00	23.53	7.06	3.33
6	5	17	4	1	29.41	23.53	5.88	17.65	8.82	2.00
7	5	17	5	0	29.41	29.41	0.00	29.41	8.82	3.33
8	9	17	6	3	52.94	35.29	17.65	17.65	15.88	1.11
9	5	17	3	2	29.41	17.65	11.76	5.88	8.82	0.67
10	5	16	4	1	31.25	25.00	6.25	18.75	9.38	2.00
11	5	17	2	3	329.41	11.76	17.65	-5.88	8.82	-0.67 *
12	6	17	6	0	35.29	35.29	0.00	35.29	10.59	3.33
	Total Promedio =				**33.50**	**24.14**	**9.36**			

Tabla 7
Resultados del examen de opción múltiple. Trimestre 00-O. Global

# Reactivo	Aciertos	Personas	AGS	AGI	p	P(GS)	P(GI)	PD	ND	RD
1	15	16	8	7	93.75	50.00	43.75	6.25	6.25	1.00
2	14	16	7	7	87.50	43.75	43.75	0.00	12.50	0.00 *
3	11	16	8	3	68.75	50.00	18.75	31.25	20.63	1.52
4	1	16	1	0	6.25	6.25	0.00	6.25	1.88	3.33
5	4	16	2	2	25.00	12.50	12.50	0.00	7.50	0.00 *
6	12	15	7	5	80.00	46.67	33.33	13.33	20.00	0.67
7	3	16	3	0	18.75	18.75	0.00	18.75	5.63	3.33
8	2	16	1	1	12.50	6.25	6.25	0.00	3.75	0.00 *
9	11	16	6	5	68.75	37.50	31.25	6.25	20.63	0.30 *
10	5	16	3	2	31.25	18.75	12.50	6.25	9.38	0.67
11	5	16	3	2	31.25	18.75	12.50	6.25	9.38	0.67
12	3	15	2	1	20.00	13.33	6.67	6.67	6.00	1.11
Total Promedio =					**45.26**	**26.84**	**18.42**			

Notas para tablas 2 a 7:

AGS	Aciertos en el grupo superior	P(GI)	Porcentaje de aciertos del grupo inferior
AGI	Aciertos en el grupo inferior	PD	Poder de discriminación
p	Grado de dificultad	ND	Norma discriminativa
P(GS)	Porcentaje de aciertos del grupo superior	RD	Relación discriminativa

Tabla 8
Calificaciones promedio

		Examen O.M	Examen O.M Sin reactivos inaceptables	Examen Tradicional
00-P	Primer Parcial	4.04	4.29	3.19
	Segundo Parcial	3.07	2.63	3.79
	Global	3.58	3.77	3.20
00-O	Primer Parcial	3.97	3.75	3.96
	Segundo Parcial	3.34	3.24	6.00
	Global	4.39	4.25	5.27

Tabla 9
Valores de *p* y coeficientes de correlación con todos los reactivos

	Examen	Valores de *p*	Coeficiente de correlación
Trimestre 00-P	1° Parcial	0.0249	0.0944
	2° Parcial	0.0972	0.3108
	Global	0.4198	0.3764
Trimestre 00-O	1° Parcial	0.9894	0.5274
	2° Parcial	0.0003	0.0490
	Global	0.1939	0.4081

Tabla 10
Valores de *p* y coeficientes de correlación sin los reactivos inaceptables

	Examen	Valores de *p*	Coeficiente de correlación
Trimestre 00-P	1° Parcial	0.0094	0.0611
	2° Parcial	0.0119	0.2949
	Global	0.2357	0.4104
Trimestre 00-O	1° Parcial	0.7642	0.4088
	2° Parcial	0.0028	-0.2516
	Global	0.2357	0.4104

