



Para citar este artículo, le recomendamos el siguiente formato:

López Campos, M., Cannella, S. y Ciancimino, E. (2010). Proyecto de evaluación mixta ALT.A.: la experiencia en un entorno mexicano de enseñanza a distancia. *Revista Electrónica de Investigación Educativa*, 12(2). Consultado el día de mes de año en: <http://redie.uabc.mx/vol12no2/contenido-lopezcannella.html>

Revista Electrónica de Investigación Educativa

Vol. 12, No. 2, 2010

Proyecto de Evaluación Mixta ALT.A.: la experiencia en un entorno mexicano de enseñanza a distancia

Alternative Mixed Assessment Project (ALT.A.): The Mexican E-Learning Experience

Mónica Alejandra López Campos (1)

monlopcam@alum.us.es

Salvatore Cannella (2)

cannella@unipa.it

Elena Ciancimino (2)

eciancimino@unipa.it

(1) Departamento de Organización Industrial y Gestión de Empresas
Universidad de Sevilla

Av. de los Descubrimientos s/n
Isla de la Cartuja, 41092
Sevilla, España

(2) Departamento de Manufactura, Gestión e Ingeniería Industrial
Universidad de Palermo

Viale delle Scienze
Parco d'Orleans, 90128
Palermo, Italia

(Recibido: 26 de agosto de 2009; aceptado para su publicación: 23 de julio de 2010)

Resumen

El objetivo de este artículo es presentar un experimento de evaluación mixta (sumativa-formativa) de preguntas formuladas por los estudiantes en un entorno de educación a distancia, llevado a cabo en la asignatura de Calidad Total de la licenciatura en Ingeniería Industrial ofrecida por una Universidad pública mexicana. Preguntas generadas por los estudiantes fueron evaluadas usando una herramienta cuantitativa diseñada especialmente: la Matriz de Observación de cuatro criterios con puntuación binaria. La realización del experimento ha mostrado: 1) cómo es posible enriquecer el proceso de evaluación y formalizar habilidades de los estudiantes difícilmente reconocibles con modalidades tradicionales de evaluación; y 2) cómo la interacción profesor-estudiante puede incrementarse significativamente mediante la técnica de evaluación mixta de preguntas formuladas en sentido inverso, es decir por los estudiantes.

Palabras clave: Educación a distancia, método de evaluación, evaluación formativa, evaluación sumativa, calidad de la educación.

Abstract

The aim of this paper is to present an experiment of mixed evaluation (summative/formative) of questions formulated by students in a distance-education environment carried out in the Total Quality Management course in the B.S. degree in Industrial Engineering offered by Mexican public universities. Questions generated by students were evaluated using a specially-designed quantitative tool: Matrix Observation of four criteria with binary scoring. The experiment showed: (1) how is it possible to enrich the evaluation process, and formalize students' skills hardly recognizable with traditional forms of assessment; and (2) how the teacher-student interaction can be increased significantly by the technique mixed evaluation of questions in reverse, i.e. by the students.

Keywords: Distance education, evaluation methods, formative evaluation, summative evaluation, educational quality.

I. Introducción

En el año 2005, en Suecia, Bergman llevó a cabo una experiencia de evaluación formativa, experimentando una prueba multi-etapas con estudiantes de doctorado. La prueba consistió en pedir a los estudiantes que formularan preguntas “buenas” o “interesantes” sobre temas determinados por el examinador (Bergman, 2005). Bergman concluyó que el modelo de evaluación a preguntas hechas por los alumnos resultaba muy útil para sacar a la luz las habilidades “ocultas” de los estudiantes y para el análisis de posibles mejoras en las condiciones de evaluación que podrían derivarse de la inversión de los roles clásicos de quién pregunta y quién responde en el proceso de enseñanza-aprendizaje.

Inspirados por el estudio de Bergman, en el año 2008 Ciancimino y Cannella llevaron a cabo un experimento piloto de evaluación mixta (formativa y sumativa) en la Universidad de Palermo (Italia). La realización de este experimento piloto representó el comienzo de una iniciativa inspirada en los principios de calidad de

la educación en el marco de la estrategia de Lisboa de la Unión Europea: el proyecto ALT.A. (Alternative Assessment).

El objetivo del proyecto ALT.A. es la creación, validación y divulgación de un marco estructurado de evaluación mixta que promueva la importancia de los mecanismos de retroalimentación profesor-estudiante en el Espacio Europeo de Educación Superior: migrar desde *evaluación del aprendizaje* hacia *evaluación para el aprendizaje*.

El experimento italiano consistió en planificar y ejecutar conjuntamente dos modalidades de examen en un curso de Master en Ingeniería de Producción. La primera modalidad de examen se basó en un clásico modelo sumativo de evaluación de fin de curso, en acuerdo a las normativas nacionales de formación. La segunda modalidad de examen se planificó tomando como referencia el modelo de evaluación mixta de preguntas formuladas por los estudiantes, también llamado de “inversión de preguntas”.

Como en el experimento de Bergman, los estudiantes tenían que formular preguntas a lo largo del curso sobre distintos bloques temáticos de la asignatura; sin embargo, a diferencia del experimento de Bergman, el profesor tenía en cada sección de evaluación formativa una herramienta específicamente concebida en virtud de las características del experimento, nombrada “Matriz de Observación”. La Matriz de Observación permitió al profesor evaluar las preguntas formuladas por los estudiantes basándose en una serie de criterios: pertenencia al tópico, terminología elegida para la enunciación de la pregunta, nivel de detalle de las preguntas y expansión del rango de temas o áreas incluidas en la pregunta.

El experimento del 2008 proveyó una importante retroalimentación al profesor, el cual pudo realizar un análisis de fortalezas/debilidades del curso, remodulando al año siguiente los bloques temáticos y adoptando la técnica alternativa de inversión de preguntas como parte integral de la metodología de evaluación.

En este artículo se presentan los resultados de la segunda fase del proyecto ALT.A.: experimento realizado en México a finales del año 2009, en un entorno de educación a distancia. El objetivo de esta segunda fase es experimentar la técnica de evaluación mixta de preguntas formuladas por los estudiantes en un entorno de educación a distancia, tanto para averiguar si la realización de este tipo de evaluaciones arroja resultados positivos que mejoren el proceso de enseñanza-aprendizaje a distancia, como para mejorar las herramientas de medición de ALT.A., con miras a su posterior aplicación en otros casos.

Esta segunda fase del proyecto ALT.A. se hizo a través de un experimento realizado en la asignatura de Calidad Total, perteneciente a la licenciatura en Ingeniería Industrial de la Universidad de Guadalajara. El proceso de interacción estudiante-profesor se realizó a través del sistema de gestión de cursos de código abierto Moodle, una herramienta docente para la integración de las tecnologías de

la información y la comunicación (TIC) en la práctica educativa (García-Valcárcel, 2008).

El experimento cumplió con los objetivos predeterminados y además mostró: 1) cómo es posible mejorar el proceso de evaluación y reconocer capacidades de los alumnos difícilmente reconocibles con modalidades tradicionales de evaluación; 2) cómo la interacción estudiantes-profesor puede ser acrecentada significativamente mediante la técnica de evaluación mixta alternativa de preguntas o cuestionarios formulados en sentido inverso, es decir por los alumnos; 3) que las herramientas de medición de ALT.A. pueden ser mejoradas para su aplicación en el entorno mexicano.

El artículo está organizado como sigue. En la sección 2 se contextualiza el modelo de evaluación mixta en el entorno del Espacio Europeo de Educación Superior. La sección 3 detalla el diseño y la ejecución del experimento mexicano. La sección 4 muestra los resultados del procedimiento de evaluación y provee las discusiones. Las conclusiones son presentadas en la sección 5.

II. El espacio europeo de educación superior y el modelo de evaluación mixta de inversión de preguntas

El modelo de evaluación mixta de inversión de preguntas se coloca entre las experiencias piloto en el Espacio Europeo de Educación Superior. Como refiere Molero (2007), desde la reunión de París en la Universidad de la Sorbona en 1998, en la cual los responsables de Educación de Francia, Alemania, Italia, y Reino Unido firmaron un acuerdo que fue el origen de los procesos de convergencia en el marco de la Educación Superior, uno de los esfuerzos de la Comunidad Europea ha sido establecer una estrategia de calidad y mejora continua del Espacio Europeo de Educación Superior e intercambiar dichas estrategias a nivel global con terceros países (Fraile Aranda, 2006).

En el 2008, en el marco estratégico actualizado para la cooperación europea en el ámbito de la educación y de la formación, la Comisión de las Comunidades Europeas señaló la mejora de la calidad y eficacia de la enseñanza y de sus resultados, entre los desafíos estratégicos a largo plazo. Contextualmente se identificaron entre los temas prioritarios para el período 2009-2010 “la total integración de las competencias transversales clave en los planos de estudio, la evaluación y las cualificaciones” (Comisión Europea, 2008).

Desde la reunión de París, las comunicaciones de la Comisión Europea han delineado el proceso evolutivo del concepto “examen universitario”: “El proceso de evaluación se ha convertido en una piedra angular para la concretización de los principios guías del Espacio Europeo” (ENQA, 2007).