Figura 1
Calificaciones de exámenes. Trimestre 00-P. Primer parcial

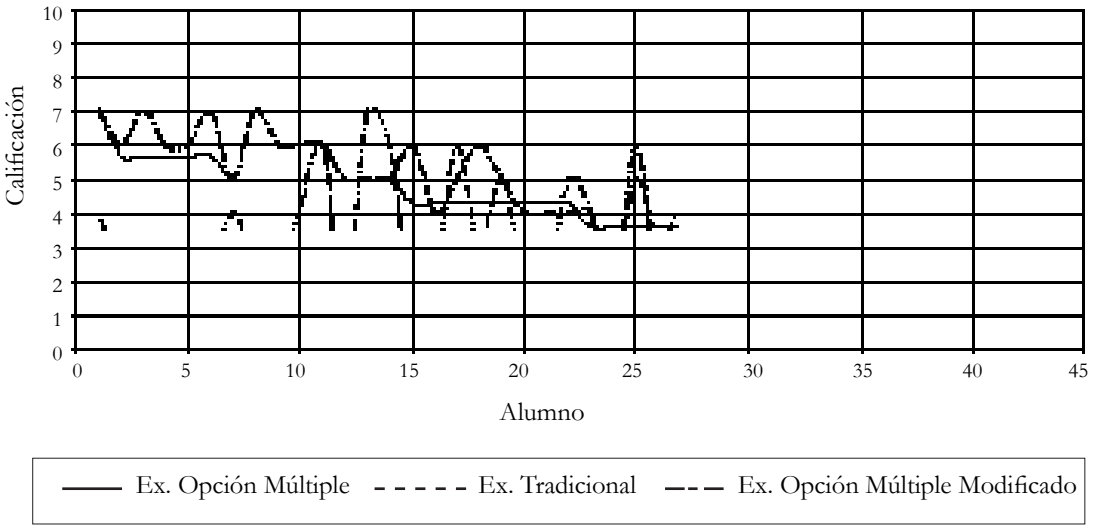

Figura 2
Calificaciones de exámenes. Trimestre 00-P. Segundo parcial

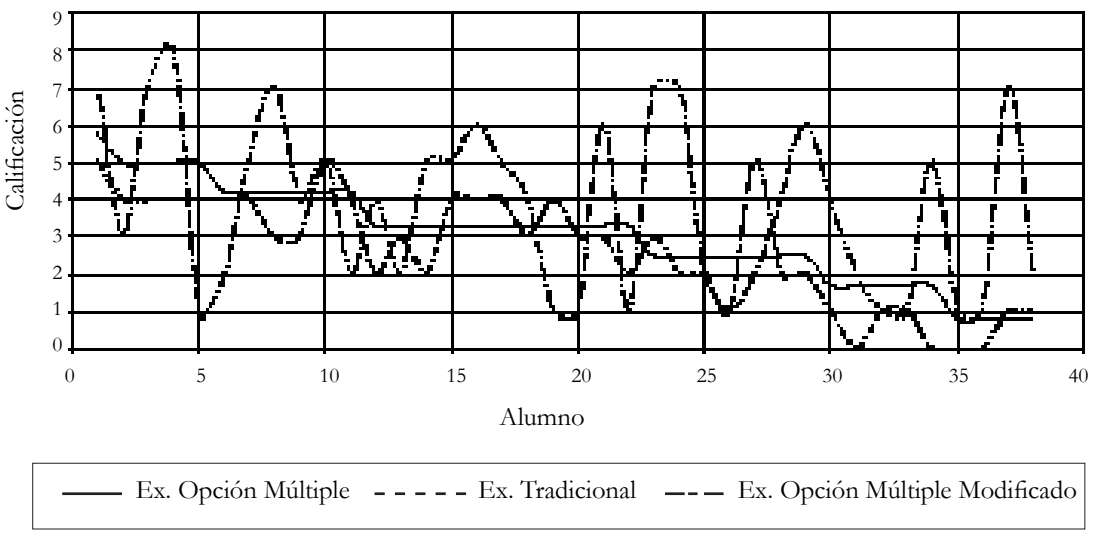

Figura 3
Calificaciones de exámenes. Trimestre 00-P. Global