En la Europa del segundo Iluminismo, del *feedback*, del pensamiento sistémico y del antirreduccionismo (Rifkin, 2004), la inversión de los binomios clásicos del

proceso evaluativo “profesor-pregunta” y “estudiante-respuesta” puede identificar una línea de demarcación entre la estructura educativa milenaria (Nettleship, 1939) y el nuevo paradigma de aprendizaje continuo, caracterizado por un énfasis en la retroalimentación entre estudiantes y docentes. En este escenario los mecanismos de retroalimentación del proceso de evaluación a preguntas formuladas por los estudiantes, son algunos de los medios para realizar estrategias de calidad y de mejora continua (Cannella, Ciancimino y López Campos, 2010).

En los últimos 20 años, en la literatura se ha subrayado cómo el proceso de evaluación de preguntas formuladas por los estudiantes puede concretamente promover un pensamiento de mayor orden y mejorar el aprendizaje (Black y Wiliam, 1998). Koch y Eckstein (1991) mostraron cómo los estudiantes a quienes se enseñó a generar preguntas para sí mismos habían alcanzado mayores resultados de aprendizaje que los estudiantes que sólo atendieron las preguntas de los profesores. King (1992) encontró que entrenar a los estudiantes para que generen preguntas específicas y luego intenten resolverlas es más efectivo que entrenarlos en otras técnicas de estudio. El generar preguntas potenciales para un examen mientras se prepara para presentarlo es una técnica muy efectiva para alcanzar un alto desempeño de los estudiantes (Foos, Maray y Tkacz, 1994) y dichas preguntas, además, son una preciosa fuente de información para detectar las particularidades de los estudiantes y para mejorar el curso (Maskill y Pedrosa de Jesús, 1997). Las sugerencias de incluir en el proceso de evaluación la generación de preguntas formalizadas por los estudiantes (Biddulph, Symington y Osborne, 1986, Dori y Herscovitz, 1999), se convierte en una muestra de la evolución darwiniana en el sistema de aprendizaje. Así, la generación de preguntas por parte de los estudiantes se reconoce como pieza integral del aprendizaje (Chin, 2001).

En la literatura las preguntas de los estudiantes se han calificado a través de diferentes métodos, la mayoría de los cuales están enfocados en las habilidades de pensamiento de mayor orden (Keeling, palacek e Ingram, 2009). Dori y Herscovitz (1999) crearon un método cuantitativo para obtener un valor indicativo del nivel de complejidad de las preguntas generadas por los estudiantes. Marbach-Ad y Sokolove (2000) desarrollaron empíricamente una taxonomía de ocho clases para categorizar las preguntas según un nivel creciente de orden de pensamiento. Chin, Brown y Bruce, (2002) clasificaron las preguntas de estudiantes en prácticas de laboratorio en preguntas de “admiración” y preguntas de “información básica”. En el estudio de Barak y Rafaeli (2004) la calidad de las preguntas fue evaluada por los estudiantes y por los profesores; los estudiantes asignaron un único valor en una escala numérica, los profesores evaluaron en base al nivel cognitivo requerido para formular la pregunta, según una modificación de la taxonomía de habilidades de pensamiento de Bloom, Engelhart, Furst, Hill y Krathwhol, (1956). Bergman (2005) evaluó las preguntas de los estudiantes con un único valor de tipo pasa/no pasa.

La relación existente entre preguntas de los estudiantes y mejora continua del sistema de aprendizaje ha empujado hacia la comprensión y formalización de las informaciones contenidas en los mecanismos de retroalimentación estudiante-profesor (Ciancimino, López Campos y Cannella, 2009). Caracterizar según criterios preestablecidos las preguntas efectuadas por los estudiantes y lograr cuantificar dichas preguntas en función de criterios orientados a la mejora continua es uno de los desafíos de la comunidad científica para contribuir a la consolidación de la calidad en el Espacio Europeo de Educación Superior, el cual, a su vez, es un marco de referencia para las reformas educativas que muchos países comunitarios y extracomunitarios habrán de iniciar en los primeros años del siglo XXI (González Rodríguez, Framiñán, León y Pérez González, 2009).

En la siguiente sección se describe el diseño y ejecución del experimento realizado este estudio.

III. Diseño y ejecución del experimento

El experimento de evaluación mixta de inversión de preguntas se ha realizado según la técnica de diseño y ejecución de la evaluación de Almond, Steinberg y Misley, (2002): la Arquitectura de Cuatro Procesos. Este marco teórico define cuatro fases: 1) Selección de actividades, 2) Proceso de presentación, 3) Proceso de respuesta y 4) Sumario de la puntuación.

En cada una se determina un proceso específico dentro de la evaluación, desde la selección de los objetivos de aprendizaje hasta la evaluación final de los exámenes.

3.1 La selección de actividades

Esta fase consiste en seleccionar y secuenciar las tareas a ser cumplidas por los estudiantes.

Para este experimento, el profesor del curso a distancia de Calidad Total seleccionó tres tópicos dentro del programa de estudio: a) Orígenes y Evolución de la Gestión de la Calidad, b) Herramientas Estadísticas para la Gestión de la Calidad, y c) Control Total de Calidad. Durante seis meses los estudiantes realizaron tareas semanales (reportes, cuestionarios, problemas numéricos) y verificaciones mensuales. Adicionalmente a las actividades tradicionales de evaluación, cada dos meses se utilizó el modelo de evaluación mixta de inversión de preguntas por cada uno de los tres temas determinados.

3.2 El proceso de presentación

El objetivo de esta fase es presentar la modalidad de evaluación a los estudiantes.

La presentación del modelo de evaluación mixta de inversión de preguntas fue realizada durante el primer mes del curso a distancia y como parte de las actividades a evaluar.

Se pidió a los estudiantes que formularan una pregunta relacionada con cada uno de los tres tópicos seleccionados en la fase 1 del marco de Almond. Mediante la plataforma Moodle se formalizaron en detalle la modalidad y la agenda de las tareas, se explicaron los criterios de evaluación de las preguntas y se aportaron ejemplos guía. Ejemplos de “buenas” y “malas” preguntas fueron proporcionados para ilustrar y distinguir la alta calidad de la baja (Sadler, 1998) y para proveer una referencia general del nivel de desempeño que se pretendía alcanzar.

3.3 El proceso de respuesta

Esta fase tiene el objetivo de proveer evidencia del nivel de conocimiento real, habilidades y capacidades de los participantes.

Este proceso fue orientado a evaluar las preguntas de los estudiantes según un método cuantitativo multi-dimensional: el modelo PTNM (Pertinencia, Terminología, Nivel, Multidisciplinariedad). La Figura 1 sintetiza el modelo.



Figura 1. Síntesis del modelo PTNM

Con este método cada una de las tres preguntas se evalúa según cuatro atributos binarios:

- a₁) Pertinencia: relevancia de la pregunta formulada respecto al tema determinado. Se evalúa si la pregunta es significativa al objeto de estudio dado.
- a₂) Terminología: idoneidad de los vocablos elegidos por los estudiantes para formular sus preguntas. Se evalúa el uso de la terminología en la redacción.
- a₃) Nivel: amplitud de comprensión del tema. Este aspecto del desempeño fue incluido para tomar en cuenta el nivel de detalle de las preguntas formuladas. Se evalúa si la pregunta es trivial o si tiene un sentido más profundo o interesante, reflejo del nivel de análisis del objeto de estudio.
- a₄) Multidisciplinariedad: expansión del rango de temas o áreas incluidas en la pregunta. Tiene que ver con la capacidad de vincular el conocimiento de otras experiencias al nuevo objeto de estudio.

En esta fase el profesor recolectó las preguntas y efectuó un análisis preliminar sobre las características de cada una en función a los criterios identificados. Al finalizar cada una de las tres sesiones de evaluación mixta de inversión de preguntas, el docente redactó en Moodle un documento para cada estudiante evaluando cualitativamente la pregunta, añadiendo un comentario detallado sobre su desempeño y consejos para mejorar su aprovechamiento. En el mismo documento el enseñante insertó su respuesta a la pregunta del estudiante.

3.4 Sumario de la puntuación

Esta fase consiste en convertir las observaciones cualitativas en un valor numérico, representando el juicio del evaluador acerca del conocimiento, habilidades y capacidades mostradas por el estudiante.

Para cada atributo el examinador asigna un valor 0 si el nivel mínimo no es alcanzado, 1 si la pregunta satisface el criterio. El resultado de la evaluación de cada pregunta es un valor entre 0 y 4, que sucesivamente se normaliza entre 0 y 1 (Ver ecuación).

$$C_j = \frac{\sum_{i=1}^4 a_i}{\sum_{i=1}^4 i}$$

La evaluación final obtenida para cada estudiante en las tres preguntas se calcula como valor promedio entre los C_j , $j=1\dots3$.

En la Tabla I se reporta la Matriz de observación (Cannella, Ciancimino y López Campos, 2010) para la evaluación de las preguntas resultantes de la experimentación.

Tabla I. Matriz de observación ALT.A.