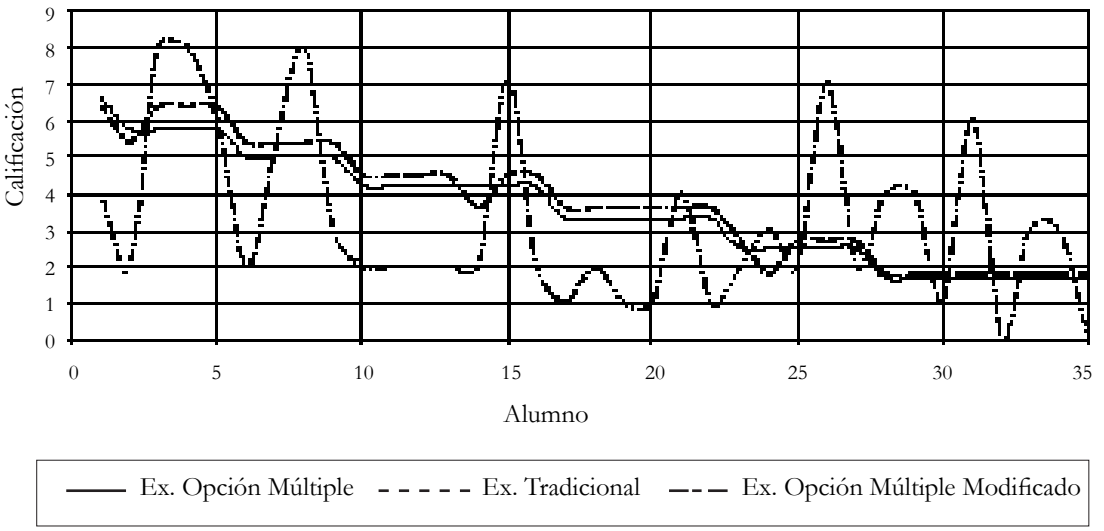

Figura 4
Calificaciones de exámenes. Trimestre 00-O. Primer parcial

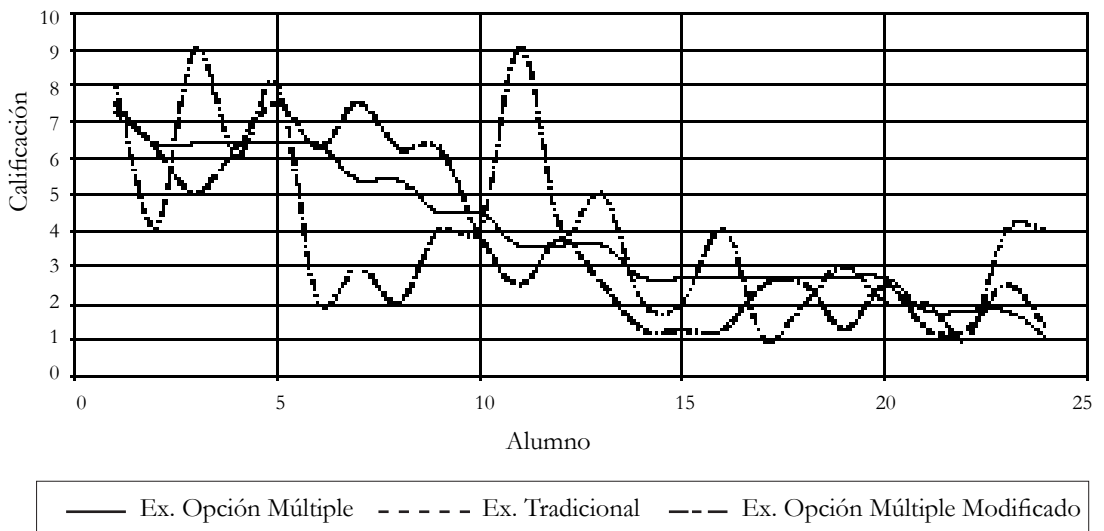

Figura 5
Calificaciones de exámenes. Trimestre 00-O. Segundo parcial

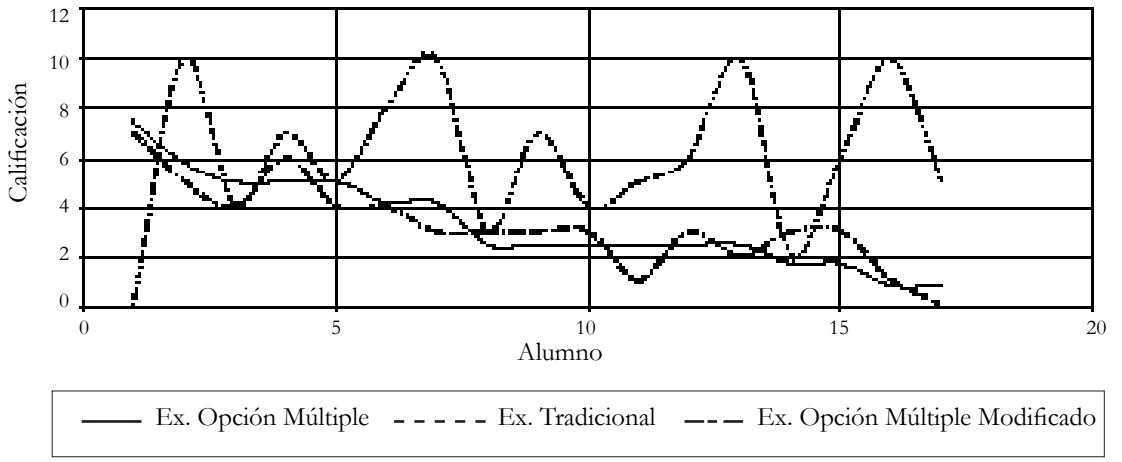

Figura 6
Calificaciones de exámenes. Trimestre 00-O. Global