Universidad:		Profesor:		Nombre del alumno:
Licenciatura:		Curso:		
Año académico:		Créditos:		Código del alumno:
Preguntas	Pertinencia	Terminología	Nivel	Multidisciplinariedad
1	—	—	—	—
2	—	—	—	—
3	—	—	—	—

En esta fase el profesor completó la Matriz de observación de cada estudiante. Los datos resultantes se recolectaron en una hoja de cálculo y a través de la

ecuación 1 se obtuvo la calificación de cada pregunta. El valor promedio de las calificaciones obtenidas en las tres preguntas fue la nota final en el experimento de evaluación mixta de inversión de preguntas.

IV. Resultados y discusiones

Este apartado se dedica a presentar y confrontar los resultados de la evaluación tradicional y de la evaluación mixta de inversión de preguntas de los treinta estudiantes mexicanos con quienes se realizó el experimento.

El primer efecto resultante de la experiencia fue un significativo incremento del grado de interacción entre los estudiantes y el docente. Las sesiones en Moodle se incrementaron un 37% con respecto a los cursos anteriores. Los estudiantes tuvieron la posibilidad de generar sus preguntas en un marco formalizado, el cual proporcionó un estímulo a expresar dudas y perplejidades. Tal estímulo se puede considerar particularmente significativo en un contexto de interacción limitada, como lo es el sistema de aprendizaje a distancia asíncrono.

Los valores resultantes de la fase 4 del marco de Almond (2002) se reportan en la Figura 2. Se cotejan las puntuaciones obtenidas por los estudiantes en la evaluación tradicional y en la evaluación mixta alternativa de inversión de preguntas. Todas las notas son normalizadas a 1.

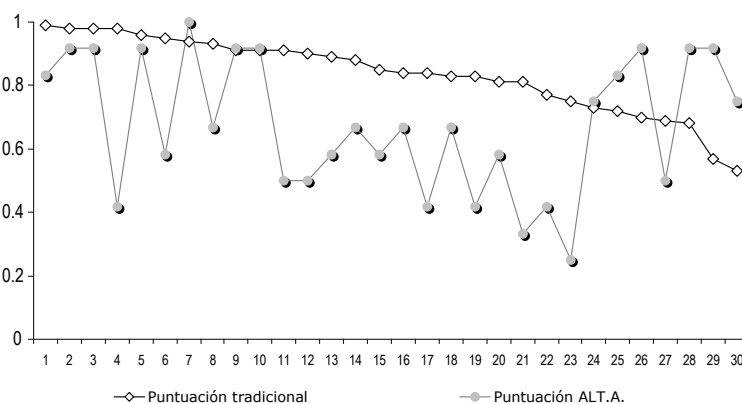


Figura 2. Puntuación tradicional y puntuación ALT.A.

En la Figura 2 se observa que es posible identificar tres distintos grupos, diferenciados por los resultados obtenidos en las dos pruebas. Los estudiantes incluidos en el primer grupo (del 1 al 10) desarrollaron bien tanto la evaluación tradicional como la mixta; el segundo grupo (11-23) tuvo un mejor desempeño en la evaluación tradicional que en la mixta; mientras que para el último grupo (del 24 al 30) se observó un mejor desempeño en la evaluación mixta; hacen excepciones las observaciones 4, 6, 8 y 27. En particular, confrontando el grupo 2 con el grupo 3, se nota que existe una tendencia opuesta entre la puntuación en la evaluación tradicional y el resultado de la evaluación mixta alternativa. Además, la

puntuación en la evaluación mixta de los estudiantes con peor desempeño en la evaluación tradicional (grupo 3) supera en valor absoluto a la puntuación obtenida por aquellos que muestran un desempeño superior en la evaluación tradicional (grupo 2).

A continuación se presenta el análisis estadístico muestral. En la Tabla II se muestran los resultados obtenidos por el cálculo de medias y desviaciones estándar muestrales y por los análisis *t* de Student y *Ji* cuadrada para la estimación de las medias y desviaciones estándar poblacionales respectivamente, tanto para la evaluación ALT.A. como para la evaluación tradicional.

Tabla II: Medias y desviaciones muestrales y su estimación poblacional para los dos tipos de evaluación

Estadístico	ALT.A.	Tradicional
Media muestral	0.6666	0.8353
Desviación estándar	0.2087	0.1211
(Prueba <i>t</i> Student)	$P(0.58 \leq \mu \leq 0.74)=95\%$	$P(0.78 \leq \mu \leq 0.88)=95\%$
(Prueba <i>Ji</i> cuadrada)	$P(0.17 \leq \sigma \leq 0.28)=95\%$	$P(0.09 \leq \sigma \leq 0.16)=95\%$

De este análisis se estima que:

- a) El 95% de los estudiantes examinados con ALT.A. tendrán una calificación entre 58 y 74; mientras que el mismo porcentaje de los estudiantes examinados con la evaluación tradicional tendrán una calificación entre 78 y 88.
- b) La desviación estándar entre las calificaciones obtenidas por el 95% de los alumnos con la evaluación ALT.A. irá de 0.17 a 0.28; mientras que el mismo porcentaje de estudiantes examinados con la evaluación tradicional tendrán una desviación estándar entre sus calificaciones de 0.09 a 0.16.

Lo anterior puede llevar a suponer que:

- a) La evaluación ALT.A. es más exigente con los alumnos, ya que requiere que los temas sean comprendidos más claramente y que sean los capaces de poner en práctica habilidades que normalmente no ejercitan con la evaluación tradicional. Por esto su calificación media es sensiblemente más baja. Un análisis de diferencia de medias, realizado con la prueba *t* de Student reveló que en un 95% de los casos la diferencia entre la evaluación ALT.A. y la evaluación tradicional es mayor a los 7 puntos: $P(0.07 \leq \mu_{x1} - \mu_{x2} \leq 0.25)=95\%$.
- b) La evaluación ALT.A. ofrece un espectro más amplio en la lectura del aprovechamiento del alumno, ya que la desviación estándar de los resultados también es sensiblemente más alta.

En la Figura 3 se reportan los porcentajes de satisfacción de los cuatro criterios adoptados para la evaluación mixta de inversión de preguntas.

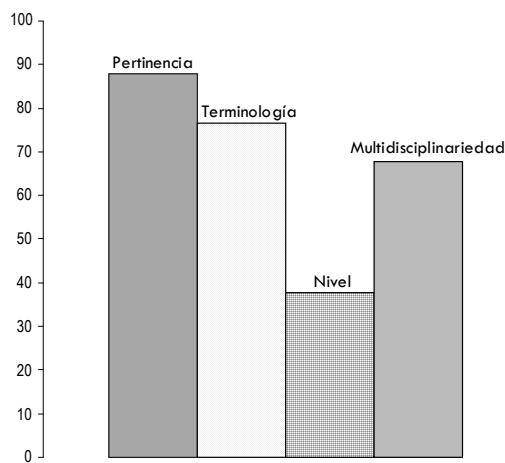


Figura 3. Satisfacción de los criterios de evaluación

En la Figura 3 se observa cómo el criterio de nivel ha sido satisfactorio en el 38% de las preguntas, contra el 88% de pertinencia, 77% de terminología y 68% de multidisciplinariedad.

Analizando conjuntamente los dos diagramas (figuras 2 y 3), se puede notar una discordancia entre la puntuación promedio obtenida por los alumnos en el examen tradicional (83% que no es baja) y el porcentaje de cumplimiento del criterio de nivel (38%, que es relativamente bajo). Este último criterio refleja la amplitud de la comprensión del tema y la profundidad de análisis del tópico. El valor relativamente bajo obtenido en el criterio de nivel podría ser indicador de una comprensión superficial de las temáticas por parte de la mayoría de los estudiantes. La discrepancia con respecto a la puntuación en el examen tradicional podría presuponer que la preparación orientada a enfrentarse con un modelo de examen tradicional no siempre es suficiente cuando se requiere un esfuerzo de conceptualización creativa, componente fundamental de la aptitud en la solución de problemas. Potenciar dicho atributo en la formación universitaria es un objetivo concreto y explícito para el Espacio Europeo de Educación Superior (Comisión de las Comunidades Europeas, 2008). Mediante la confrontación de resultados entre el examen tradicional y la evaluación mixta de inversión de preguntas, se puede suponer que la mayoría de los estudiantes involucrados en la experiencia han tenido obstáculos para convertir su conocimiento en capacidades y competencias.

A modo de recapitulación, los resultados de este experimento son los siguientes:

- 1) La utilización de la técnica de inversión de preguntas incrementó la interacción entre el profesor y los estudiantes, así como la cantidad de tiempo que los alumnos pasaban dentro de la plataforma Moodle.

- 2) Algunos estudiantes que no tenían buen desempeño con la evaluación tradicional pudieron utilizar otras cualidades y conocimientos para tener buenos resultados en la asignatura, por medio de la evaluación mixta.
- 3) La técnica de inversión de preguntas parece ser más exigente con los alumnos, ya que tiende a generar calificaciones promedio más bajas y a su vez genera más amplitud de resultados que la evaluación tradicional, ya que tiende a reportar mayores diferencias en las calificaciones de alumnos de un mismo grupo.
- 4) A pesar de que en la evaluación tradicional algunos alumnos tengan buen desempeño, la inversión de preguntas es una técnica que los obliga a ejercitar habilidades que no siempre dominan, teniendo así un desempeño no igualmente bueno, sobre todo para plantear preguntas de buen nivel.

En síntesis, los resultados subrayan que la evaluación mixta de inversión de preguntas provee al educador de una herramienta de retroalimentación estructurada, que ayuda a incrementar la interacción con los estudiantes y que puede proporcionar información adicional sobre sus logros de aprendizaje, habilidades y capacidades. Además el cotejo entre las puntuaciones del examen tradicional y del examen mixto de inversión de preguntas representa un elemento concreto a disposición del enseñante para desarrollar un análisis estructurado de fortalezas/debilidades de la asignatura en cuestión.

V. Conclusiones

En este artículo se presentaron los resultados de la segunda fase del proyecto ALT.A.: el experimento de evaluación mixta de inversión de preguntas en un entorno de educación a distancia. El experimento se realizó en la asignatura de Calidad Total de la licenciatura en Ingeniería Industrial de una universidad pública mexicana. El modelo de evaluación ALT.A. consistió en requerir a los estudiantes que generaran preguntas de temas específicos del programa de la asignatura. Las preguntas fueron evaluadas usando un sistema de soporte para la evaluación: el modelo PTNM por medio de la Matriz de Observación, una herramienta cuantitativa basada en una evaluación binaria de distintos criterios.

La segunda fase del proyecto ALT.A. arrojó varios resultados, los cuales se pueden sintetizar en dos conclusiones básicas, cumpliendo así los objetivos de este experimento: buen nivel.

- 1) La adopción de un examen formativo, y en particular su formalización en un marco estructurado capaz de proporcionar información cuantitativa, puede representar una herramienta concreta de mejora del sistema de aprendizaje, en cuanto proporciona información complementaria a la clásica retroalimentación contenida en los exámenes tradicionales. Particularmente

en un entorno de educación a distancia, los beneficios de una metodología alternativa de evaluación como ALT.A. se perciben ampliamente.

- 2) La formalización de un examen basado sobre la propositiva participación de los alumnos ha permitido un incremento de la interacción profesor-estudiante. Los estudiantes se sintieron estimulados para expresar sus dudas en un contexto de interacción limitada, como lo es el sistema de aprendizaje a distancia asíncrono. En este sentido, ALT.A. se convirtió en una herramienta de ayuda para el aprendizaje y coparticipación por parte de los estudiantes.
- 3) La tercera fase de este proyecto involucrará un experimento estructurado en la Universidad de Sevilla (España), con un grupo de control. Las herramientas de medición, en especial la Matriz de Observación, serán mejoradas de acuerdo con la experiencia a distancia ALT.A. (Apéndice A) y serán armonizadas con los objetivos de aprendizaje definidos por el Marco Europeo de Cualificaciones (conocimientos, habilidades y competencias) adoptado por el Parlamento del Concilio Europeo el 23 de Abril del 2008 (Comisión Europea, 2008).

VI. Apéndice

Observaciones del método ALT.A. y algunas ideas para mejoras futuras

Esta sub sección está dedicada a presentar los comentarios del profesor tanto a propósito de la herramienta cuantitativa propuesta para evaluar las preguntas de los estudiantes, como con respecto al método en sí mismo:

- a) El uso del lenguaje y la calidad de la redacción podría ser un elemento independiente de evaluación más, y no sólo penalizable desde el criterio del uso de la terminología. En particular, en el contexto mexicano, las cuestiones de ortografía y sintaxis podrían ser materia importante para evaluar dentro de la Matriz de Observación dado que, según la experiencia del docente, estos temas son frecuentemente un punto de mejora para los estudiantes.
- b) Responder las preguntas de los estudiantes resultó más interesante y útil para el profesor y para los alumnos que sólo realizar la evaluación (sumativa) del cumplimiento de los criterios, esto podría ser formalmente incluido en el procedimiento ALT.A. Cabe decir que el hecho de que el curso se llevara a cabo de forma virtual y asíncrona facilitó este proceso.
- c) Sería deseable tener un grado intermedio de evaluación entre 0 y 1. Haciendo esto, el análisis del cumplimiento de los criterios arrojaría más información que la simple calificación cumple/no cumple. Además, un tercer grado facilitaría al examinador la tarea de calificar.

- d) Se sugiere, además, asignar diferentes pesos a los criterios de la evaluación. En particular, el criterio de Nivel debería pesar más que los restantes.
- e) Se observó que es necesario especificar de forma muy explícita a los alumnos e insistir en que las preguntas que ellos elaboren no formarán parte de futuros exámenes.

Referencias

- Almond, R., Steinberg, L. y Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. *The Journal of Technology, Learning, and Assessment*, 1(5), 1-63.
- Barak, M. y Rafaeli, S. (2004). On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning. *International Journal of Human Computer Studies*, 61(1), 84-103.
- Bergman, B. (2005, julio). Interactive assessment. A new element in the examination process. *Proceedings of 8th Quality Management and Organisational Development Conference*, 959-966.
- Biddulph, F., Symington, D. y Osborne, R. (1986). The place of children's questions in primary science education. *Research in Science and Technological Education*, 4(1), 77-88.
- Black, P. y Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Bloom, B. S., Engelhart, M. B., Furst, E. J., Hill, W. H. y Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals, Handbook 1: The cognitive domain*. Nueva York: Longmans Green.
- Cannella, S., Ciancimino, E. y López Campos, M. (2010). Mixed e-Assessment: an application of the student-generated questions technique. *IEEE EDUCON*, April 14-16, Madrid.
- Chin, C. (2001). Learning in science: What do students' questions tell us about their thinking? *Education Journal*, 29(2), 85-103.
- Chin, C., Brown, D. E. y Bruce, B. C. (2002). Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*, 24(5), 521-549.
- Ciancimino, E. y Cannella, S. (2008). The Alternative Assessment Experiment. *The International Journal of Learning*, 15(2), 295-300.

Ciancimino, E., López Campos, M. y Cannella, S. (2009). ALTA project. The e-learning experience. *12th QMOD*. August 27-29, Verona, Italia.

Comisión Europea (2008). Recommendation of the European Parliament and of the Council of 23 April 2008 on the establishment of the European Qualifications Framework for lifelong learning. *Official Journal of the European Union*, C111/01. Bruselas: Autor.

Dori, Y. J. y Herscovitz, O. (1999). Question-posing capability as an alternative evaluation method: analysis of an environmental case study. *Journal of Research in Science Teaching*, 36(4), 411-430.

European Association for Quality Assurance in Higher Education (2007). *Report on standards and guidelines for quality assurance in the european higher education area* (2a. ed.). Helsinki: Autor.

Foos, P. V., Mora, J. J. y Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology*, 86(4), 567-576.

Fraile Aranda, A. (2006). El sistema universitario europeo como modelo posible para la educación superior latinoamericana. *Revista Electrónica de Investigación Educativa*, 8(1).

García-Valcárcel, A. (2008). La tutoría en la enseñanza universitaria y la contribución de las TIC para su mejora. *Revista Electrónica de Investigación y Evaluación Educativa*, 14(2). Consultado el 1 de julio de 2009, de: http://www.uv.es/RELIEVE/v14n2/RELIEVEv14n2_3.htm.

González Rodríguez, P. L., Framiñán Torres, J. M., León Blanco J. M., Pérez González, P. (2009, 2-4 de septiembre). Experiencias en la aplicación de un nuevo sistema de evaluación formativa en el ámbito del Espacio Europeo de Educación Superior. 3rd International Conference on Industrial Engineering and Industrial Management. Barcelona.

Keeling, E. L., Polacek, K. M. e Ingram, E. L. (2009). A statistical analysis of student questions in a cell biology laboratory. *CBE Life Sciences Education*, 8(2), 131-139.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, 27(1), 111-126.

Koch, A. y Eckstein, S. G. (1991). Improvement of reading comprehension of physics texts by students' question formulation. *International Journal of Science Education*, 13(4), 473-485.

Marbach-Ad, G. y Sokolove, P. G. (2000). Can undergraduate biology students learn to ask higher level questions? *Journal of Research in Science Teaching*, 37(8), 854-870.

Maskill, R. y Pedrosa de Jesus, H. (1997). Pupils' questions, alternative frameworks and the design of science teaching. *International Journal of Science Education*, 19(7), 781-799.

Molero López Barajas, D. (2007). Rendimiento académico y opinión sobre la docencia del alumnado participante en experiencias piloto de implantación del Espacio Europeo de Educación Superior. *Revista Electrónica de Investigación y Evaluación Educativa*, 13(2). Consultado el 1 de julio de 2009 de: http://www.uv.es/RELIEVE/v13n2/RELIEVEv13n2_2.htm.

Nettleship, R. L. (1939). *The Theory of education in Plato's Republic*. Londres: Oxford University Press.

Rifkin, J. (2004). *The European dream*. Cambridge: Polity press.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.